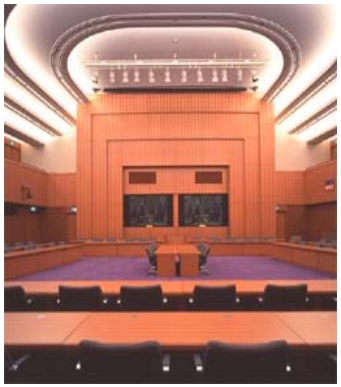
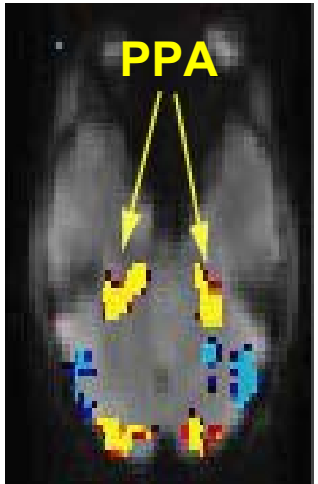
A scene-centered representation of gist

Aude Oliva

Brain & Cognitive Sciences
Massachusetts Institute of Technology
Email: oliva@mit.edu http://cvcl.mit.edu

VSS 2007 Symposium: Natural Scene Understanding: Statistics, Recognition and Representation





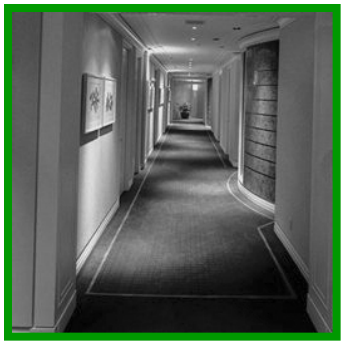
Demo
Remember the pictures

The classical RSVP task
Potter (1975)

You do not comprehend “everything”

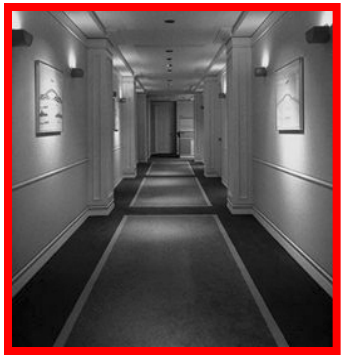
You have seen these pictures



You were tested with these pictures

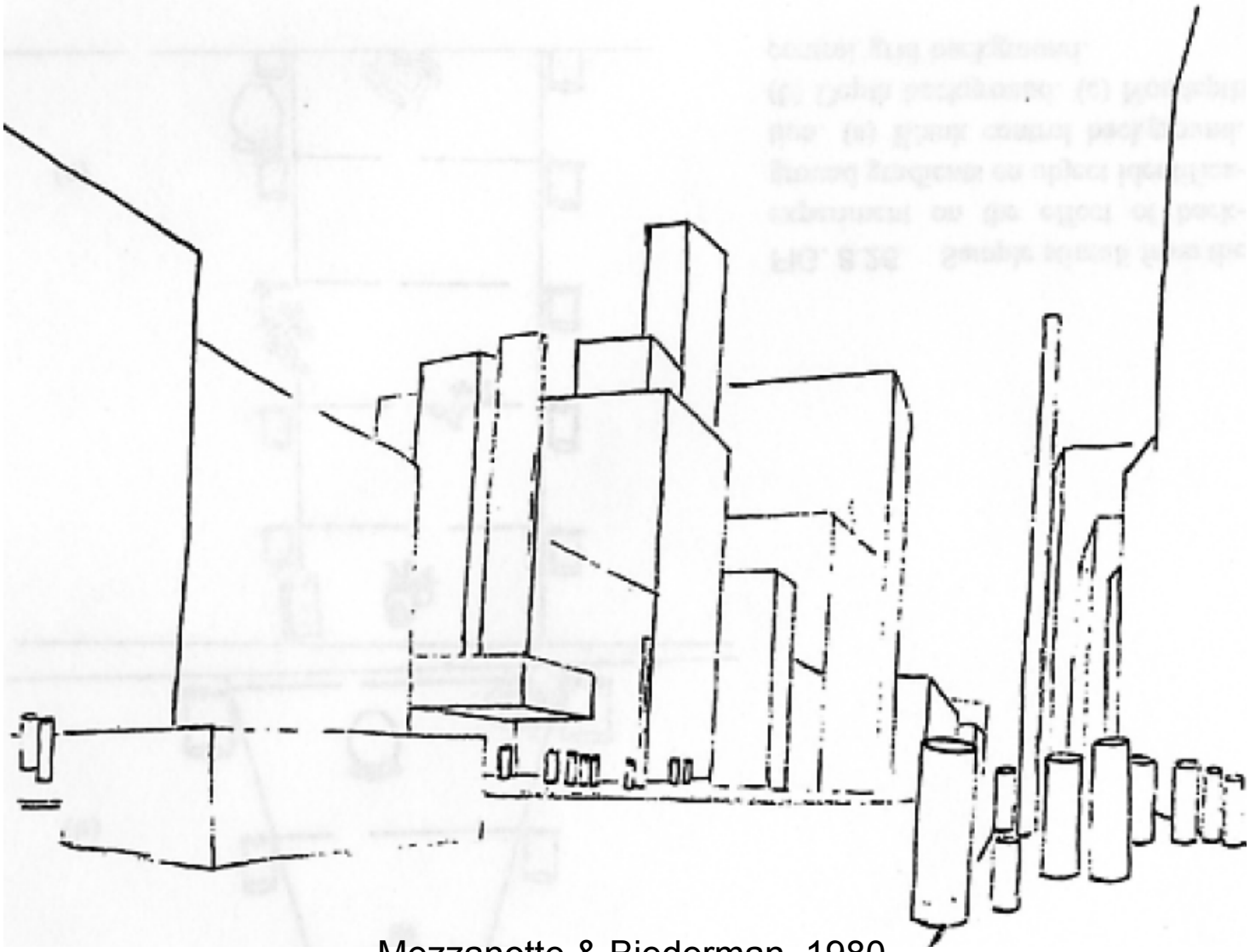


- Can we identify the gist of a scene from other information than objects?

Reducing the objects
Enhancing the scene

Reducing the objects
Enhancing the scene



Mezzanotte & Biederman, 1980

Recognizing the Forest before the Trees

Navon (1977)

How do we recognize the forest in the first place?

Hints of Globality: Spatial Structure

Forests are “enclosed”






Beaches are “open”






Spatial Envelope Representation

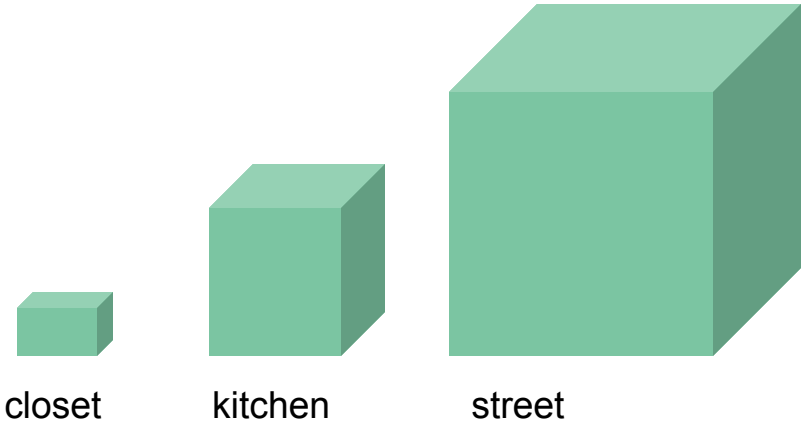
Global Properties diagnostic of the space the scene subtends provide the basic level of the scene

(1) Boundary of the space

Mean depth

Openness

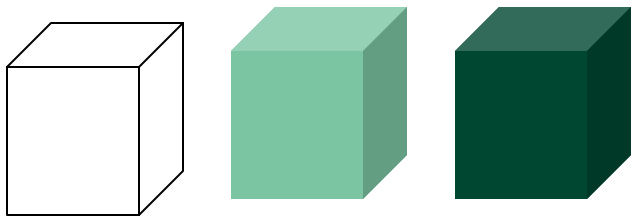
Perspective



(2) Content of the space

Naturalness

Roughness



Oliva & Torralba (2001, 2002, 2006)

Spatial Envelope Representation

Global Properties diagnostic of the space the scene subtends provide the basic level of the scene

(1) <u>Boundary of the space</u>

Mean depth

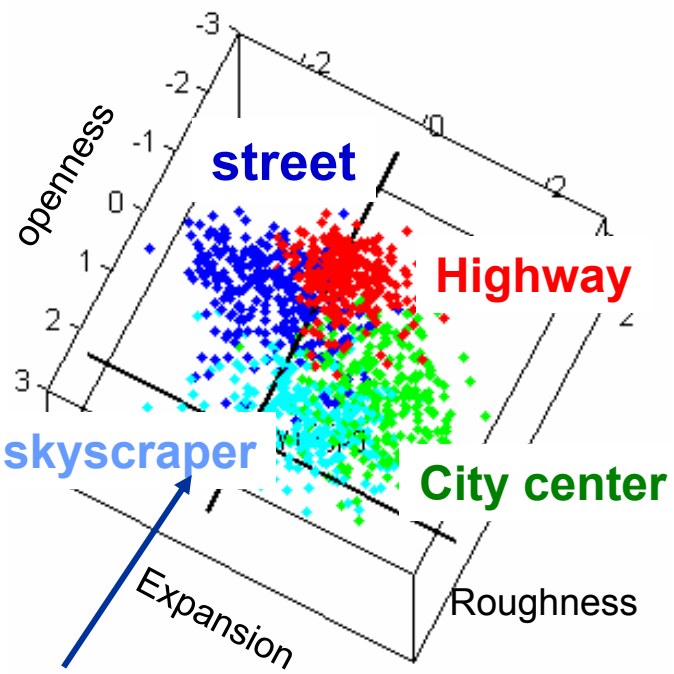
Openness

Perspective

(2) <u>Content of the space</u>

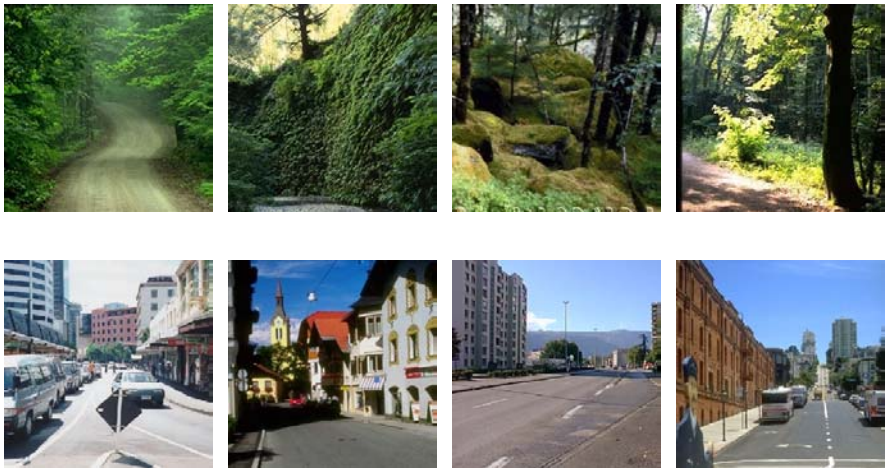
Naturalness

Roughness



Oliva & Torralba (2001, 2002, 2006)

Spatial Envelope Representation



Proposal 1: "Scenes of the same feather flock together"

- Scenes of the same category membership share similar spatial envelope properties

Proposal 2: "Faster than you can say feed-forward "

- Spatial envelope properties arise from low level building blocks

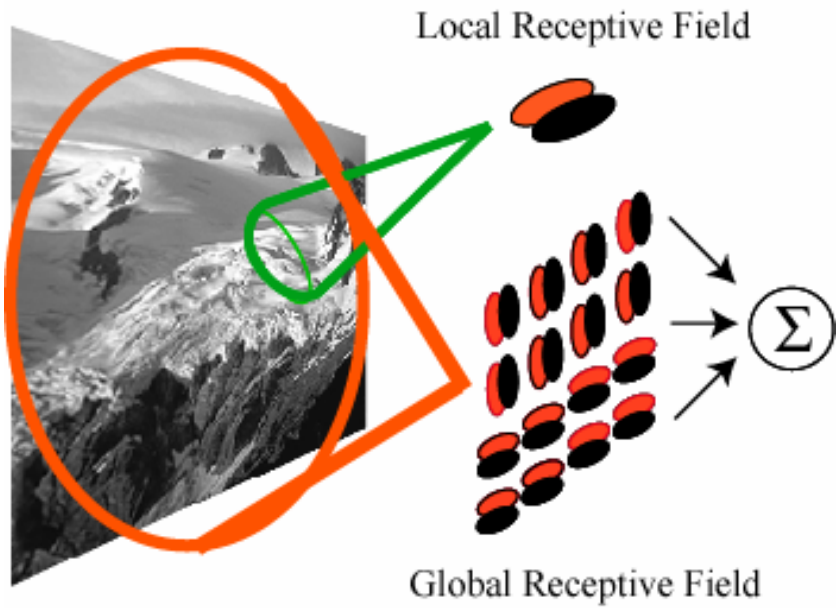
Oliva & Torralba (2001)

Degree of Openness

Given human ranking of how *open to enclosed* a given scene image is, the goal is to find the low level features that are correlated with “openness”

From open scenes

to closed scenes



High degree of Openness

Lack of texture

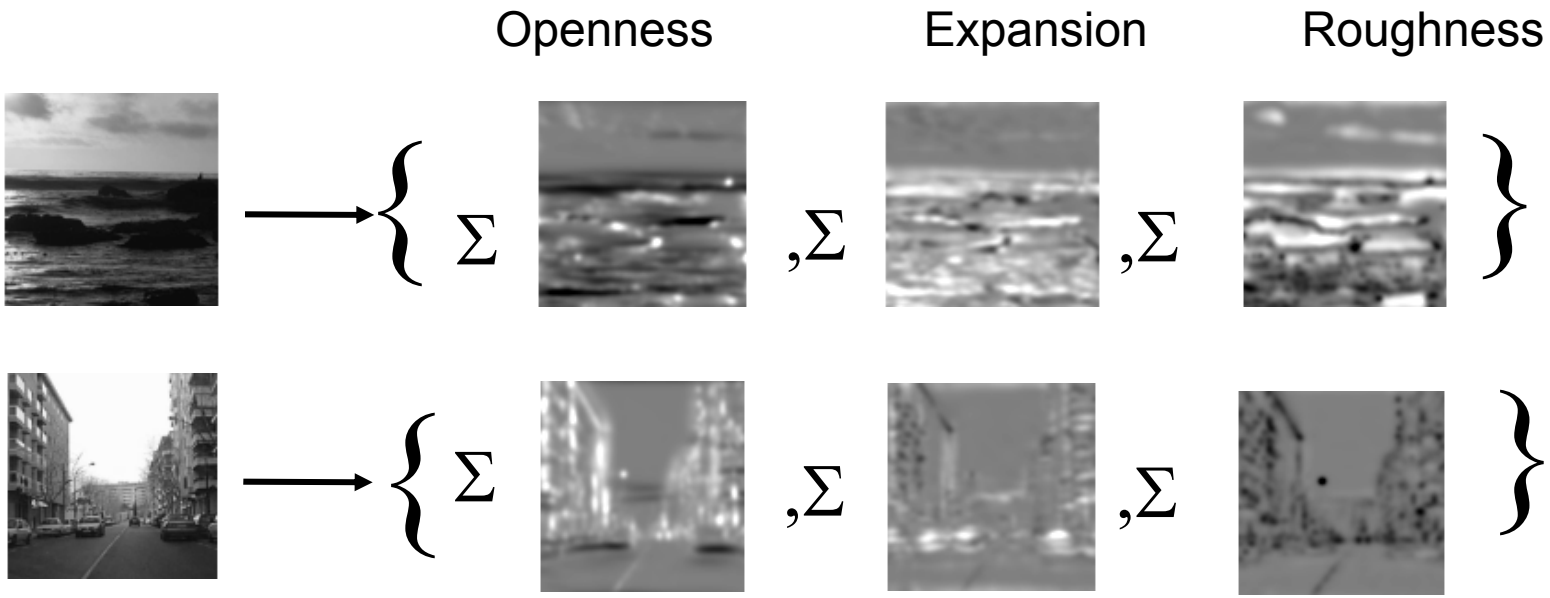
Low spatial frequency horizontal

High spatial frequency isotropic texture

Oliva & Torralba (2001, 2006)

Spatial Envelope Representation

- A scene image is represented by a vector of values for each spatial envelope property.
- For instance:



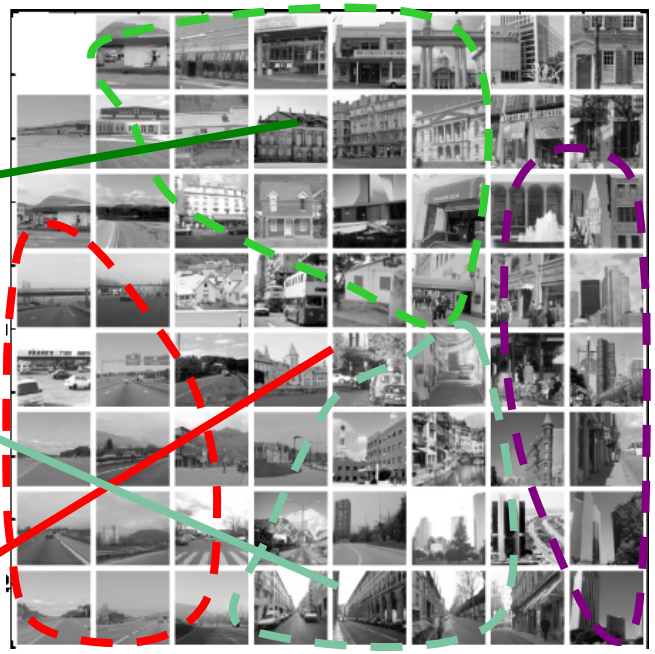
Oliva & Torralba (2001)

Categorization of Urban Scenes

Confusion Matrix
Classification of prototypical scenes (400 / category)

	Highway	Street	City centre	tall building
Highway	**91.6**	4.8	2.7	0.9
Street	4.7	**89.6**	1.8	3.4
Centre	2.5	2.3	**87.8**	7.4
Tall Building	0.1	3.4	8.5	**88**

Local organization:
correct for 86 % images
(4 similar images on 7 K-NN)



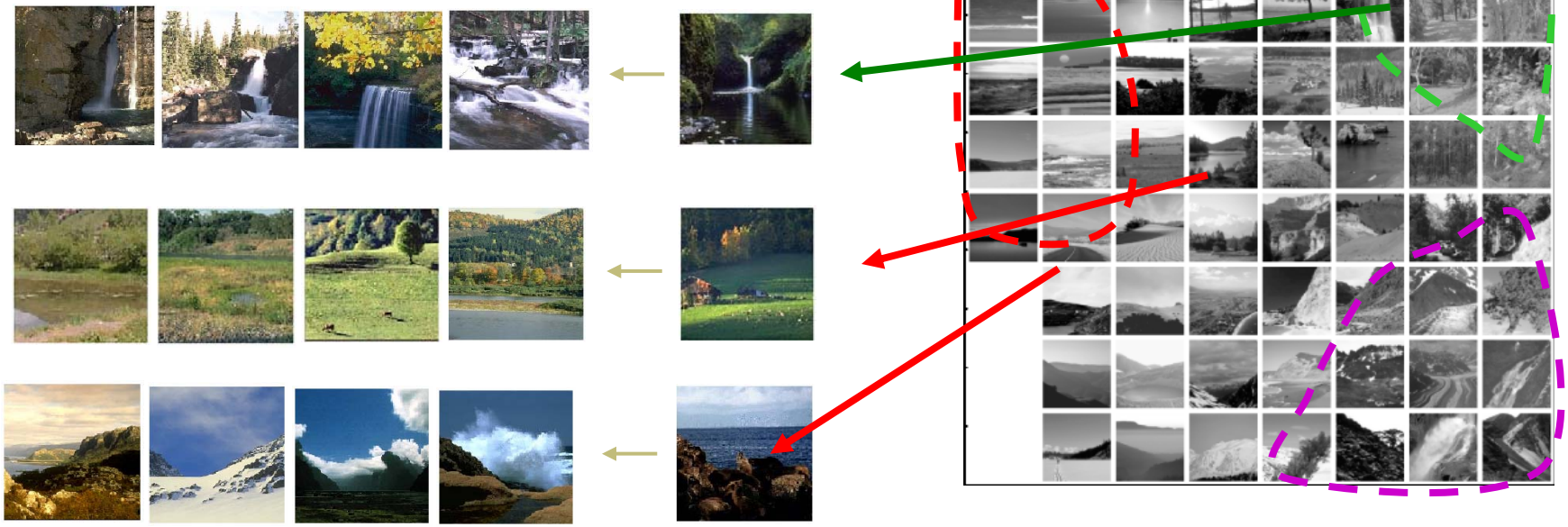
Oliva & Torralba (2001)

Categorization of Natural Scenes

Confusion Matrix
Classification of prototypical scenes (400 / category)

	Coast	Countryside	Forest	Mountain
Coast	**88.6**	8.9	1.2	1.3
Countryside	9.8	**85.2**	3.7	1.3
Forest	0.4	3.6	**91.5**	4.5
Mountain	0.4	4.6	3.8	**91.2**

Local organization:
correct for 92 % images
(4 similar images on 7 K-NN)



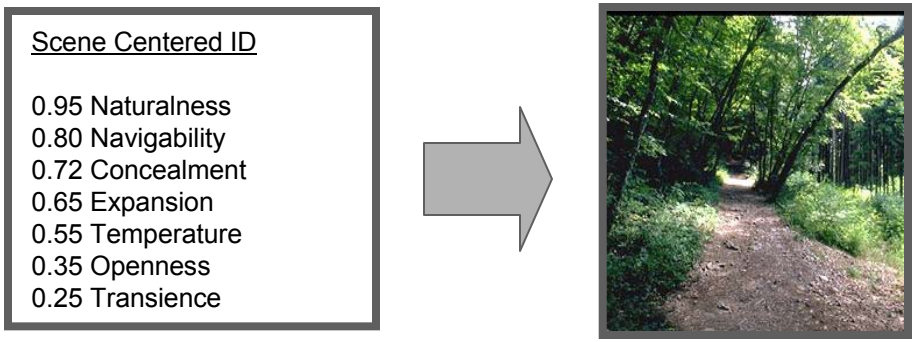
Oliva & Torralba (2001)

Psychological Validity

Can a scene-centered representation predict human scene basic level recognition ?

Global properties

"forest"



In collaboration with Michelle Greene

Greene & Oliva (2006, submitted)

Approach: Errors Prediction

Two scenes with a similar scene-centered representation but different categorical membership should be confused with each other (false alarm)

Closed space
Low navigability

Open space
High navigability

Coast

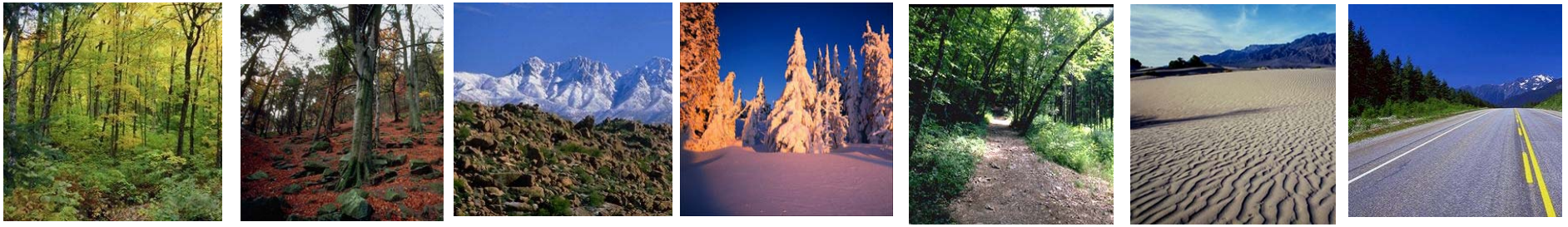
Forest

Field

Greene & Oliva (2006, submitted)

Experiment 1: Scene centered representation of natural images

Potential for Navigation



Difficult to walk through ⟷ Easy to walk

Mean depth

Small volume ⟷ large volume

Greene & Oliva (2006, submitted)

Scene-centered categorical similarities

	Desert	Field	Forest	Lake	Mount	Ocean	River
Desert							
Field	0.7						
Forest	0.1	0.1					
Lake	0.5	0.6	0.2				
Mount	0.3	0.3	0.3	0.6			
Ocean	0.4	0.4	0.1	0.7			
River	0.0	0.0	0.4	0.4	0.4	0.5	
Waterfall	0.0	0.0	0.3	0.3	0.3	0.4	0.7

Matrix of similarity between Scene Categories

Greene & Oliva (2006, submitted).

Exp. 2: Fast scene categorization task

Is it a forest?

30 msec image + mask

Exp. 2: Fast scene categorization task

Is it a forest?

30 msec image + mask

Exp 2 Result: False Alarms matrix

	Desert	Field	Forest	Lake	Mount	Ocean	River
Desert							
Field	0.29						
Forest	0.11	0.16					
Lake	0.07	0.15	0.09				
Mount	0.16	0.16	0.10	0.21			
Ocean	0.11	0.16	0.07	0.25			
River	0.09	0.13	0.16	0.19	0.14	0.21	
Waterf	0.06	0.06	0.11	0.09	0.14	0.13	0.29

Matrix of false alarms between Scene Categories

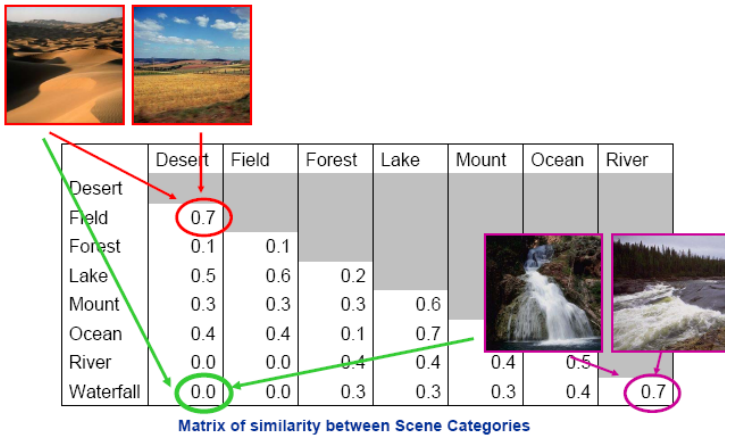
Greene & Oliva (2006, submitted).

Scene-centered representation predicts human false alarms

Scene-Centered Representation

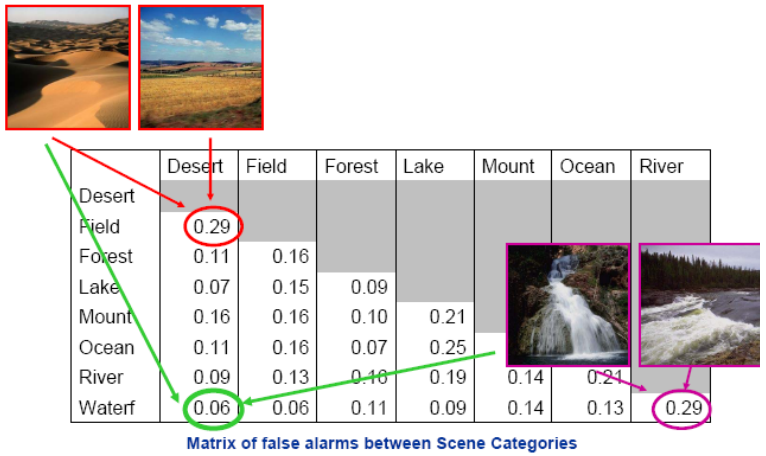
0.76

False alarms Scene categories



	Desert	Field	Forest	Lake	Mount	Ocean	River
Desert							
Field	0.7						
Forest	0.1	0.1					
Lake	0.5	0.6	0.2				
Mount	0.3	0.3	0.3	0.6			
Ocean	0.4	0.4	0.1	0.7			
River	0.0	0.0	0.4	0.4	0.4	0.5	
Waterfall	0.0	0.0	0.3	0.3	0.3	0.4	0.7

Matrix of similarity between Scene Categories



	Desert	Field	Forest	Lake	Mount	Ocean	River
Desert							
Field	0.29						
Forest	0.11	0.16					
Lake	0.07	0.15	0.09				
Mount	0.16	0.16	0.10	0.21			
Ocean	0.11	0.16	0.07	0.25			
River	0.09	0.13	0.10	0.19	0.14	0.21	
Waterf	0.06	0.06	0.11	0.09	0.14	0.13	0.29

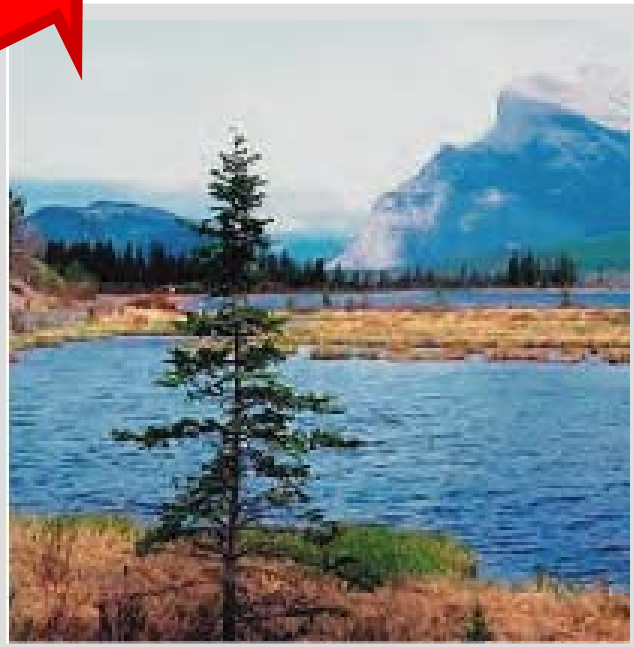
Matrix of false alarms between Scene Categories

Image analysis (distance of each distractor to the target category) shows the same high correlation.

Greene & Oliva (2006, submitted).

How far can we go in explaining scene recognition without objects?

A lake



Scene-Centered Representation

100% natural space
66% open space
64% perspective
74% deep space
68% cold place

Object-Centered Representation

23% sky
35 % water
18% tree
12 % mountain
% grass

Greene & Oliva (2006, submitted).

Exp 3: How *sufficient* is a scene-centered representation?

Method: Compare a naïve Bayes classifier to human performance.

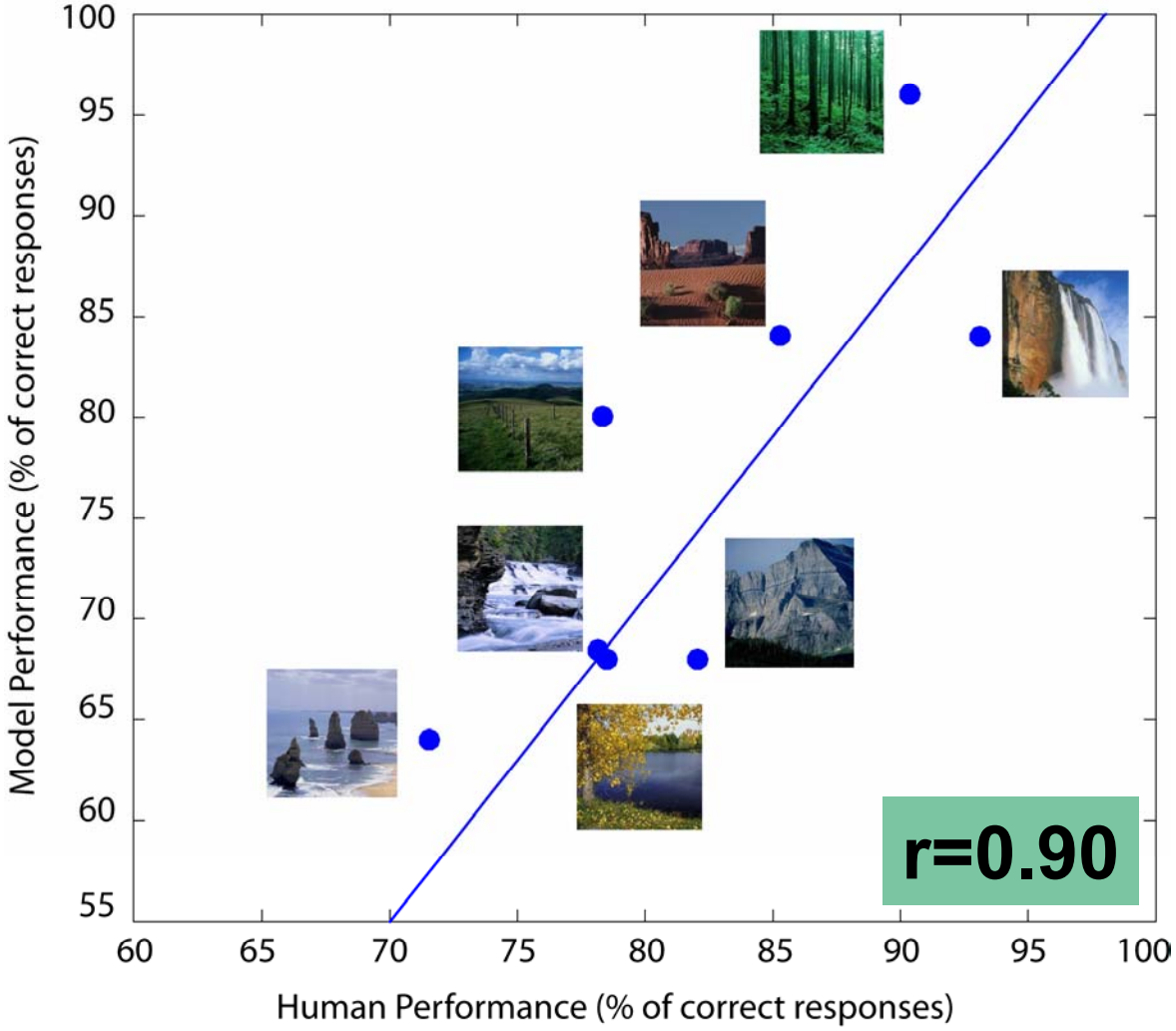
Given a novel image	Scene-centered Signature	Probable Semantic Class

0.36 Camouflage
0.38 Movement
0.94 Navigability
0.99 Temperature
0.89 Openness
0.68 Expansion
0.83 Mean depth

→ "desert"

Greene & Oliva (2006, submitted).

A scene-centered classifier predicts fast scene categorization



Greene & Oliva (2006, submitted).

A scene-centered classifier predicts human false alarms

Given a misclassification of the classifier, at least one human observer made the same false alarm in 87% of the images

river

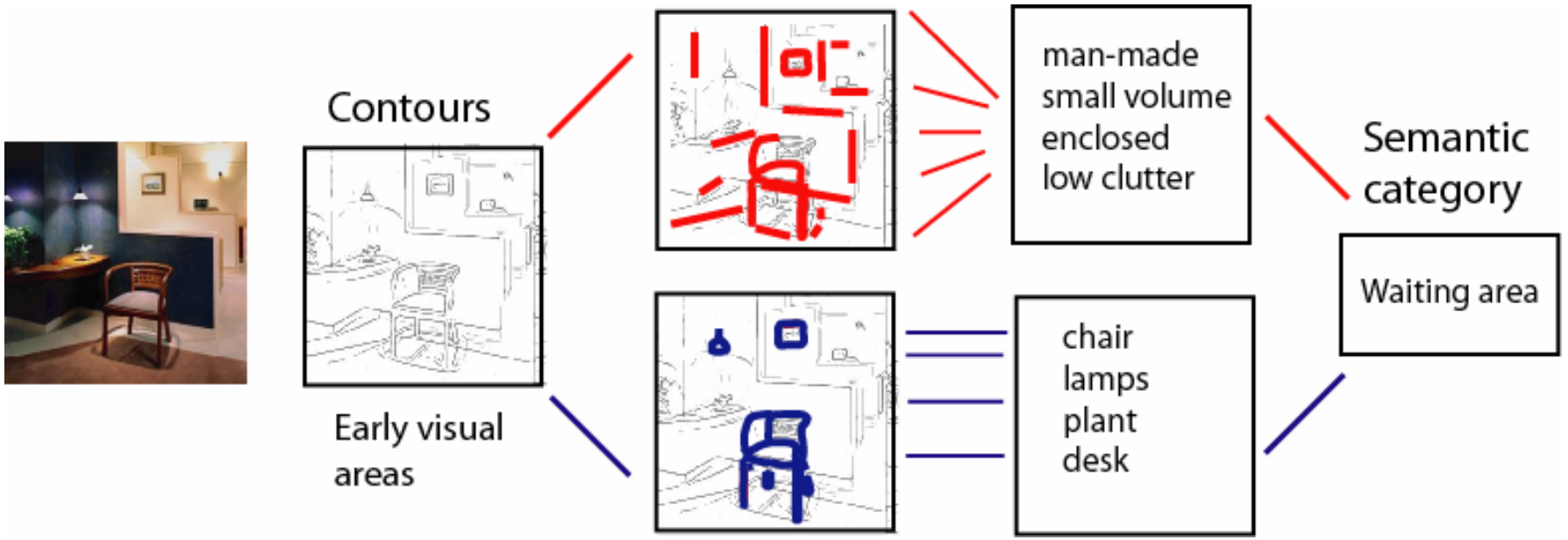
ocean (oups..)"

desert

field (oups..)"

Greene & Oliva (2006, submitted).

Scene Gist Representation Framework



References available at **http://cvcl.mit.edu/publications.htm**

Greene, M.R., & Oliva, A. (submitted). Scene Gist from Global Properties: Seeing the Forest without representing the Trees.

Greene, M.R., & Oliva, A. (2006). Natural scene categorization from the conjunction of ecological global properties. Proceeding of the Cognitive Science Meeting, Vancouver, August 2006.

Oliva, A. & Torralba, A. (2006). Building the Gist of a Scene: The Role of Global Image Features in Recognition. Progress in Brain Research: Visual perception, 155, 23-36.

Torralba, A., & Oliva, A. (2003). Statistics of Natural Images Categories. Network: Computation in Neural Systems, 14, 391-412.

Torralba, A., & Oliva, A. (2002). Depth estimation from image structure. IEEE Pattern Analysis and Machine Intelligence, 24,1226-1238.

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: a Holistic Representation of the Spatial Envelope. International Journal in Computer Vision, 42, 145-175.

Sponsors
National Science Foundation
CAREER award - IIS Program
NEC award in computers and communication