BIL 722 – Advanced Topics in Computer Vision

Nov. 1, 2012

Aykut Erdem
Dept. of Computer Engineering
Hacettepe University

Tutorial

The Role of Context

The Role of Context

"*Some entity Z can have certain properties, when Z is viewed in isolation, which change when Z is viewed in some context. Alternately, an entity Z is seen as one thing in context A and another in context A*"

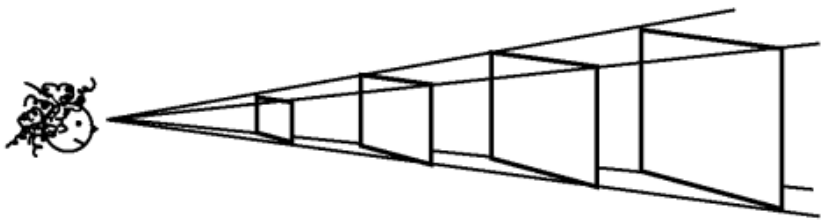
The use of context in pattern recognition, G.T. Toussaint, Pattern Recognition, 10(3): 189- 204 (1978)

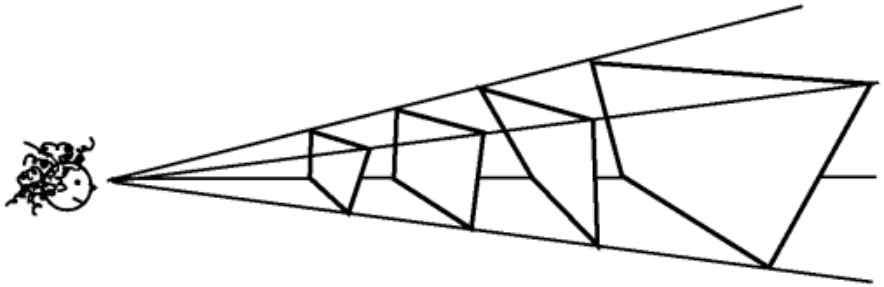
12

14

Visual Perception

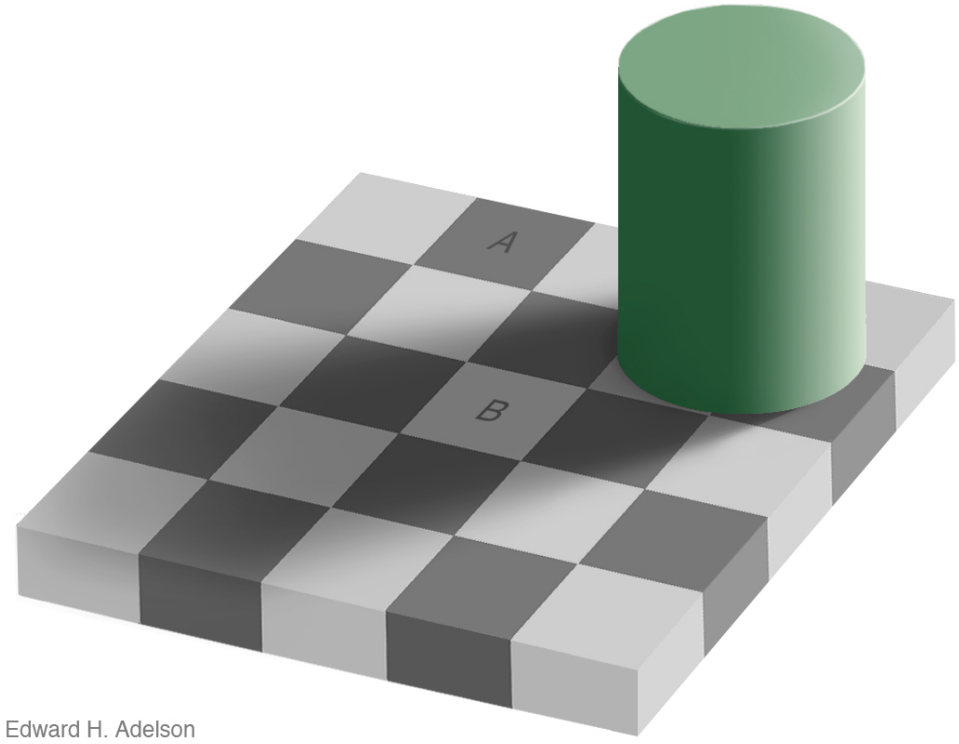
- Vision requires solving ill-posed problems.
- Images are both complicated and highly ambiguous.
- Many different interpretations are possible for the same visual input





Figures: Steven Pinker, How the Mind Works, 1997

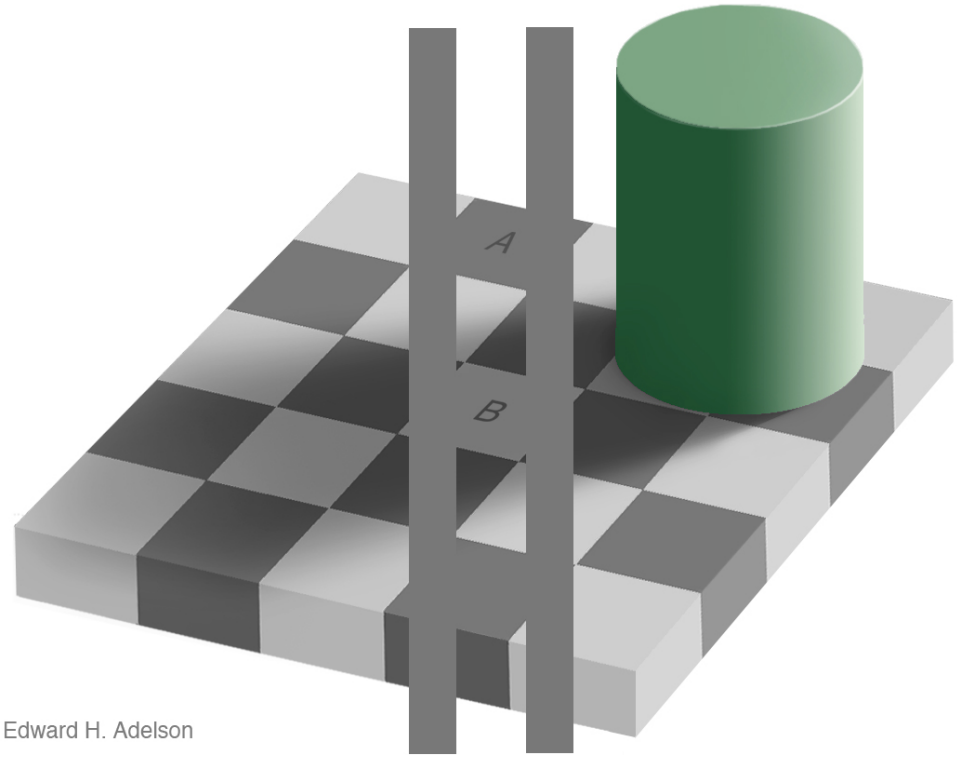
Brightness perception



Edward Adelson

http://web.mit.edu/persci/people/adelson/illusions_demos.html

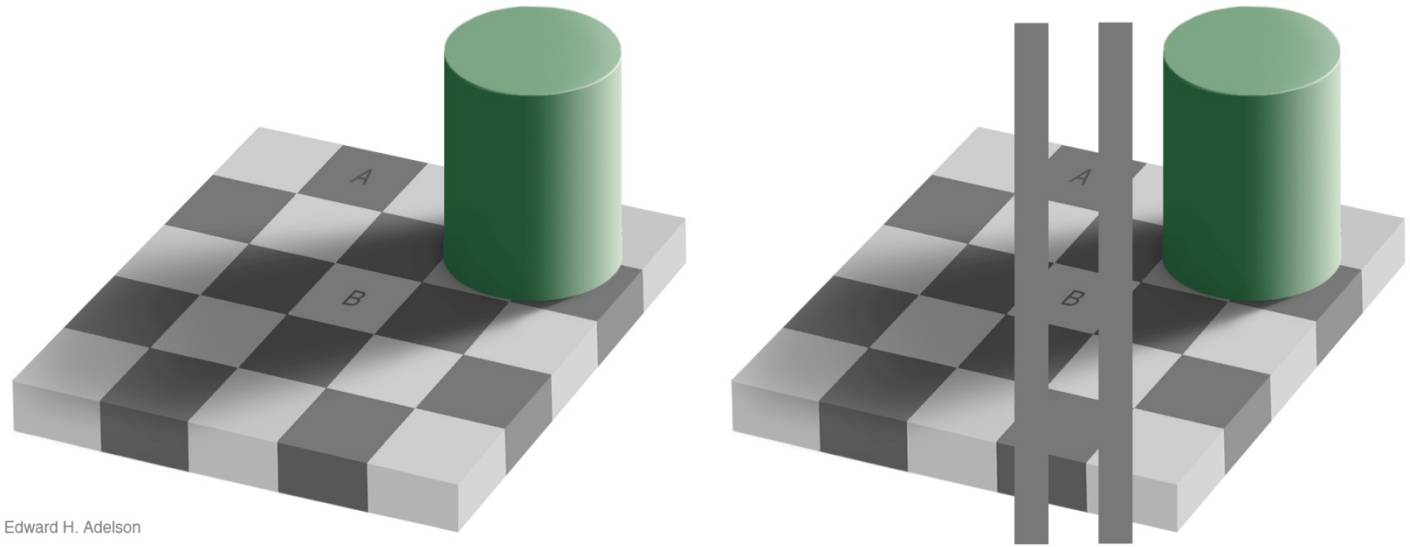
Brightness perception



Edward Adelson

http://web.mit.edu/persci/people/adelson/illusions_demos.html

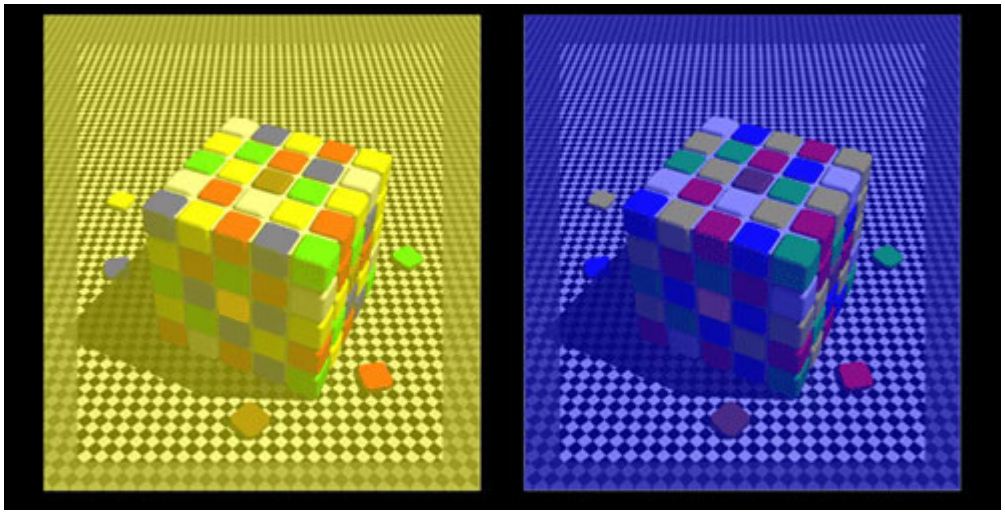
Brightness perception



Edward Adelson

http://web.mit.edu/persci/people/adelson/illusions_demos.html

Color perception

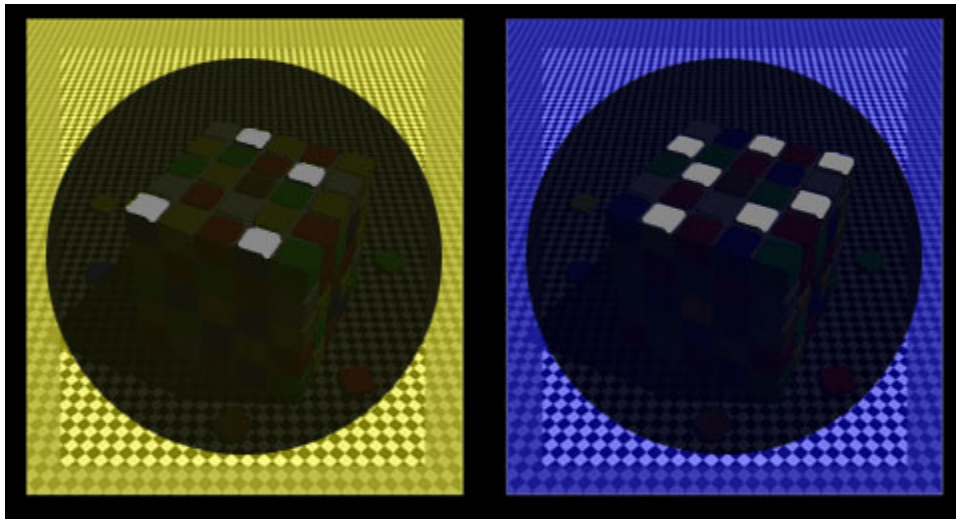


Look at blue squares

Look at yellow squares

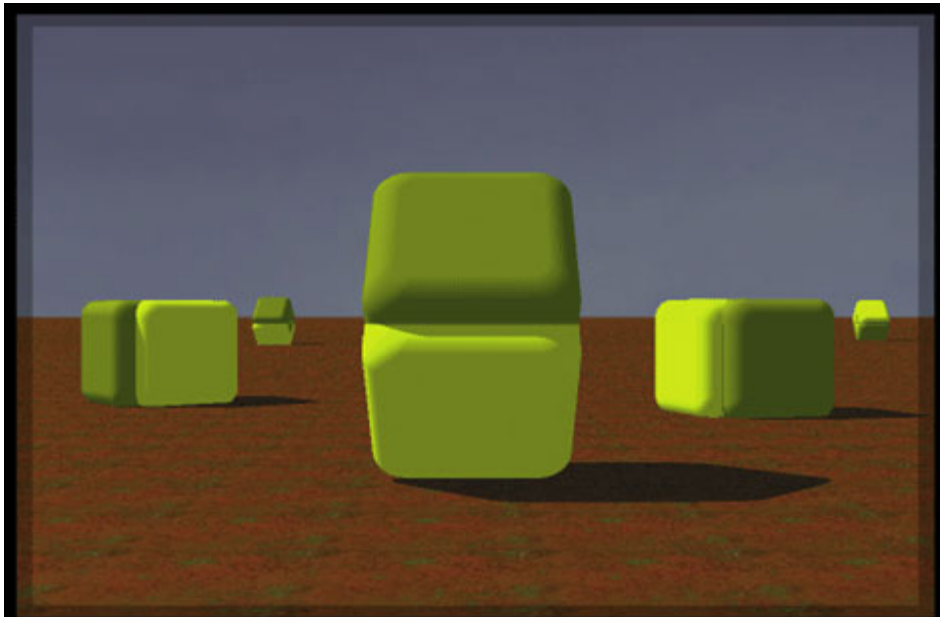
http://www.lottolab.org/articles/illusionsoflight.asp

Color perception



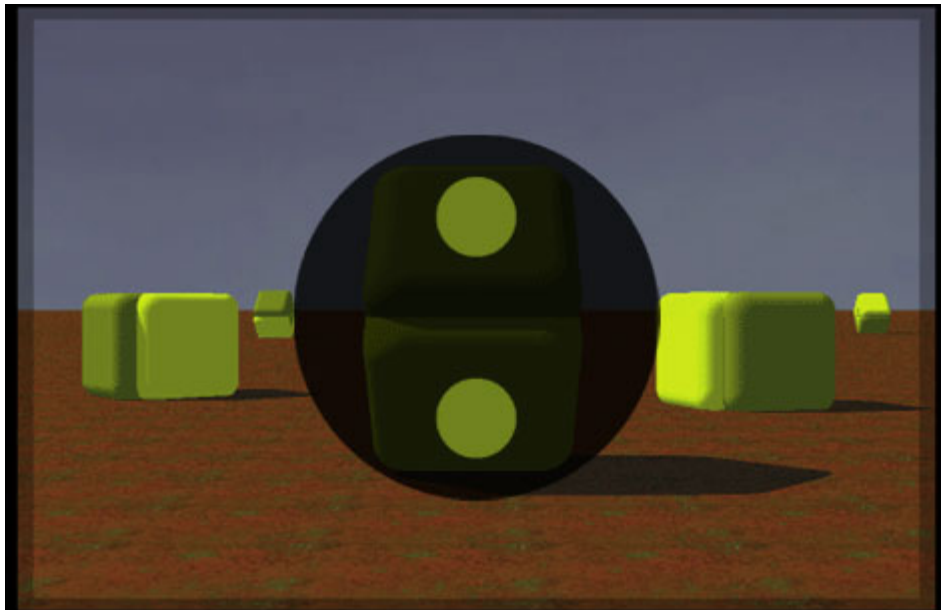
http://www.lottolab.org/articles/illusionsoflight.asp

Color perception



http://www.lottolab.org/articles/illusionsoflight.asp

Color perception



http://www.lottolab.org/articles/illusionsoflight.asp

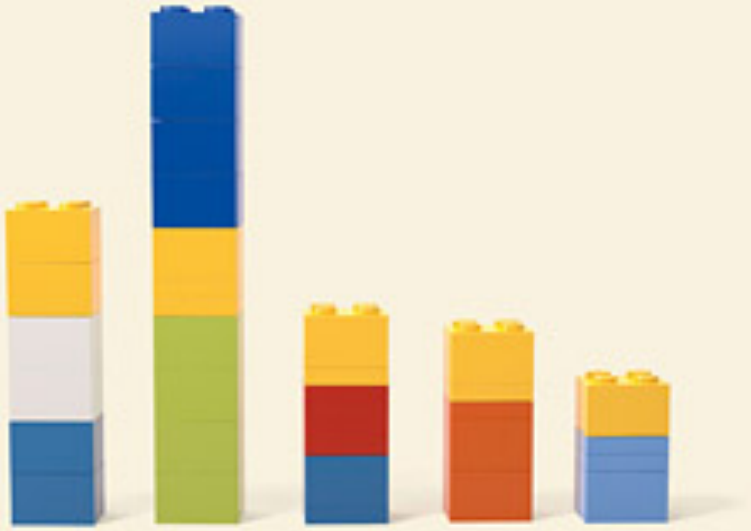
Context provides clues

- What is this?

Context provides clues

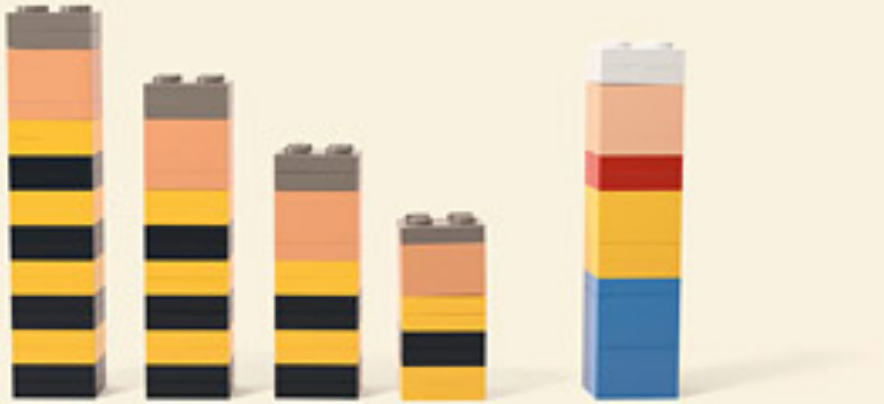
IMAGINE

- Now can you tell?



Lego Imagine Campaign,
Agency: Jung Von Matt

Context provides clues



Lego Imagine Campaign,
Agency: Jung Von Matt

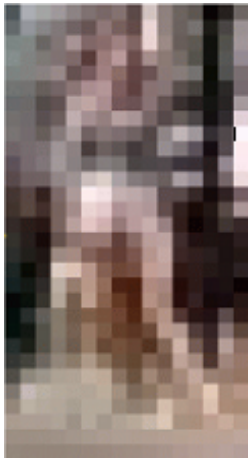
Context provides clues

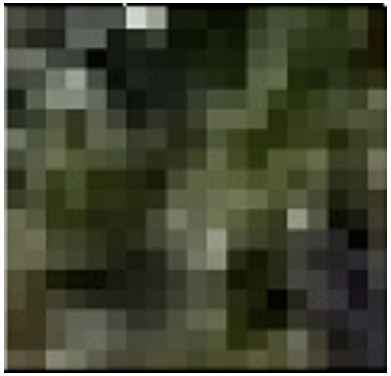
Lego Imagine Campaign,
Agency: Jung Von Matt

Context provides clues

Lego Imagine Campaign,
Agency: Jung Von Matt

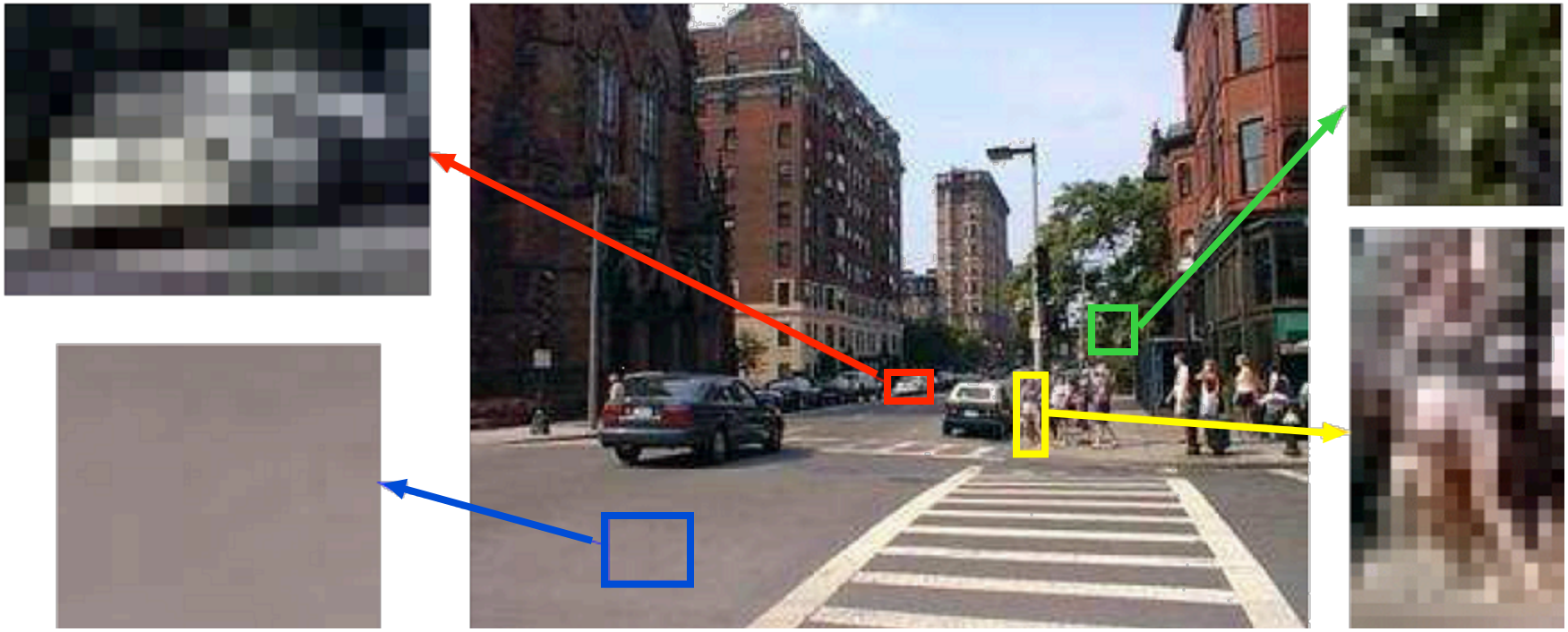
Why context is important?





Slide credit: D. Hoiem

Context resolves ambiguity



Slide credit: D. Hoiem

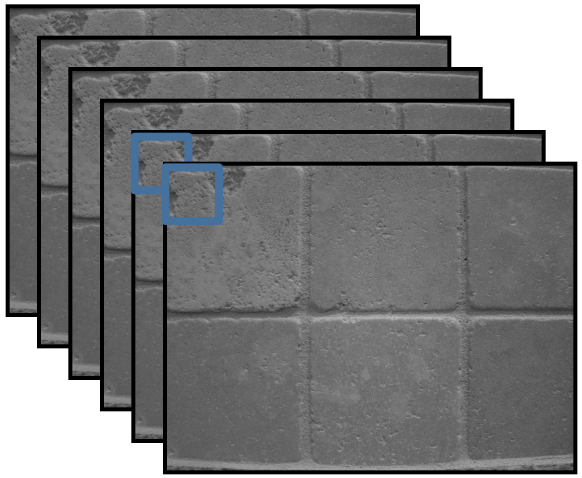
Types of context

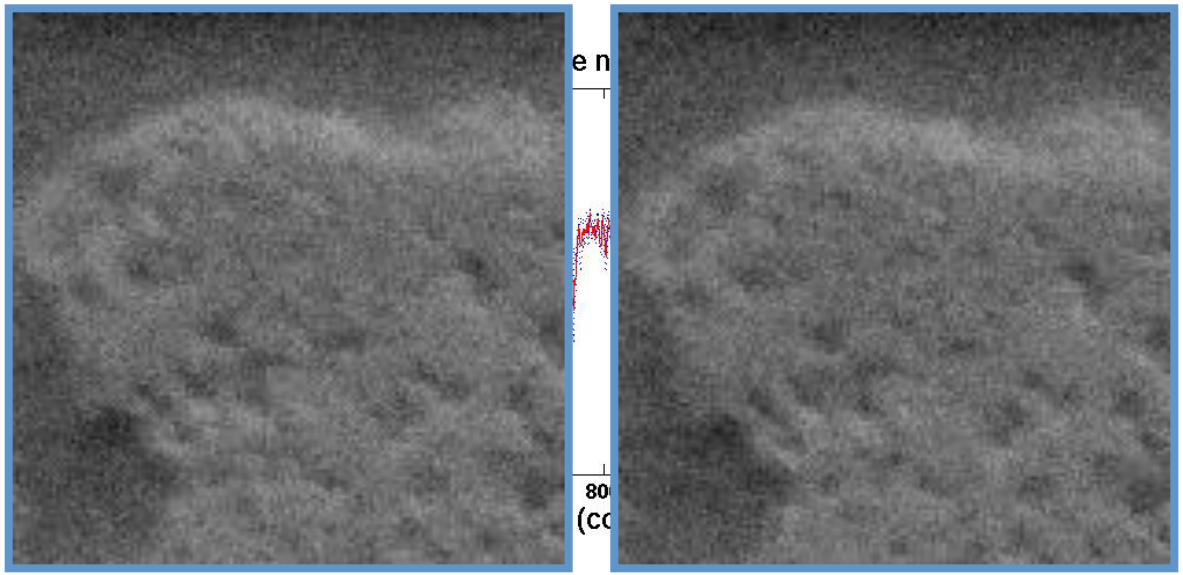
- **Local pixels**
  - window, surround, image neighborhood, object boundary/shape, global image statistics
- **2D Scene Gist**
  - global image statistics
- **3D Geometric**
  - 3D scene layout, support surface, surface orientations, occlusions, contact points, etc.
- **Semantic**
  - event/activity depicted, scene category, objects present in the scene and their spatial extents, keywords
- **Photogrammetric**
  - camera height orientation, focal length, lens distortion, radiometric, response function
- **Illumination**
  - sun direction, sky color, cloud cover, shadow contrast, etc.
- **Geographic**
  - GPS location, terrain type, land use category, elevation, population density, etc.
- **Temporal**
  - nearby frames of video, photos taken at similar times, videos of similar scenes, time of capture
- **Cultural**
  - photographer bias, dataset selection bias, visual cliches, etc.

from Divvala et al. CVPR 2009

Slide credit: D. Hoiem

Noise Reduction

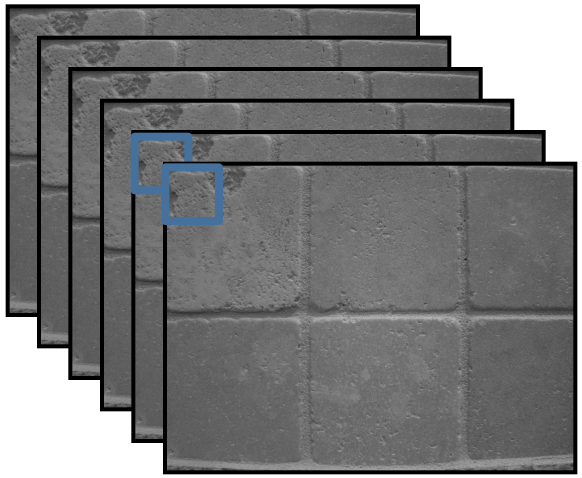


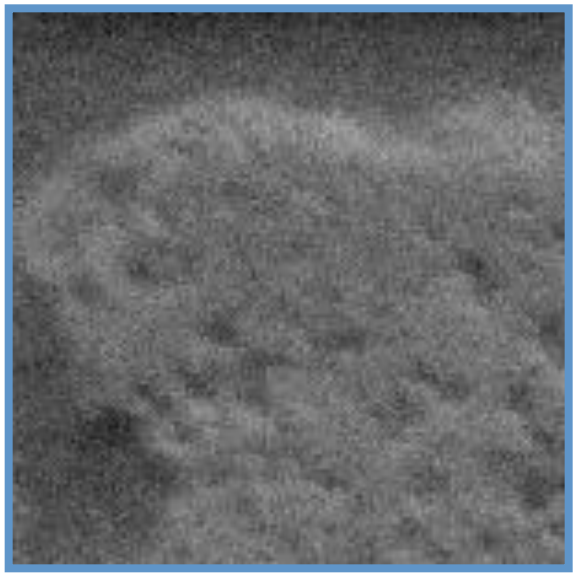


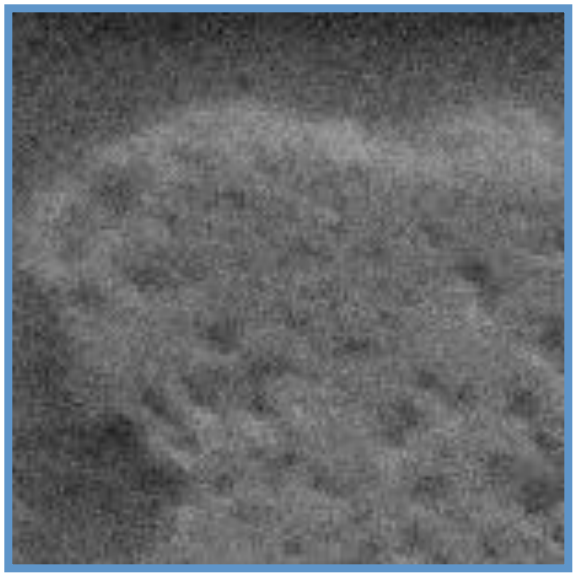
- Make multiple observations of the same <u>static</u> scene
- Take the average
- Even multiple images of the same static scene will not be identical.

Adapted from: K. Grauman

Noise Reduction





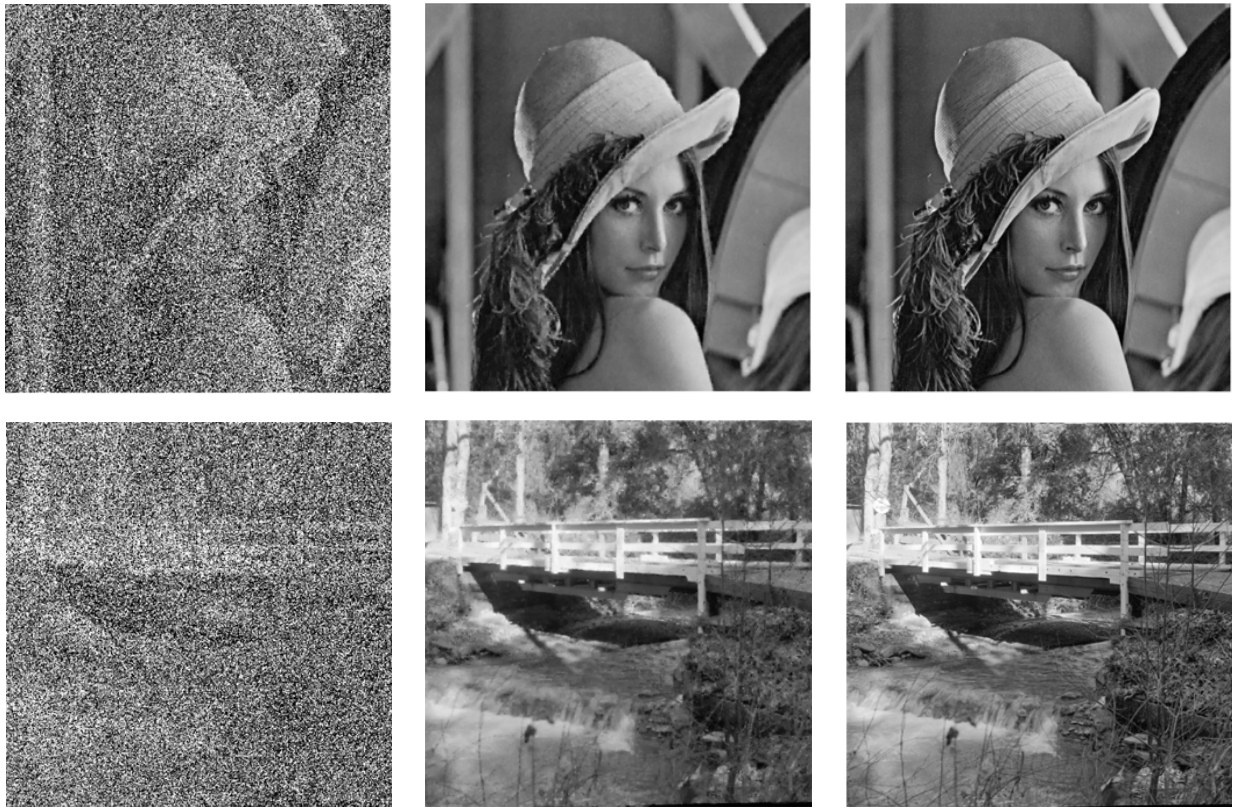


- Make multiple observations of the same static scene
- Take the average
- Even multiple images of the same static scene will not be identical.
- What if we can't make multiple observations?
  What if there's only one image?

Adapted from: K. Grauman

Image Smoothing

- Images are corrupted with 70% salt-and-pepper noise

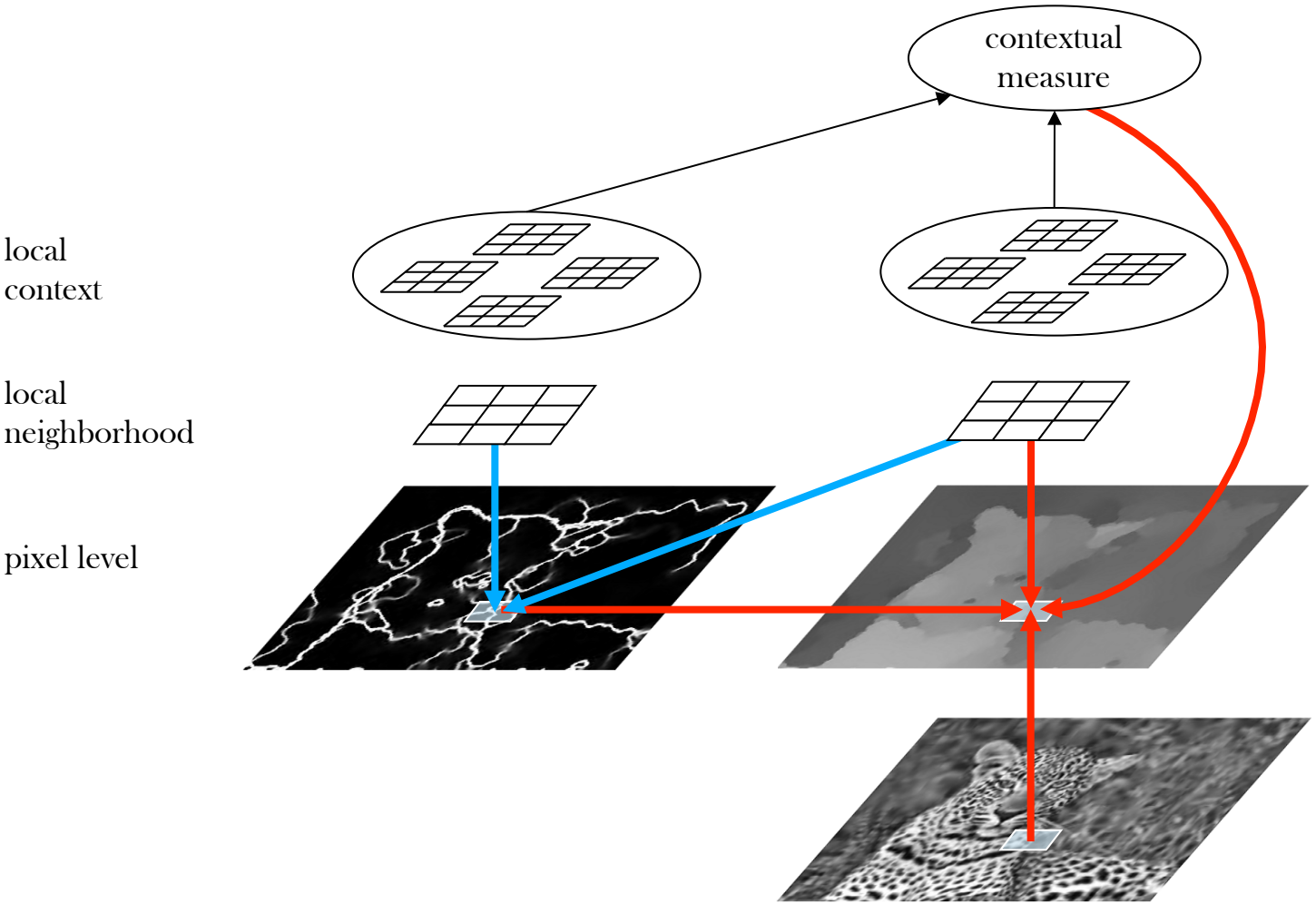


Noisy input | Recovered image | Original image

What do these examples demonstrate?

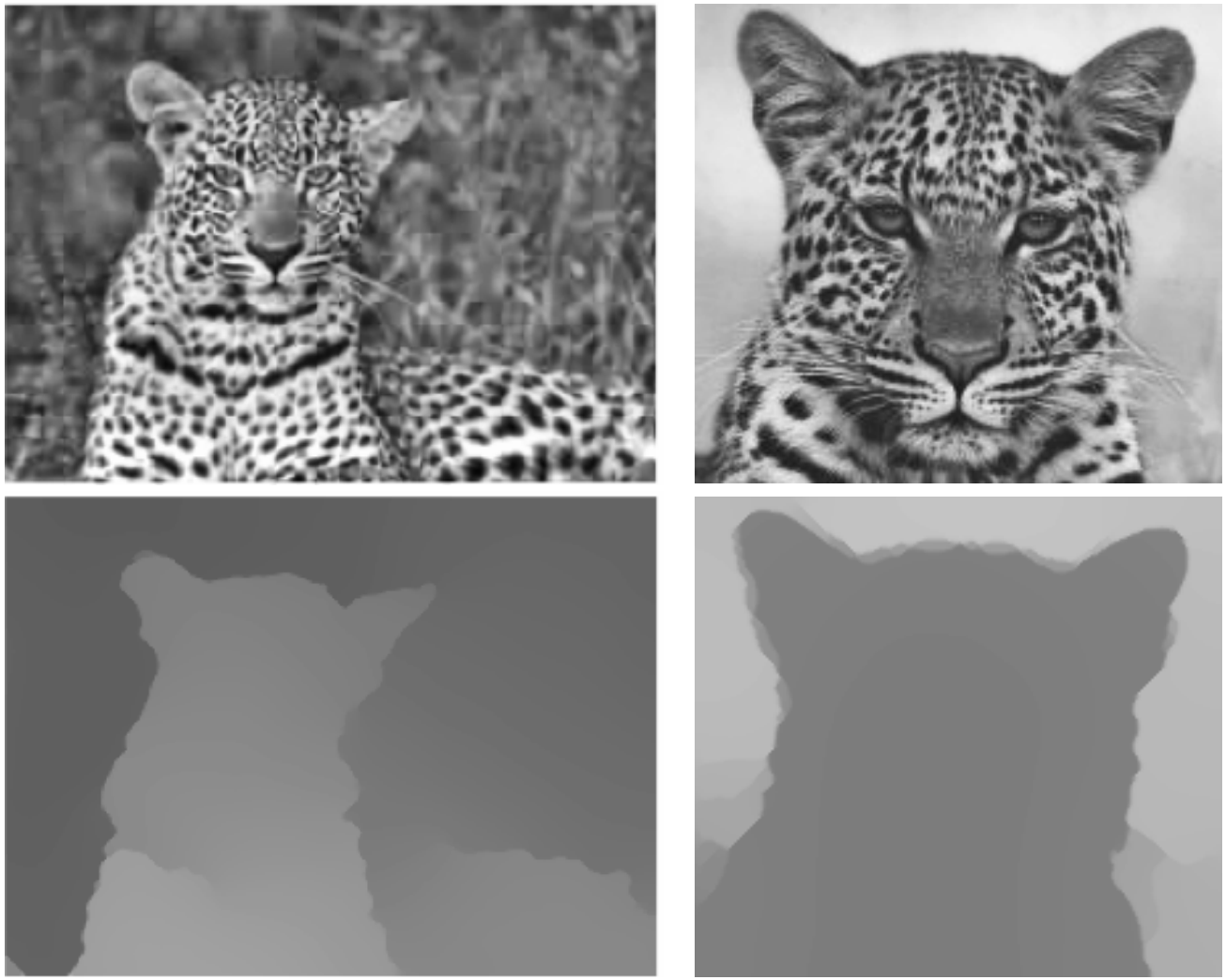
R. H. Chan et al., Salt-and-Pepper Noise Removal by Median-Type Noise Detectors and Detail-Preserving Regularization. IEEE TIP 2005

Context Guided Image Filtering



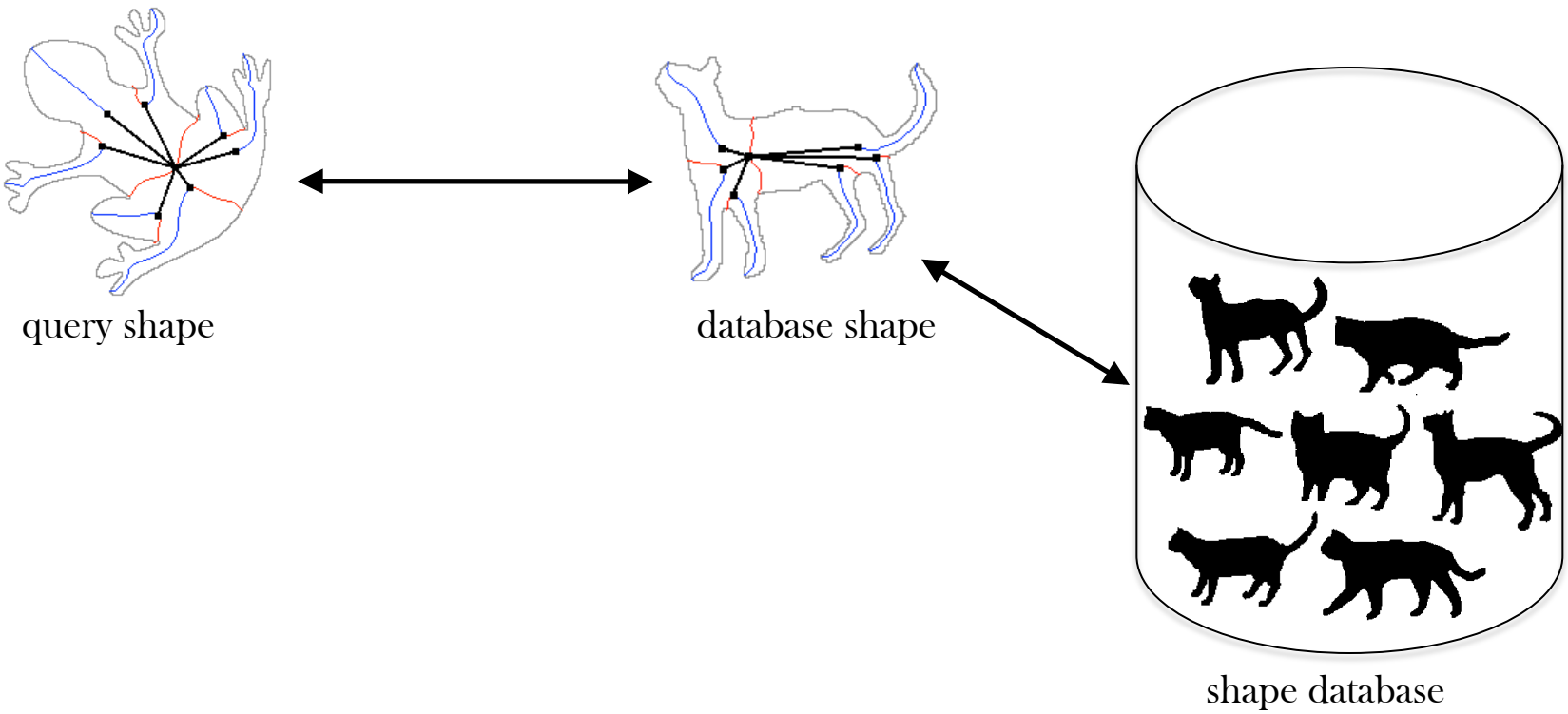
E. Erdem and S. Tari, Mumford-Shah Regularizer with Contextual Feedback, JMIV, 2009

Context Guided Image Filtering



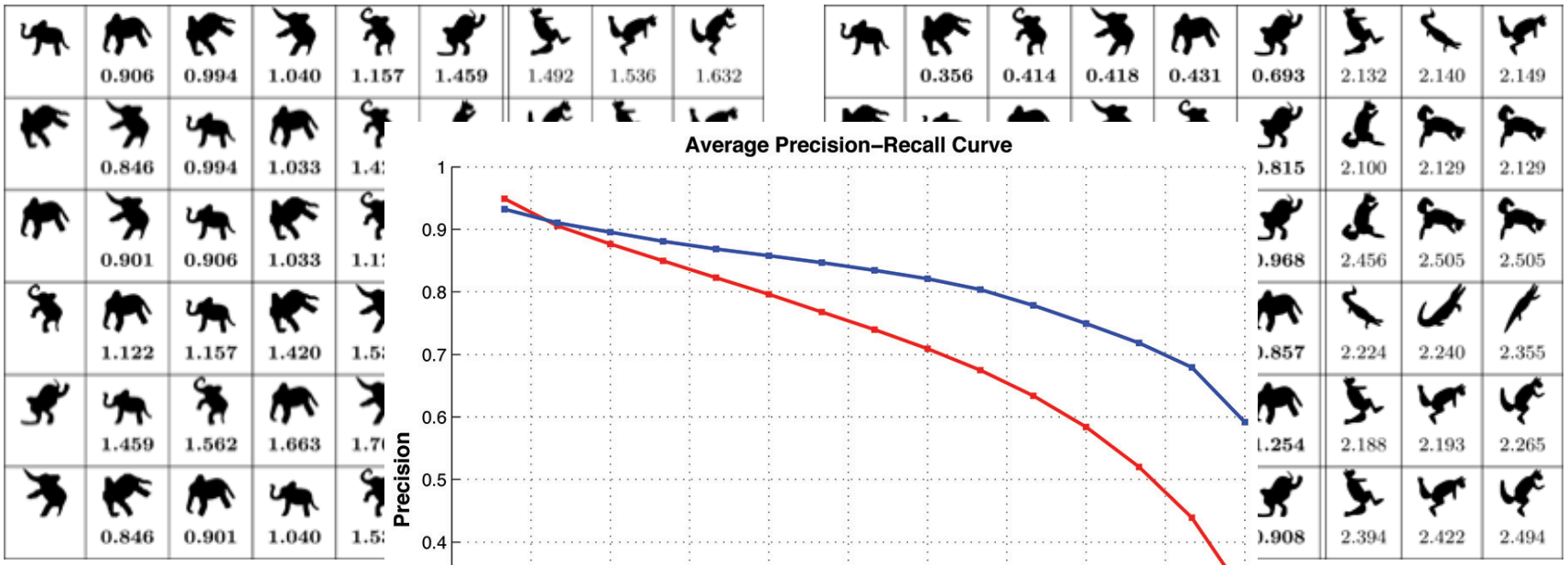
E. Erdem and S. Tari, Mumford-Shah Regularizer with Contextual Feedback, JMIV, 2009

Context-Driven Shape Matching

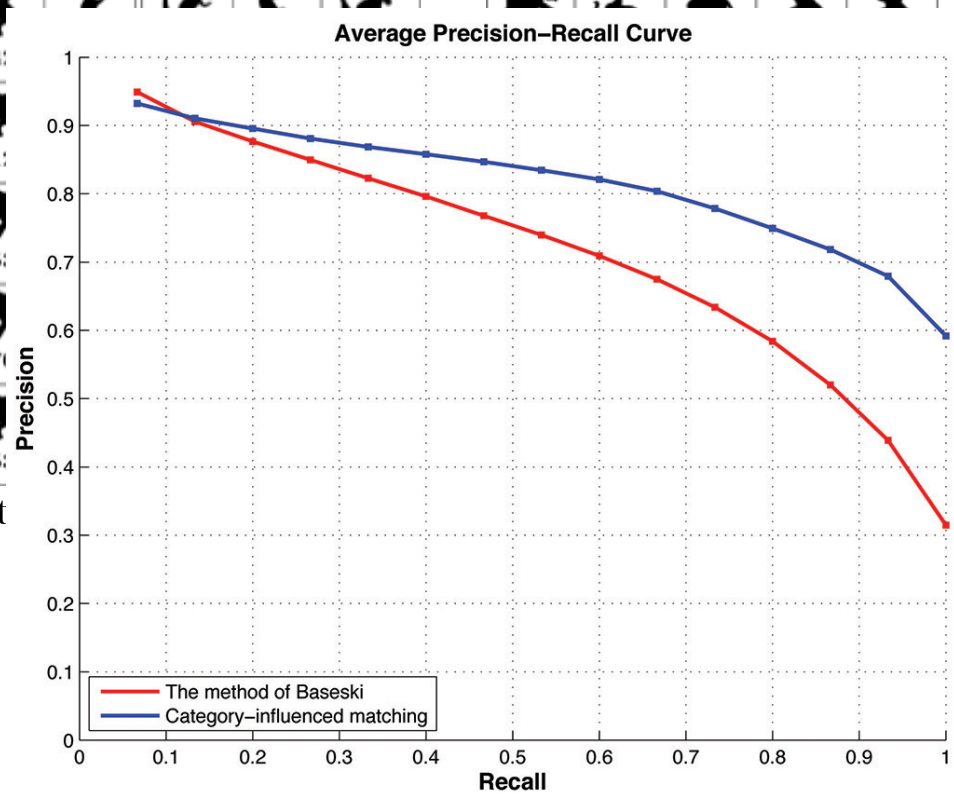


E. Baseski, A. Erdem and S. Tari, Dissimilarity Between Two Skeletal Trees in a Context, Pat. Recog., 2009

Context-Driven Shape Matching



Results without context



E. Baseski, A. Erdem and S. Tari, Dissimilarity Between Two Skeletal Trees in a Context, Pat. Recog., 2009

Cultural context

Who is Mildred? Who is Lisa?

A. Gallagher and T. Chen, Estimating Age, Gender and Identity using First Name Priors, CVPR 2008

Slide credit: D. Hoiem

Cultural context

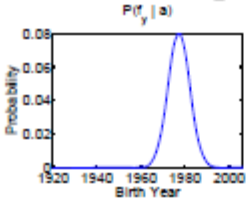
Age given Appearance

Age given Name

Mildred

$$P(f_g|f_a) = \begin{bmatrix} 0.563 \\ 0.437 \end{bmatrix}$$

$$P(f_g|n = \text{Mildred}) = \begin{bmatrix} 0.999 \\ 0.001 \end{bmatrix}$$



3.88

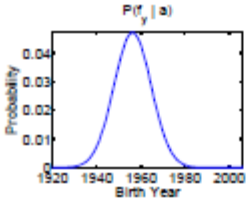
3.88

4.77

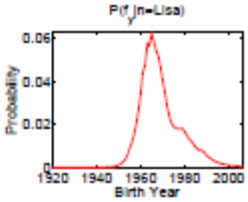
Lisa

$$P(f_g|f_a) = \begin{bmatrix} 0.687 \\ 0.313 \end{bmatrix}$$

$$P(f_g|n = \text{Lisa}) = \begin{bmatrix} 0.998 \\ 0.002 \end{bmatrix}$$



6.70

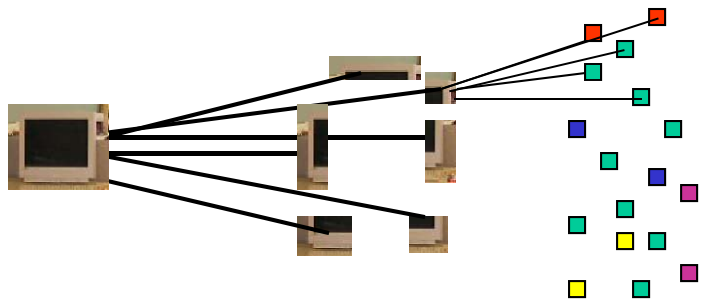


A. Gallagher and T. Chen, Estimating Age, Gender and Identity using First Name Priors, CVPR 2008

Slide credit: D. Hoiem

Object Representations

Inside the object
(intrinsic features)



Object size

Global appearance **Parts** **Pixels**

Agarwal & Roth, (02), Moghaddam, Pentland (97), Turk, Pentland (91),Vidal-Naquet, Ullman, (03)

Heisele, et al, (01), Agarwal & Roth, (02), Kremp, Geman, Amit (02), Dorko, Schmid, (03)
Fergus, Perona, Zisserman (03), Fei Fei, Fergus, Perona, (03), Schneiderman, Kanade (00), Lowe (99)
Etc.

Slide credit: A. Torralba

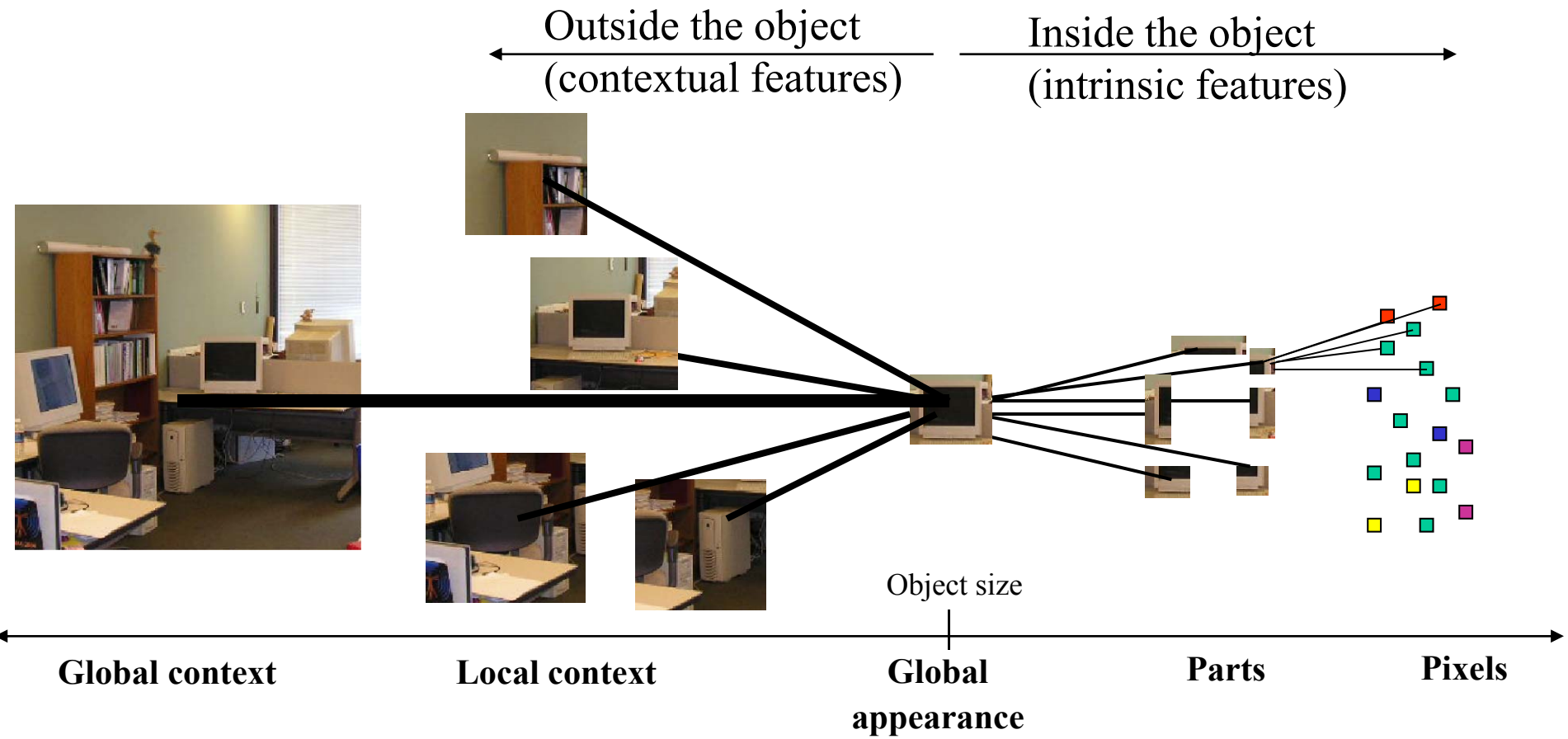
What are the hidden objects?

Slide credit: A. Torralba

What are the hidden objects?

Slide credit: A. Torralba

Object Representations



Kruppa & Shiele, (03), Fink & Perona (03)

Carbonetto, Freitas, Barnard (03), Kumar, Hebert, (03)

He, Zemel, Carreira-Perpinan (04), Moore, Essa, Monson, Hayes (99)

Strat & Fischler (91), Murphy, Torralba & Freeman (03)

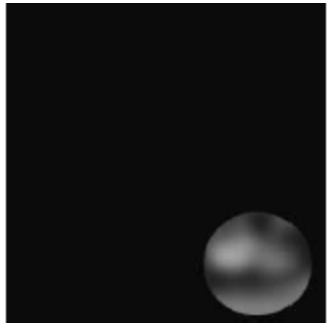
Agarwal & Roth, (02), Moghaddam, Pentland (97), Turk, Pentland (91),Vidal-Naquet, Ullman, (03)

Heisele, et al, (01), Agarwal & Roth, (02), Kremp, Geman, Amit (02), Dorko, Schmid, (03)
Fergus, Perona, Zisserman (03), Fei Fei, Fergus, Perona, (03), Schneiderman, Kanade (00), Lowe (99)
Etc.

Slide credit: A. Torralba

Sometimes context is the major component of recognition

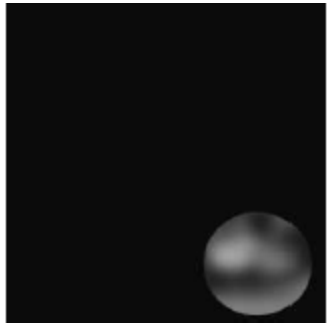
- What is this?



Slide credit: D. Hoiem

Sometimes context is the major component of recognition

- What is this?



- Now can you tell?

Slide credit: D. Hoiem

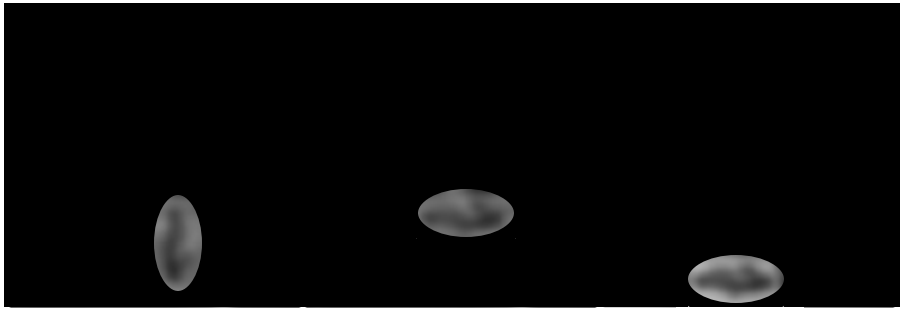
More Low-Res

- What are these blobs?

Slide credit: D. Hoiem

More Low-Res

- The same pixels! (a car)



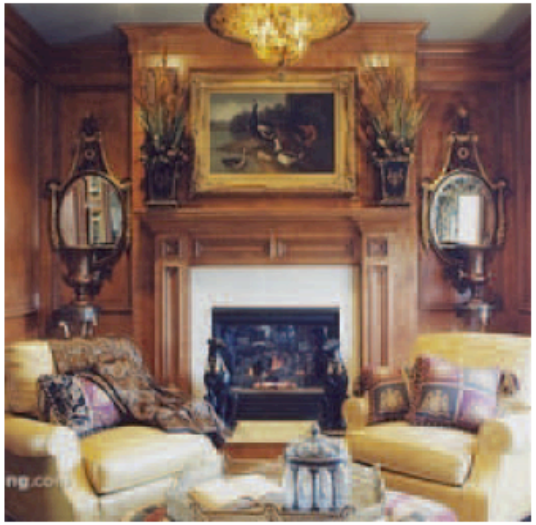
Slide credit: D. Hoiem

Global Context

- The gist of the scene
- In a glance, we remember the meaning of an image and its global layout but some objects and details are forgotten

Slide credit: A. Torralba

Which are the important elements?



Different content (i.e. objects), different spatial layout

Slide credit: A. Torralba

Which are the important elements?

Similar objects, and similar spatial layout

Different lighting, different materials, different "stuff"

Slide credit: A. Torralba

Global Context

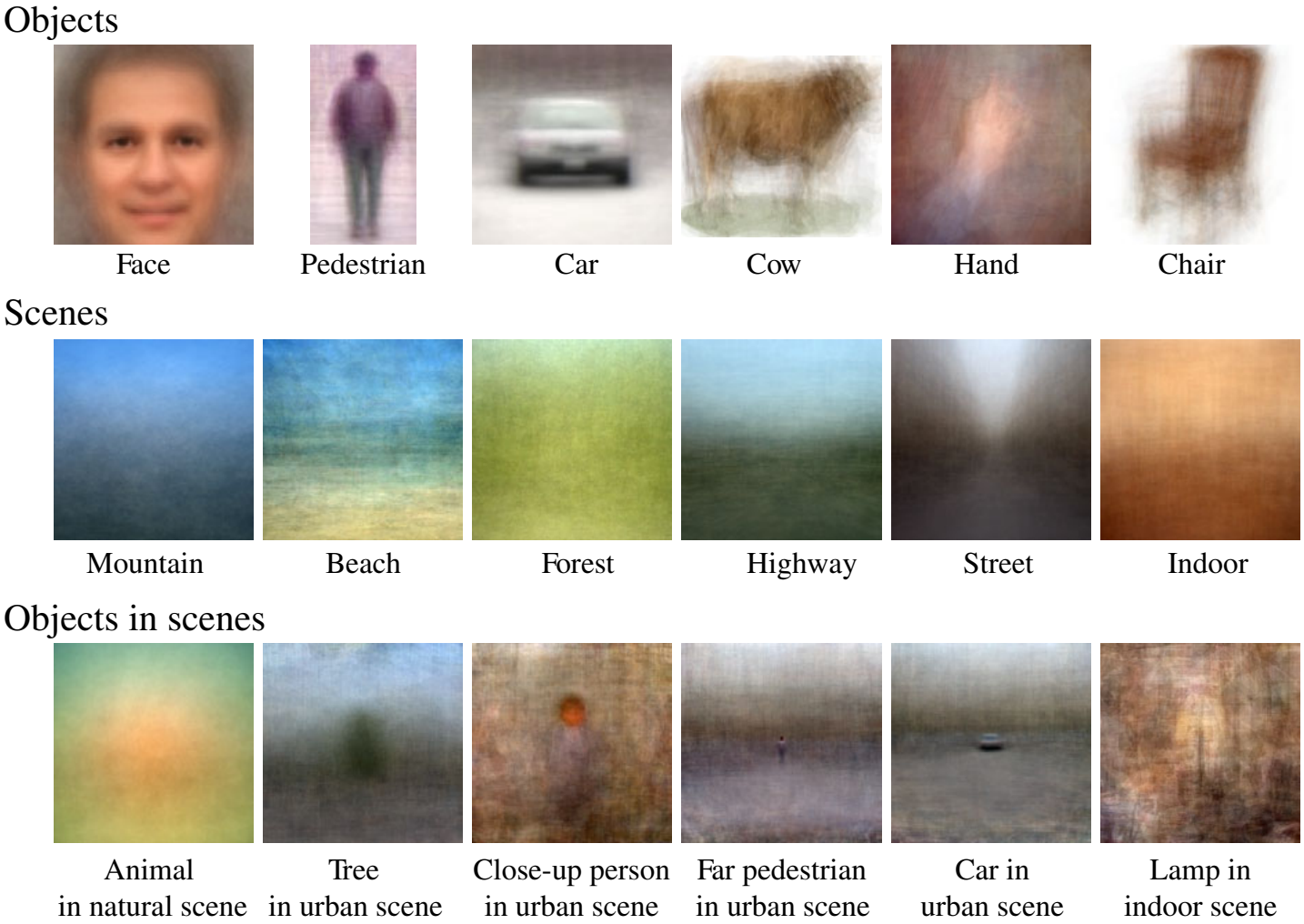


Figure 1. Averaged pictures of categories of objects, scenes and objects in scenes, computed with 100 exemplars or more per category. Exemplars were chosen to have the same basic level and viewpoint in regard to an observer. The group objects in scenes (third row) represent examples of the averaged peripheral information around an object centred in the image.

Oliva & Torralba 2003

Global Context

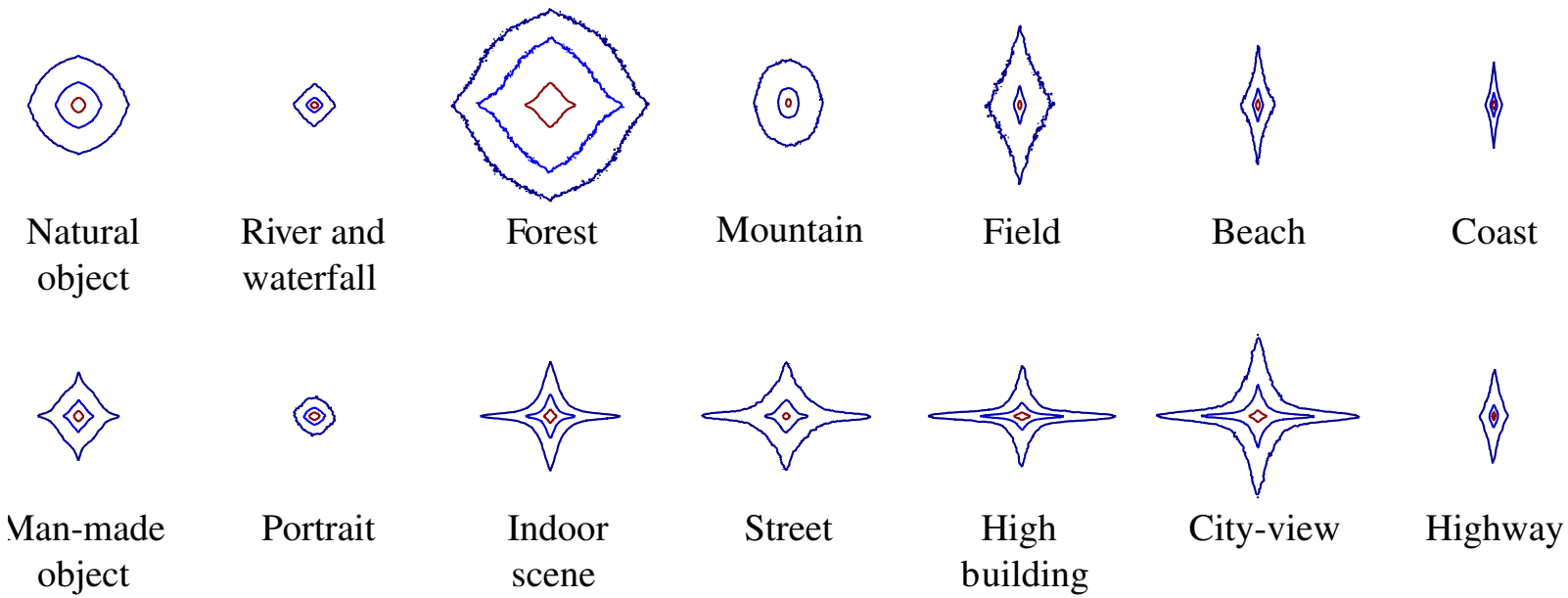
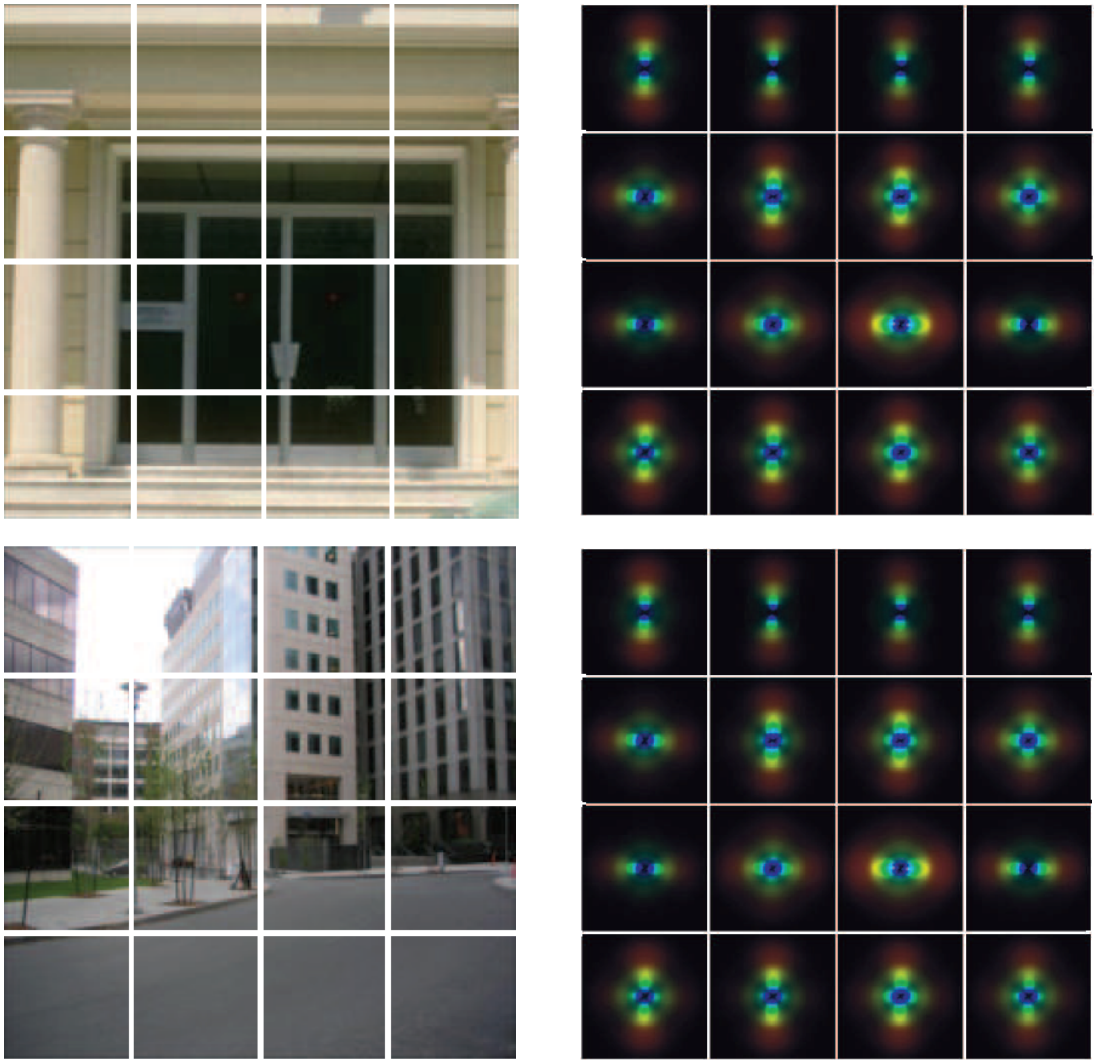


Figure 3. Spectral signatures of 14 different image categories. Each spectral signature is obtained by averaging the power spectra of a few hundred images per category. The contour plots represent 60, 80 and 90% of the energy of the spectral signatures (energy is obtained by adding the square of the Fourier components). The size of the spectral signature is correlated with the slope (α). A large value of α produces a fast decay of the energy at high spatial frequencies, which produces a smaller contour. The overall shape is a function of both $\alpha(\theta)$ and $A(\theta)$.

Oliva & Torralba 2003

Gist descriptor

Oliva and Torralba, 2001



- Apply oriented Gabor filters over different scales
- Average filter energy in each bin

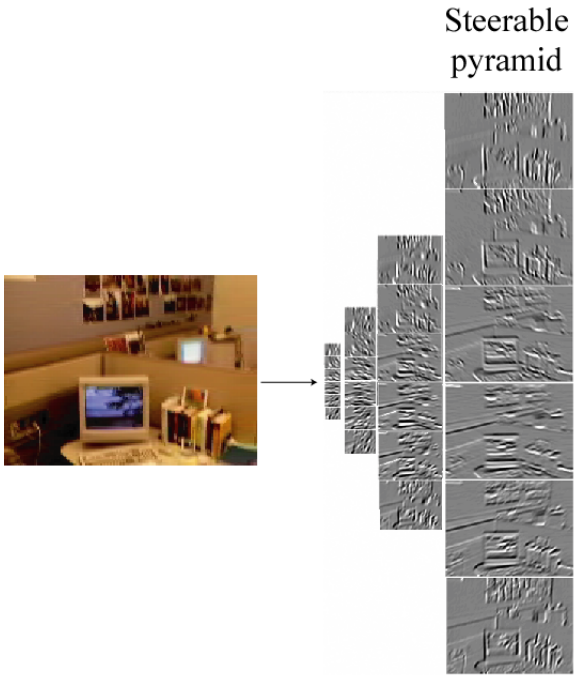
8	orientations
4	scales
x 16	bins
512	dimensions

Similar to SIFT (Lowe 1999) applied to the entire image

M. Gorkani, R. Picard, ICPR 1994; Walker, Malik. Vision Research 2004; Vogel et al. 2004; Fei-Fei and Perona, CVPR 2005; S. Lazebnik, et al, CVPR 2006; ...

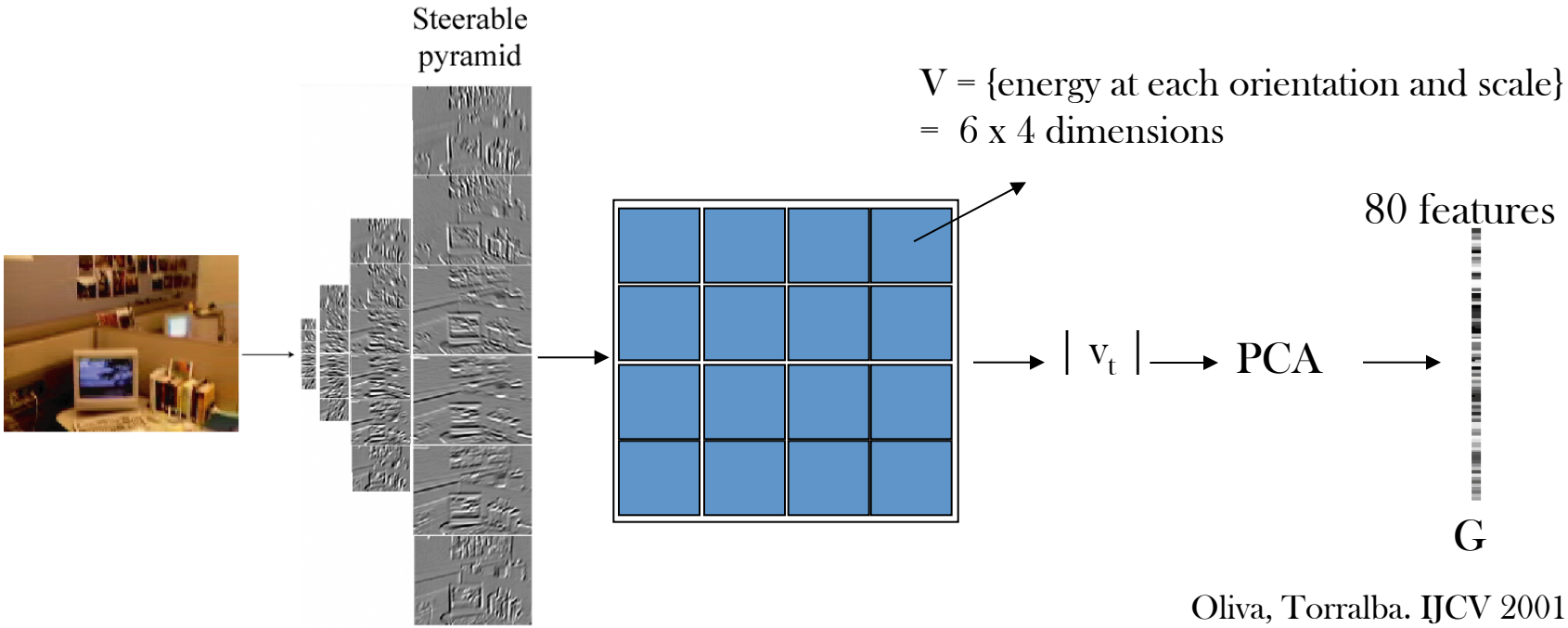
Slide credit: A. Torralba

Gist descriptor



Slide credit: A. Torralba

Gist descriptor



Oliva, Torralba. IJCV 2001

Slide credit: A. Torralba

Global Context

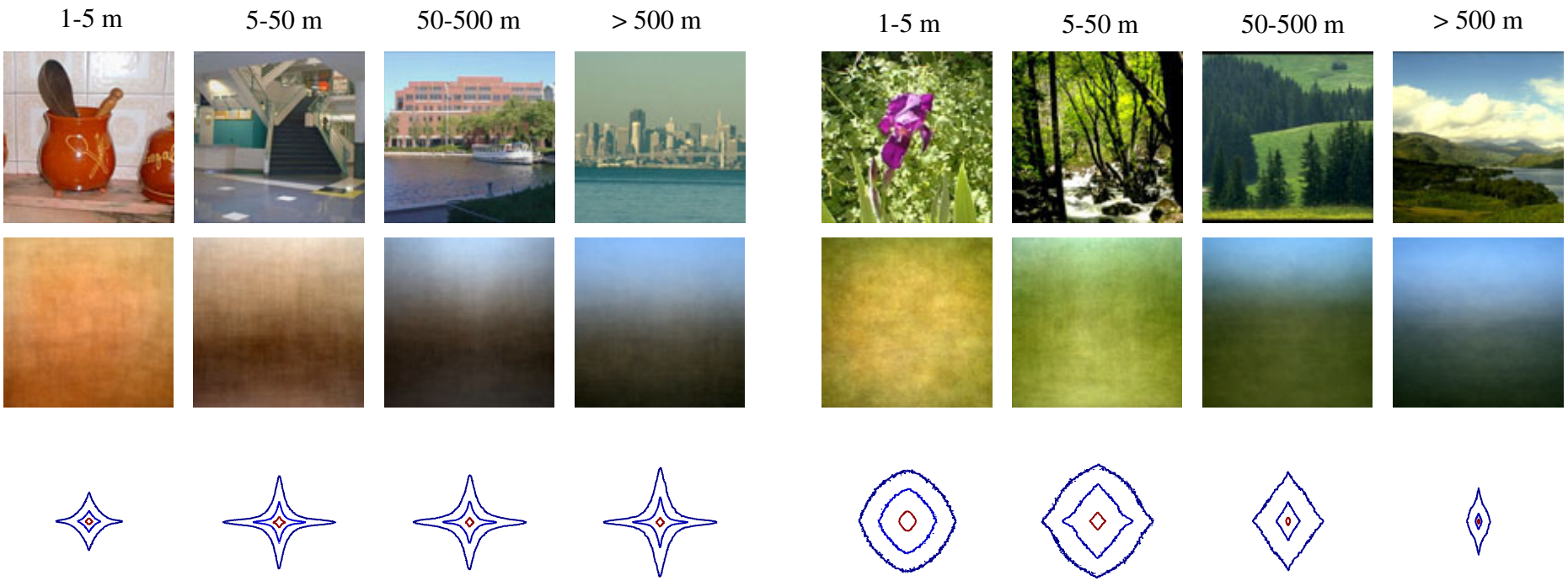
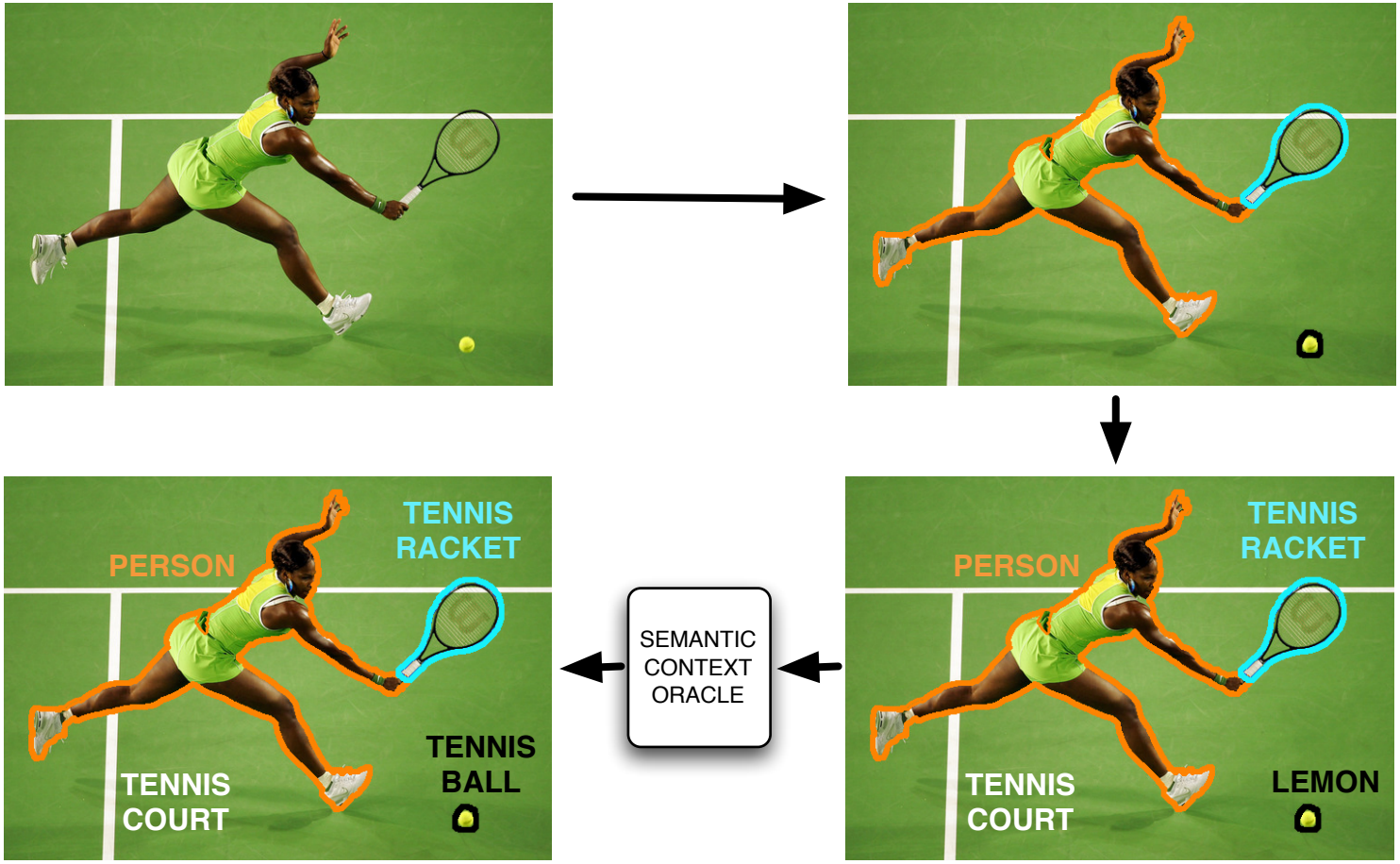


Figure 4. Averaged spatial images and spectral signatures as a function of scene scale. Scene scale refers to the mean distance between the observer and the principal elements that compose the scene. Each image average and spectral signature was calculated with 300–400 images.

Figure 15. Examples of images and the selected regions that are expected to contain faces based on contextual features. The regions are selected according to global image statistics and not to the actual presence of the object of interest.

Oliva & Torralba 2003

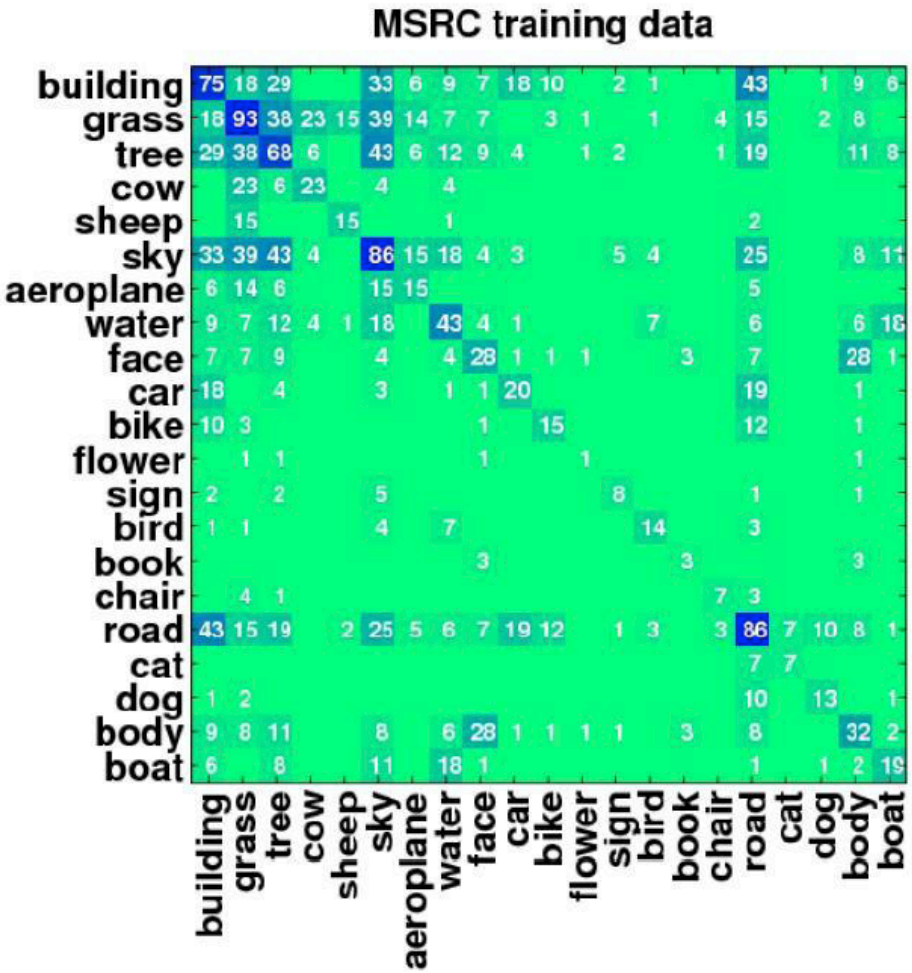
Semantic Context



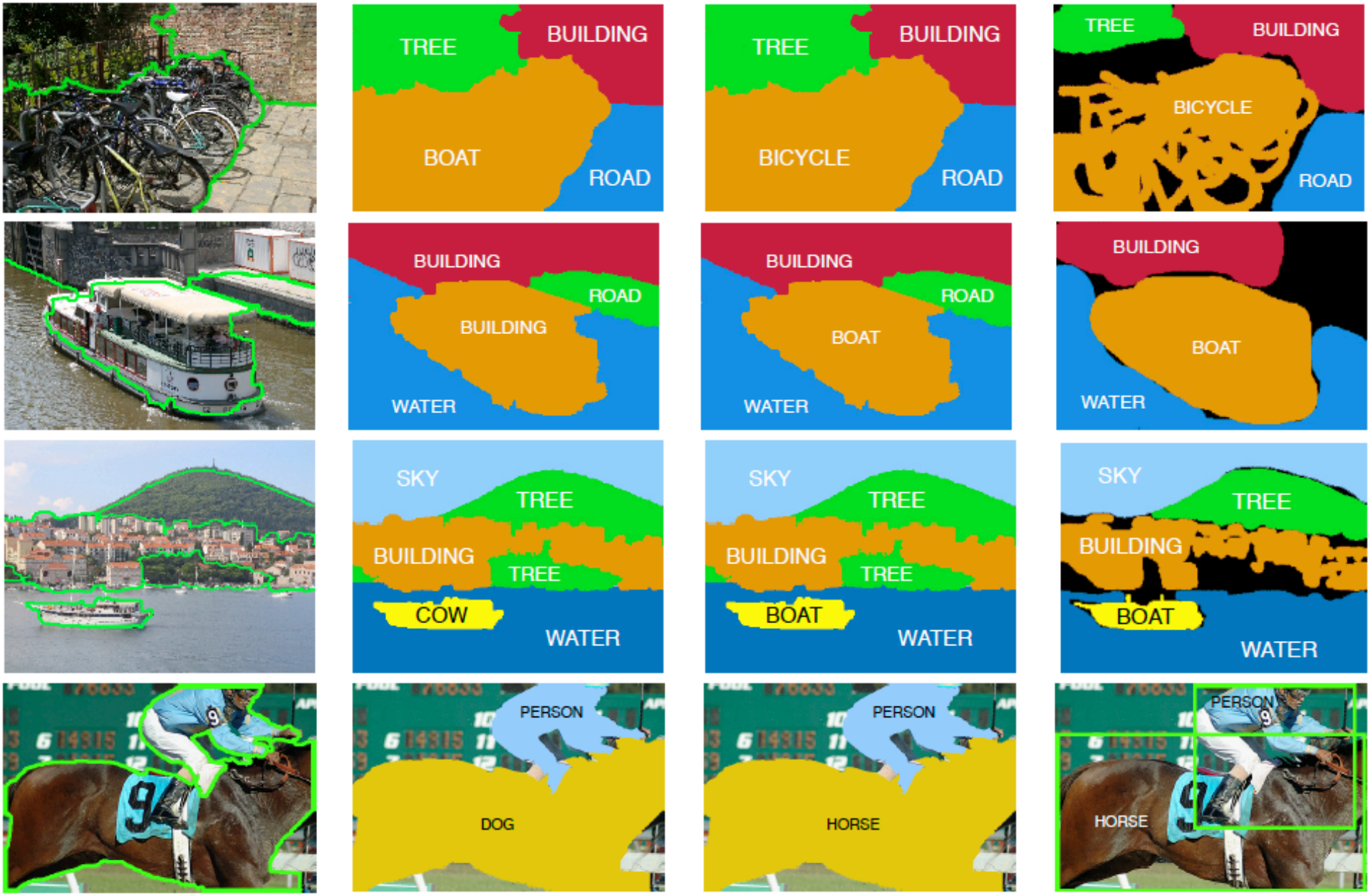
A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora and S. Belongie. Objects in Context. ICCV 2007

Semantic Context

$F(c_i=m_i, c_j=m_j)$ = co-ocurrence matrix on training set (count how many times two objects appear together).



A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora and S. Belongie. Objects in Context. ICCV 2007



A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora and S. Belongie. Objects in Context. ICCV 2007

Geographic Context

Where in the world?

J. Hays and A. Efros, im2gps: Estimating Geographic Information from a Single Image. CVPR 2008.

Slide credit: J. Hays

Scene Matches

J. Hays and A. Efros, im2gps: Estimating Geographic Information from a Single Image. CVPR 2008.

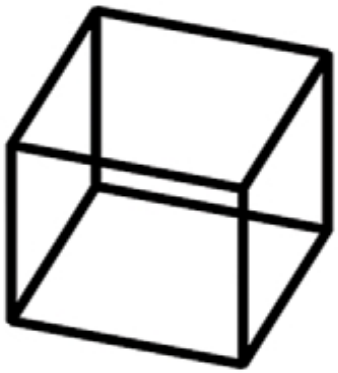
Slide credit: J. Hays

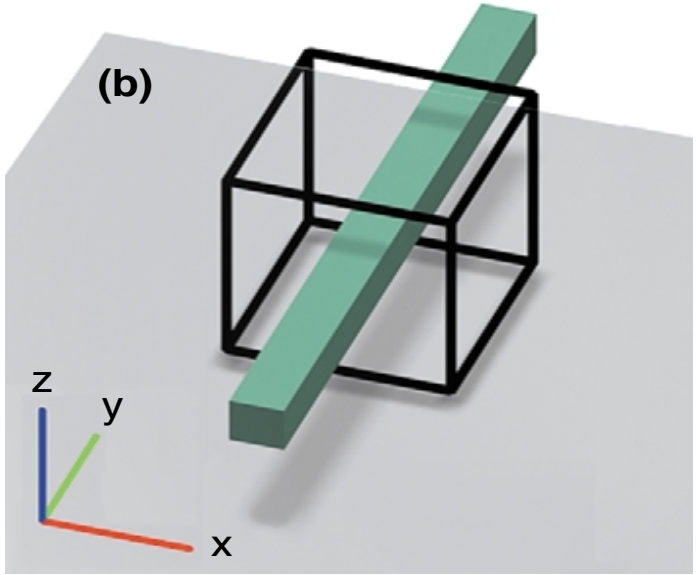
J. Hays and A. Efros, im2gps: Estimating Geographic Information from a Single Image. CVPR 2008.

Slide credit: J. Hays

Crossmodal Context

- Crossmodal interactions to disambiguate unreliable sensory input



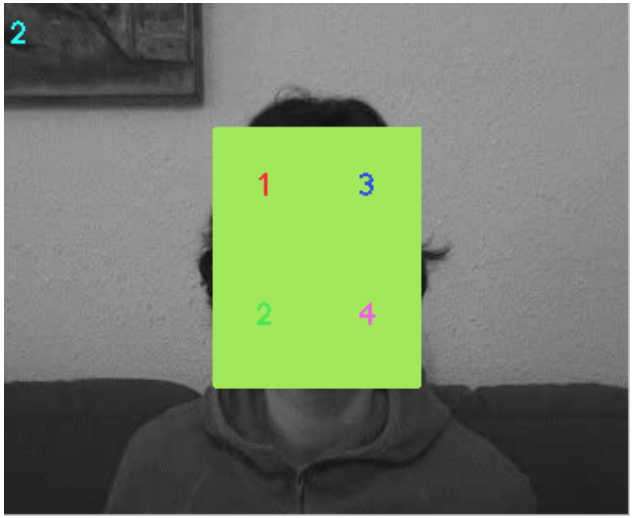


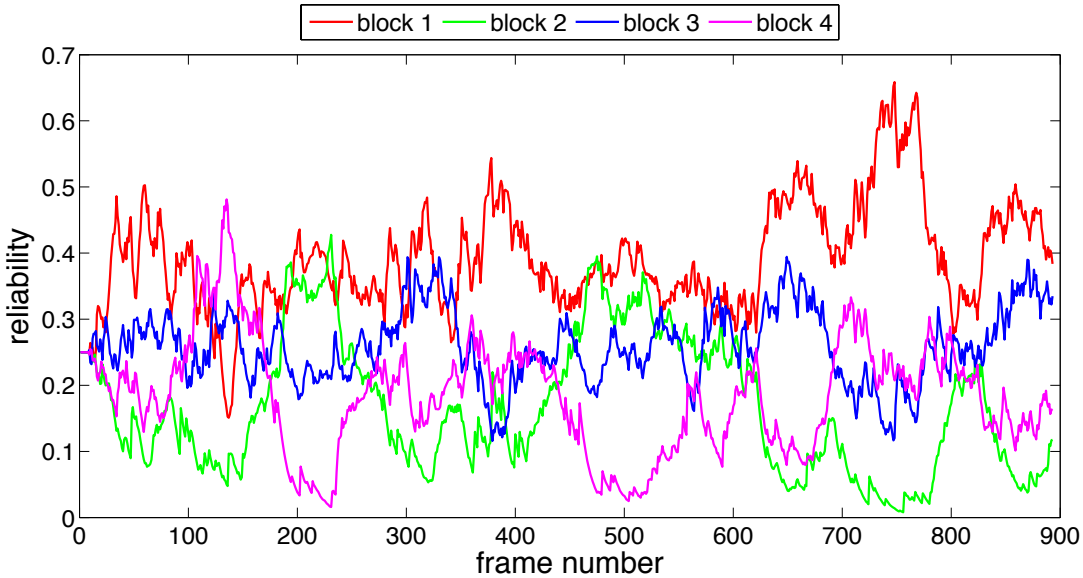
introduction of an occlusion cue and a shadow resolves the ambiguity

Necker cube images taken from Ernst and Bülthoff, Trends Cog. Sci., 2004

Crossmodal Context in Tracking

- Template patch is divided into four fragments.
- Reliability of a fragment increases when the fragment is visible, and decreases when it is hidden.





E. Erdem *et al.*, Fragments Based Tracking with Adaptive Cue Integration, CVIU 2012

Crossmodal Context in Tracking

- Woman goes behind cars and becomes occluded.
- Her pose undergoes some changes along the sequence.
- Tracking is successful even only few fragments have a good reliability.

E. Erdem *et al.*, Fragments Based Tracking with Adaptive Cue Integration, CVIU 2012