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)
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

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

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

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

Look at blue squares

Look at yellow squares

Content © 2008 R. Beau Lotto
http://www.lottolab.org/articles/illusionsoflight.asp
Color perception
Color perception
Color perception
Context provides clues

• What is this?
Context provides clues

• 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?
Context resolves ambiguity
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.
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.

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

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

Context-Driven Shape Matching

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

A. Gallagher and T. Chen, Estimating Age, Gender and Identity using First Name Priors, CVPR 2008
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.
What are the hidden objects?
What are the hidden objects?
Object Representations

Outside the object (contextual features)  
Inside the object (intrinsic features)

Global context  
Global appearance  
Local context  
Parts  
Pixels

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.
Sometimes context is the major component of recognition

• What is this?
Sometimes context is the major component of recognition

• What is this?

• Now can you tell?
More Low-Res

- What are these blobs?
More Low-Res

• The same pixels! (a car)
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
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
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.
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 ($\alpha$). A large value of $\alpha$ 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

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


Slide credit: A. Torralba
Gist descriptor

Steerable pyramid

Slide credit: A. Torralba
Gist descriptor

\[ V = \{ \text{energy at each orientation and scale} \} = 6 \times 4 \text{ dimensions} \]

Oliva, Torralba. *IJCV* 2001

Slide credit: A. Torralba
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.
Abstract
In the task of visual object categorization, semantic context can play a very important role in reducing ambiguity in objects' visual appearance. In this work, we propose incorporating semantic object context as a post-processing step into any off-the-shelf object categorization model. Using a conditional random field (CRF) framework, our approach maximizes object label agreement according to contextual relevance. We compare two sources of context: one learned from training data and another queried from Google Sets. The overall performance of the proposed framework is evaluated on the PASCAL and MSRC datasets. Our findings conclude that incorporating context into object categorization greatly improves categorization accuracy.

1. Introduction
Object categorization has been an active topic of research in psychology and computer vision for decades. Initially, vision scientists and psychologists formulated hypotheses about models of object categorization and recognition [7, 8, 24]. Subsequently, in the past 10 years or so, object recognition and categorization have become very popular areas of research in computer vision. With two general models emerging, generative and discriminative, the newly developed algorithms aim to adhere to the original modeling constraints proposed by vision scientists. For example, the hypothesis put forth by Biederman et al. [1] suggests five classes of relations between an object and its setting that can characterize the organization of objects into real-world scenes. These are: (i) interposition (objects interrupt their background), (ii) support (objects tend to rest on surfaces), (iii) probability (objects tend to be found in some contexts but not others), (iv) position (given an object is probable in a scene, it often is found in some positions and not others), and (v) familiar size (objects have a limited set of size relations with other objects). Classes (i, ii, iv, and v) have been addressed fairly well in the models proposed by the computer vision community [2, 5, 23]. Class (iii), referring to the contextual interactions between objects in the scene, however, has received comparatively little attention.

Existing context-based methods for object recognition and classification consider global image features to be the source of context, thus trying to capture object class specific features. In [9, 14, 25, 28], the relationship between context and object properties is based on the correlation between the statistics of low-level features across the image that contains the object, or even the whole object category. Semantic context among objects has not been explicitly incorporated into existing object categorization models. Semantic context requires access to the referential meaning of the object [1]. In other words, when performing the task of object categorization, objects' category labels must be assigned with respect to other objects in the scene, assuming there is more than one object present. To illustrate this

For simplicity we will use context and semantic context interchangeably from now on.

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

Geographic Context

Where in the world?

Scene Matches

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