CSE597E : Visual Salience and Object Segmentation

Today

- Logistics (schedule; what is required)
- Overview of topics
- Next week’s readings
Seminar Goals

This seminar will explore cues that might be used (by people and/or machines) to parse the visual scene into “objects”.

- explore concepts: salience, attention, object segmentation, recognition
- become more familiar with literature on psychology of vision (special emphasis on computational/statistical models)
- review relevant work in vision literature
- can psychological theory and vision practice be combined?

My bias: Figuring out how people do it is not the goal. I want to develop computer vision algorithms to do it. But... so far, we only know that people can do it and computers can’t (yet). Therefore, it makes sense to at least consider what cues/methods people might be using.
Course Logistics

Ideally, we will meet once per week, for a three-hour block of time. Tues and Thurs are not good (Bob teaches in morning / Yanxi in afternoon). How is Wednesday 1-4? 11:15-2:15?????????? see vicki
Course Logistics

The format of this course is a “reading group.” Students will be responsible for reading each week’s papers and being prepared to discuss the material. Each paper will have one student (or two students if it is a really long paper) assigned to do an initial summary and then lead the discussion.

Important: this course will succeed or fail based on our ability to get interesting discussions going. Everyone has to come prepared and to participate.
Initial Topic Ideas

Models of Visual Attention (in psych and computer vis)

From Features to Objects (semantic gap; “binding”)

Perceptual Organization Cues (including symmetry!)

Visual Search (relevance to tracking)

Image Statistics (e.g. natural image statistics and work on “gists”)
What is Attention?

William James, Principles of Psychology, 1890

“Everyone knows what attention is.”

Actually, he went on to elaborate a bit...

“Everyone knows what attention is. It is the taking possession by the mind in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought...It implies withdrawal from some things in order to deal effectively with others."
Role of Attention

- computational efficiency
- enhances signal to noise ratio
- facilitates processing of nearby/relevant features
- binds features into integrated object representations (Treisman)
- prevents illusory conjunctions of features

example: red circles and green squares...
Attention Metaphors

Location-based (Spatial) vs Object-based

- Spotlight: location-based
- Zoom lens: modified spotlight, location based
- Object-based: sticks to object.
Some Attention Models

Treisman and Gelade, Feature-Integration Theory (FIT)
Wolfe, Guided-Search [builds on FIT]
Itti and Koch, Saliency Maps and Winner Take All
Moser&Sitton, Saliency Maps via Neural Nets
Bundesen, Theory of Visual Attention (TVA)
Logan, CODE Theory of Visual Attention (CTVA) [builds on TVA]
Attention Models
Common Theme...

FIGURE 1. An attentional saliency map constructed from bottom-up and top-down information
Itti and Koch Salience Maps

Figure 3.10: Comparison of the saliency map to the map of spatial frequency content.
My Interest in Salience Maps: Autonomous Boat (Roboat)

Joint work with Northrop-Grumman, PGH
Finding Interesting Objects

Scenario: Roboat traverses a river on a mission. When it returns, we would like to see a pictorial log of all the “interesting” things it passed on the way.

Motivation for finding interesting scene features:

• unusual observations
• focus of attention
• use as landmarks
Determining Feature Salience

Approach: determine salience as difference from the norm.

Features / cues we examine while running:

- Contrast
- Color
- Motion
- Higher-Level Features
Example: High Contrast Corner Features

Original image

Corner features

gradients

2x2 sums over local windows

\[
\begin{align*}
\sum g_x \cdot g_x & \quad \sum g_x \cdot g_y \\
\sum g_x \cdot g_y & \quad \sum g_y \cdot g_y
\end{align*}
\]

(same approach as Shi and Tomasi)
Sample Corner Points

highest contrast corner feature is shown as a yellow crosshair
Automated Cameraman!

In each of these examples, the computer is deciding where to point the camera based on the location of the highest contrast corner/blob feature within its field of view.
Another Example: Color Salience

Is “red” a salient color, apriori?

Well, no. Clearly context is important.
Bayesian Surprise

http://ilab.usc.edu/research/

Quantifying the Wow! - A formal Bayesian theory of surprise

Why do some events immediately catch our attention while others just become ignored as background clutter? To address this question, we have developed and tested a new formal Bayesian theory of surprise, in collaboration with Prof. Pierre Baldi at the University of California at Irvine. Crucially, our theory emphasizes how, for something to become surprising, it much change your beliefs about the world. This new definition provides for the first time a formal mathematical framework with which surprise may be quantitatively measured. This theory complements Shannon's theory of information by emphasizing the effects an event may have onto the subjective beliefs of an observer, while Shannon’s theory emphasizes measuring the intrinsic objective complexity and predictability of the event. In experiments with human subjects, we found that surprise as defined by our theory is the strongest known attractor of human attention and gaze: observers looked towards surprising events in television and video game clips significantly more reliably than they looked towards simply colorful, contrasted, moving, or informative events. This work has widespread applications ranging from video surveillance to web search and advertising design.

Web page: See our Bayesian Theory of Surprise Home Page.

Selected Publications:
- More on the iLab publication server on Bayesian Theory of Surprise
Bayesian Surprise
http://ilab.usc.edu/research/

Input

Entropy

Surprise

Humans

KL = 0.151 ± 0.005

KL = 0.241 ± 0.006

KL = 0.679 ± 0.011
Measuring Attention/Saliency via Eye Gaze
Doug Decarlo at Rutgers: using eye tracking data to generate “meaningful” abstractions of images.
Visual Search

Lots of good work by Jeremy Wolfe

General idea: Try to determine what principles/features guide deployment of human attention by asking people to find a particular object embedded in a field of “distractors”, while measuring response time.
Fig. 2. Search for vertical (0 deg) among horizontal bars is easy (a), even if the items are defined by properties other than luminance contrast: texture (b), motion, depth, etc. Search for 0 deg among 5 deg tilted bars is hard (c), even though perceptual orientation-discrimination thresholds are much lower than 5 deg and cortical cells are sensitive to differences in orientation of less than 5 deg. Search for 0 deg among ±20 deg is hard (d), even though search for 0 deg among +20 would be easy.
Fig. 1. The core research task in visual search is to explain why some search tasks are easier than others. Finding the target blue–yellow–red ‘molecule’ is trivial in (a) because of the unique red element. Search is much less efficient in (b) because no unique feature defines the target and because we are particularly bad at search for targets defined by conjunctions of multiple colors.
find the objects that are different.

some differences just POP-OUT at you. some don’t.
claim... these get harder as you go left to right...

look for orange circle
Intriguing Idea  [citation?]  

When feature values of target are linearly separable from those of the distractors, visual search is very efficient.

Echos our work on feature selection for tracking 
Collins, Liu and Leordeanu 
Yin and Collins
Our Feature Selection Work

training frame

foreground

background

test frame

sorted variance ratio

0 5 10 15 20 25 30 35 40 45 50
Example: Feature Ranking

Best

Worst
More Sample Feature Rankings

Object/background designation  | Likelihood from most discriminative feature | Likelihood from least discriminative feature
---|---|---
R-2G+2B  | R+2G+2B  | -R-G+2B
-R+2G  | R-2B  | R-2G+2B
-R+2G  | R-2B  | R-2G+2B
-R+2G  | R+G-B  | 2R-B
Inattention

Sometimes you can be looking right at something and not be seeing it...

Related concept is change blindness. This may not be relevant to us, but there are some cool demos at

http://nivea.psycho.univ-paris5.fr/ASSChtml/ASSC.html
Interesting Question

Treisman says attention is the mechanism by which features get bound (grouped) into different objects (OVERSIMPLIFICATION ALERT)

But... object-based attention presumes that features are grouped into objects before attention takes place

How can this grouping happen?

Possibly via Gestalt principles of perceptual organization

  e.g. proximity, similarity, continuity, symmetry, common fate

Note: Relationship to study of figure/ground perception.
What is an object?

Tentative definition (Bob’s):

A coherent blob of stuff, that can be separated from the background, and moved around.
Coherence in color

Sample image

Sample likelihood map
Common Fate
Dissimilarity to Background

Wenmiao Lu, Yap-Peng Tan, Weiyun Yau, CVPR 2001
Occlusion Boundaries

(Yan Li at CMU)
Movement

Background subtraction

Layered Representation
Moveable
What is an object?

Tentative definition:

A coherent blob of stuff, that can be separated from the background, and moved around.

Self-similarity
Common fate

Dissimilarity
Occlusion boundaries

Moving
Moveable
Figure Ground Perception
Symmetry and P.O.

Figure 7.25
Symmetry and figure ground. Look to the left and to the right, and observe which colors become figure and which become ground. (Adapted from Hochberg, 1971.)
Symmetry and P.O.

Novel or unfamiliar figures look different in different orientations, provided that we view them naïvely and do not mentally rotate them. The reason may be the way in which a figure is “described” by the perceptual system. The colored figure at left could be described as a closed shape resting on a horizontal base with a protrusion on its left side and an indentation on its right side. The colored figure adjacent to it, although identical, would be described as a symmetrical shape resting on a curved base with a protrusion at the top. The first black figure could be described as a quadrilateral resting on a side. The black figure at right would be described as a diamondlike shape standing on end.
Occlusion and “Completion”

occlusion plays a fundamental role...

Figure 1. Examples of (a) amodal completion in space and (b) amodal integration in space and time.
Gestalt Grouping

Figure-ground separation; perceptual organization; figural completion; …
occlusion at work in visual search...

find horizontal pairs of triangles
Temporal Completion

Tunnel Effect demos:

http://pantheon.yale.edu/~bs265/demos/tunnel-CD.html

correct spatio-temporal motion is more important to determination of object “persistence” than sameness of surface material properties.
Natural Image Statistics

Figure 2. White-noise image, created by independently assigning the intensity of each pixel a random value, contains no statistical order and looks nothing like the natural scenes one is used to seeing.
Natural Image Statistics

Figure 6. Synthetic image that preserves the two-point correlations found in natural scenes appears curiously “natural.” But this image lacks the sharp discontinuities in intensity that are so commonly seen at the edges of objects.
Figure 2. White-noise image, created by independently assigning the intensity of each pixel a random value, contains no statistical order and looks nothing like the natural scenes one is used to seeing.

Figure 6. Synthetic image that preserves the two-point correlations found in natural scenes appears curiously "natural." But this image lacks the sharp discontinuities in intensity that are so commonly seen at the edges of objects.
optimal basis for sparse encoding of natural images. Looks like Gabor filters! More important, looks like known biological visual receptive fields.

Figure 7 Example basis functions derived using sparseness criterion (see Olshausen & Field 1996).
Statistics of Image Categories

Torralba. Work on “gists” (what type of scene is it)

Objects
- Face
- Pedestrian
- Car
- Cow
- Hand
- Chair

Scenes
- Mountain
- Beach
- Forest
- Highway
- Street
- Indoor

Objects in scenes
- Animal in natural scene
- Tree in urban scene
- Close-up person in urban scene
- Far pedestrian in urban scene
- Car in urban scene
- Lamp in indoor scene
presumably these are different enough that you could classify the type of scene of a new image...
using it to classify scenes containing/not containing different types of objects...

Figure 12. Average intensity and spectral signatures of sets of images constrained to contain specific objects. Image statistics can be predictions of the presence/absence of particular objects in the scene.
Figure 13. Performance in object prediction. For each object category we show performance for prediction of presence (left bar) of the objects and prediction of absence (right bar).
my question: to what extent does this presume a human took the picture? That is, did learn more about how photographers frame their shots than they did about what different types of objects look like?

**Figure 14.** Illustration of images organized according to the predicted likelihood of presence of animals using image statistics. Images in the centre are ambiguous in terms of image statistics and do not produce reliable predictions. Images located at the extremes provide reliable predictions of absence of animals or presence of animals.
Papers for Next Week

http://www.cse.psu.edu/~rcollins/CSE597E/

[PSYCHOLOGY]

[COGNITIVE SCIENCE]

[NEUROSCIENCE]

[COMPUTER VISION]