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

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-Snaech [builds on FIT]
- Itti and Koch, Salience Maps and Winner Take All
- Moser & Sitton, Salience Maps via Neural Nets
- Bundesen, Theory of Visual Attention (TVA)
- Logan, CODE Theory of Visual Attention (CTVA) [builds on TVA]

Attention Models
Common Theme...

Itti and Koch Salience Maps

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

\[
\text{min eigenvalue } \sum_{x,y} (g_{x+1,y} - g_{x,y})^2 \\
\text{2x2 sums over local windows}
\]

(same approach as Shi and Tomasi)

Sample Corner Points

Highest contrast corner features are shown as yellow crosshairs

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, a priori?

Well, no. Clearly context is important.
Bayesian Surprise
http://ilab.usc.edu/research/

Quantifying the Mind: A formal Bayesian theory of surprise

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

Measuring Attention/Salience 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.

Visual Search Tests

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 (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.

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

Example: Feature Ranking

Best

Worst
More Sample Feature Rankings

Object/background designation
Likelihood from most discriminative feature
Likelihood from least discriminative feature

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

Common Fate

Dissimilarity to Background

Occlusion Boundaries
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 7.25
Symmetry and figure ground. Look to the left and to the right, and observe which colors become figures and which become ground. (Adapted from Hochberg, 1971.)
Occlusion and “Completion”

occlusion plays a fundamental role...

Figure 1. Examples of (a) normal completion in space and (b) normal organization in space and time.

Gestalt Grouping

Figure-ground separation; perceptual organization; figural completion; ...

occlusion at work in visual search...

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

Natural Image Statistics

Figure 1. Natural image that possesses the low-pass characteristics found in natural scene appears (artificial) “natural.” This image lacks the sharp discontinuities in intensity that are so common seen at the edges of objects.
Natural Image Statistics

optimal basis for sparse encoding of natural images. Looks like Gabor filters. More important, looks like known biological visual receptive fields.

Figure 1. Example basis functions derived using a popular criterion (see Olshausen & Field, 1997).

Statistics of Image Categories

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

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 5. Spectral signatures of 15 different image categories. Each spectral signature is obtained by averaging the power spectra of a few hundred images in the category. The columns show coverage of 80, 90, and 90% of the energy of the spectral signatures (energy is obtained by adding the square of the Fourier components). The rate of the spectral signature is computed with the slope of a plot of a larger value of r (a factor by which the energy at high spatial frequencies is multiplied). A small r value produces a sharp decay of the energy at high spatial frequencies, which produces a smaller contrast. The overall shape is that of both r and s.

Figure 13. Performance in object prediction. For each object category we show performance for prediction of presence (left) and of absence (right).
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?