## **Overarching Principle**

Vision systems evolve to obtain information about the environment that is relevant for the tasks the organism must perform in order to survive and reproduce.

**Corollary**: The design of a vision system is constrained by the tasks it performs, by the physical/statistical properties of the environment, and by various biological factors.

## Approach to Handling Many of the Difficult Problems

- 1. Efficiently encode the attributes of retinal images in small regions with a foveated visual system.
- 2. Combine the measured local attributes into groups, categories or objects using mechanisms based on the physical laws and statistical facts of natural scenes, and on past experience.



Simple example demonstrating the consequences of context problem.



If the problem is not solved (at least partially) recognition of previously encountered objects is blocked. Perceptual grouping and segmentation mechanisms play a central role in solving the context problem for most natural stimuli.



Wertheimer and the other Gestalt psychologists were the first to fully appreciate the fundamental importance of grouping mechanisms for perception. **Proximity**: objects that are nearby tend to be grouped together. **Similarity**: objects that are similar tend to be grouped together. **Good continuation**: contour elements that are consistent with a smooth contour tend to be grouped together. **Closure**: Contours that are consistent with a closed form tend to be grouped together.



Using demonstrations similar to these they showed that grouping is based in part upon similarities along a number of stimulus dimensions.



**Good Continuation**: contour elements that are consistent with a smooth curve tend to be grouped together.



Gestalt principle of "closure." Because of good continuation the two straight line segments in A tend to look like a pair of crossing "sticks". Because of "closure" the same two line segments tend to be split at the middle to become parts of two "butterfly wings."



Perceptual grouping also makes use of principles that are based upon the three-dimensional properties of the environment. For example, these line segments are grouped into two boxes and a cylinder. Object corners occluding a background object tend to form an "L junction" or an "Arrow junction." Object corners that do not occlude a background object tend to form a "Fork junction." Occluded contours of an object tend to form "T junctions" with the contours of the occluding object.

Why are these grouping principles used by the brain? What other grouping principles might the brain use? How might the brain implement these principles? A good starting point is to examine the statistical relationship between the natural environment and the images formed in the eye (i.e., measure Bayesian statistics).

Although Gestalt psychologists and early computer vision scientists recognized the importance of grouping principles, they did not explicitly try to make the link to the properties of natural environment. Egon Brunswik was the first perception scientist to examine and think through the formal connection.





Bayesian statistics describe the probability of particular environmental properties given the observed image properties. These statistics can be useful for characterizing the information relevant for natural tasks, where the goal is to make accurate inferences about the environment. Unfortunately, Bayesian statistics are more difficult to measure than absolute statistics because ground-truth information about the environment must be obtained.



After deciding on a natural task, select physical scene properties potentially relevant to that task (often the selected properties are the things one wants to estimate or identify in the performing task), select potentially relevant image properties, measuring likelihood and prior probability distributions. To measure the quality of the information and how it might be used, determine performance of Bayesian (optimal) observer.



One specific approach to measuring the Bayesian statistics is to analyze images that have been hand segmented by human observers. I will describe a few examples of this approach from our lab.



To motivate the natural scene statistics approach to the study of pattern vision, consider three example tasks.





Our first step in measuring these Bayesian statistics was to extract small-scale edge elements from natural images with an automatic algorithm. 20 representative images (close-ups, distant shots, forests, mountains, ocean, sky, water, fields, animals). Each image analyzed separately.



Each red pixel in the right image is a edge element location. The orientation of each element was measured but is not shown here. Two observers then assigned edge elements to physical contours (sources); observers regarded boundary contours, lighting contours and surface marking contours as distinct. This assignment information was assumed to provide approximate ground truth.



Decimated edge samples with orientation shown.



The geometrical and contrast-polarity relationship between two edge elements is given by 4 parameters. Once images are hand segmented it is straight forward to estimate the likelihood and prior probability distributions. In the specific task we consider next, the prior probabilities are forced to be equal, so the relevant function is the likelihood ratio distribution which is plotted on the right. The reference is in the middle; distance is given by the ring, direction by the angle around the ring, orientation difference by the orientation of the plotted line segment, polarity by the particular half of the diagram, and likelihood ratio by the color of the plotted line segment.

For an earlier version of this analysis (without contrast polarity) see Geisler, Perry, Super & Gallogly (2001) *Vision Research*, 41, 711-724.

These average Bayesian pair-wise statistics make it to possible to determine optimal performance in the contour completion task.



In the contour completion task, a pair of edge elements is selected at random from a natural image and an occluder is placed between them. The task (the display is shown in B) is to indicate whether the pair of elements is from the same or different physical contour, where the prior probability is 0.5.

Three occluder diameters. No feedback is given for the first 600 trials, then 600 trials with feedback, then 600 trials with no feedback.



Comparison of human (symbols) and ideal (solid curves) performance, with (green) and without (red) contrast polarity information. Human efficiency is high and parallel to ideal. Average data for four observers (two experienced, two naïve).



Practice with feedback does not lead to improvements in performance. If anything performance gets worse. Unpracticed observers have excellent knowledge of the contour statistics of natural scenes. When they get feedback they may try to make trial-to-trial adjustments to their decision criteria which leads to non-optimal performance.



The pair-wise natural scene statistics can also be used to generate predictions for contour integration experiments (although it is not an ideal Bayesian observer). Here is an experiment we carried out several years ago to compare with the predictions from natural scene statistics. Computer vision researchers (e.g., Parent & Zuker 1989; Sha'ashua & Ullman 1988; Jacobs 1996) proposed algorithms for solving the contour integration problem. Hayes Field, Hess & Hayes (1993) were the first to do careful psychophysical work using this kind of task, and they raised awareness of the importance of the task in the biological vision science community.



Made parametric measurements for all four dimensions.



Groups obtained using pair-wise natural image statistics.



A Bayesian grouping rule based on natural scene statistics predicts human contour grouping performance quite well (correlation of about 0.9, as shown in next slide).





Image contours can occur for a number of entirely different physical reasons. They can be the result of surface boundaries, surface markings or shading. There can be little doubt that many perceptual tasks depend critically upon identifying whether a contour is a surface boundary, a marking or a shadow.

We have recently begun a systematic program to measure Bayesian statistics in the world of close-up foliage.

## Why Close Up Foliage?

"Divide and conquer": Foliage is a major component of the natural environment; other components will be studied later

Foliage comprises almost all of the natural environment of monkeys such as the macaque—the primary animal model for human vision

Close up foliage images are easy to hand segment accurately, making it relatively easy to measure Bayesian statistics

Statistics of distance foliage can be measured/inferred by reducing image scale



The statistics were measured from 2D (monocular) images obtained with a calibrated 36-bit camera. The camera is calibrated so that it gives us images where each pixel is described by the L, M and S cone responses. I won't describe the details of the calibration, but mention that it is sufficiently accurate for our purposes. foliage images using a 36-bit-per-pixel camera that allowed us to estimate L and M cone responses with a precision of 0.2% and S cone responses with a precision of 1%. (Usually, the variation in the image properties swamps the variation due to camera error.)

We have obtained close up images for large variety of foliage. There are a very large number of possible perceptual tasks we can consider using this image database. We have just considered a few so far.



This step does not add or subtract information, but makes the distribution of values more Gaussian and the dimensions a bit more statistically independent.



Here is a typical image captured with the calibrated camera. To obtain approximate ground truth, all the objects within or touching a region of interest were hand segmented. The region of interest (orange circle) was picked so there would be about 40 objects from each image.



Here is an example of the segmentation for 1 image. All the leaves and branches are segmented as separate objects, occlusions are marked, and even though some occluded objects have many detachments, the detachments are all linked together as single objects.



Close up of a leaf showing segmented surface boundary contours, shadow contours, and surface marking contours. These segmentations are assumed to provide approximate ground truth, the necessary ingredient for measuring Bayesian statistics.

We currently have over 2000 leaf objects segmented from over 60 images. The results I will describe next are based on all of the more than 2000 segmented leaf objects.



For the contour classification task, we measured Bayesian statistics for three local image measures. In general, the power of these statistics depend upon the length of contour that is analyzed. Here I focus on short contours, 32 pixels in length, which correspond on average to about 15% of the median length of a leaf's perimeter. There are many stimulus attributes that could be examined we are starting by looking at the simplest static monocular attributes.



Here are the results for the intensity difference measure. For simplicity I have plotted the joint distributions for only the luminance and red-green signals (blue-yellow signals are not shown).

The black line shows the optimal linear classification boundary.

The numbers indicate the percent classification accuracy. These give a lower bound on how useful the intensity difference information is the world of foliage for classifying contour boundaries.



In the world of foliage, contrast changes are larger across surface boundaries than across the other two categories, and larger across shadow contours than



In the world of foliage, surface markings are more bar-like than the other two categories.



Performance when all the measures are combined and when using single linear decision bound. Apparently, in the world of foliage it is possible to use three simple measures (at a small scale—32 pixels length) and obtain fairly good classification performance. The toughest problem is distinguishing shadow contours from surface boundary contours. We have not yet tested whether humans (or monkeys) can perform better with short contour segments. They may very well perform better, and we are currently exploring other potential sources of information.





Performance	
<u>Length</u>	Accuracy
16	81%
32	87%
64	93%
128	97%
*The median length of a continuous surface boundary is 97 pixels	

Performance ranges from 81% for contours 16 pixels long to 97% for contours 128 pixels long. Contiguous surface boundary contours have a median length of 97 pixels, so in the world of foliage it is possible to largely solve the foreground-background problem with these two rather simple image measures.

Duane Albrecht

Rob Frazor

Alison Crane

Jeff Perry

David Ing

Tony Wilson

Boaz Super

Don Gallogly

Tom Arnow