Five Difficult Problems for Vision Systems

_Illumination problem_
The illumination of scenes is highly variable and complex.

_Depth problem_
The images in the eyes are two-dimensional projections of the three-dimensional environment.

_Context problem_
Objects often appear in a complex and varying context of other objects.

_Viewpoint problem_
Objects are rarely seen from the same viewpoint.

_Category complexity problem_
The specific objects that define a category are often quite different.

_Fundamental Biological Constraints_
Limited neural resources, dynamic ranges, and physical space

Most natural tasks involve dealing with one or more of these difficult general problems. Furthermore, the solutions that the visual system can come up with in natural tasks are constrained by various fundamental 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.
Five Difficult Problems for Vision Systems

Illumination problem
The illumination of scenes is highly variable and complex.

Depth problem
The images in the eyes are two-dimensional projections of the three-dimensional environment.

Context problem
Objects often appear in a complex and varying context of other objects.

Viewpoint problem
Objects are rarely seen from the same viewpoint.

Category complexity problem
The specific objects that define a category are often quite different.

Fundamental Biological Constraints
Limited neural resources, dynamic ranges, and physical space
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.
Schreech owls have feathers that match the features (bark) of oak trees (the trees they frequently inhabit). If their features cannot be grouped separately from the background they are more difficult to recognize.
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.
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.

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.
Natural Systems Analysis

1. Identify and characterize natural tasks

2. Measure and analyze relevant environmental properties (e.g., measure natural scene statistics) and biological constraints

3. Determine how to exploit the measured environmental properties to perform natural tasks optimally, given the biological constraints

4. Formulate hypotheses and test them in physiological and behavioral studies that capture the essence of the natural task
The structure of many natural tasks is to estimate some property of the environment from the proximal stimulus encoded by the sensory organs.
2. Relevant Natural Scene Statistics

Joint probability of environmental state and stimulus

\[ p(\omega, s) = p(s|\omega)p(\omega) \]

which specifies the posterior probability of an environmental state for any given proximal stimulus

\[ p(\omega|s) \]
Make the estimate that maximizes the gain averaged over the probability of the different states of the environment, given the observed stimulus.

$$\hat{\omega}_{opt} = \max_{\hat{\omega}} \left[ \sum_{\omega} \gamma(\omega, \hat{\omega}) p(\omega | s) \right]$$
The contour occlusion task is an important subtask in many natural vision tasks.
1. Contour Occlusion Task

Are the contours in the patches from the same or different physical contour?

- $\omega$ = 'same' or 'different'
- $s = (d, \phi, \theta, \rho)$
- $\hat{\omega}$ = 'same' or 'different'
- $\gamma(\omega, \hat{\omega}) = 'maximize accuracy' $


In other words, we are only considering the stimulus features represented in this display.

\begin{itemize}
\item $\omega = \text{'same' or 'different'}$
\item $s = (d, \phi, \theta, \rho)$
\item $\hat{\omega} = \text{'same' or 'different'}$
\item $\gamma(\omega, \hat{\omega}) = \text{'maximize accuracy'}$
\end{itemize}
Measure the joint distribution of the state of the environment and the stimulus features under consideration. To do this we first extracted edge elements from a diverse collection of natural images. For example let’s zoom in on the indicated patch.
The yellow dots the locations of the extracted edge elements. We then assigned these elements to physical sources. See arrows at the bottom. These assignments allow us to estimate the joint probability distributions.
Bayesian Ideal Observer: Contour Occlusion Task

Optimal decision rule:

if \( p(c|d,\phi,\theta,\rho) > p(\sim c|d,\phi,\theta,\rho) \) then respond "c"

Equivalent optimal decision rule:

if \( \frac{p(\phi,\theta,\rho|d,c)}{p(\phi,\theta,\rho|d,\sim c)} > \frac{p(\sim c|d)}{p(c|d)} \) then respond "c"

or,

if \( I(\phi,\theta,\rho|d) > \beta(d) \) then respond "c"
Here are the measured statistics; b shows the distance dependent likelihood ratio and c shows the distance dependent decision criterion. (15500+ bins)

Once we have these distributions we can determine the optimal decision rule for performing the contour occlusion task (given only the geometry and contrast polarity features).
For a given distance and same polarity the elements should be grouped if the
direction and orientation difference fall within the black contour.
The last step in a natural systems analysis is to formulate hypotheses and test them in behavioral or neural experiments. 1800 trials per subject with effectively no feedback.
1. Histogram of human and ideal observer performance for hits, misses, false alarms and correct rejections, for each geometrical and contrast relationship and occlusion diameter.

2. The proportion of responses in each bin is similar for human and ideal (no obvious parts of the stimulus space where humans differ from optimal based on average scene statistics).

3. This shows that naïve humans without feedback have a good sense of natural contour statistics. There is no time to discuss these results further, but they rule out most (all) existing models of contour interpolation. The scene statistics suggest a family of models.
Has the natural systems analysis approach really told us much new that we did not already know? Did we simply show that the principle of good continuation or David Field’s “association field” has a physical basis in natural scenes? Would the parameter free predictions of a simple constraint like relatable edges that was hypothesized without measuring scene statistics work just as well? Nope.
An experiment to measure human ability to detect contours, where the only information available in the stimulus for performing the task is the geometry of the edge elements.
Examples of the different dimensions of contour shape that were tested in the experiment in previous slide. In each case, the contour was embedded in a dense background of random contour elements.
The connected contours on the right are the contour groups obtained using the statistics of contours in natural images. Notice how they roughly match the contours you “see” in the left side. In the forced choice experiment, we predict that humans pick the interval with the longest group of edge elements.
There is a high correlation between human ability to detect contours and the model (hypothesis) based directly upon the statistics of contour geometry in natural images. The neural circuits in the brain that perform contour grouping are unknown at this time (although there are some hints from neurophysiological data).