COMP 546
Lecture 18
Perceptual Organization, Object Recognition
Thurs. March 23, 2017
What is perceptual organization?

An image is an array of intensities (RGB).

We don’t perceive the individual points. Instead we perceive groups, parts, hierarchies, ....

How does the visual system organize an image?
Gestalt Rules of Grouping

‘Gestalt’ is a German word for (take) shape, form.

Wertheimer, Kohler, Koffka, ... (Germany 1920’s & early 1930’s) came up with a set of grouping rules.

More demos than experiments.
Proximity

Items that are closer together tend to be grouped.
Good continuation (curve)

Group points along smooth curves.
Here, an alternative would be to see two curves in contact.
Similarity (luminance/color)

Group similar items e.g. color.
Similarity (Size)

Group items of the similar size (or orientation, ...)

Similarity (Orientation)
Similarity (Motion)

Powerpoint animation (not available in PDF)
Closure (and illusory contours)

Often shapes are missing parts (either regions or boundaries). Here we fill in the missing parts and perceive three disks and two triangles. [Kanizsa 1955]
Convexity

The occluded shape on the right is more likely to be seen as a single shape than the one on the left. This can be shown, for example, using depth discrimination by stereo disparity (Liu 1999).
Gestalt theory assumes that items are given. Careful: not so! The items themselves must be grouped out of their pixels.
Uniform Connectedness
[Palmer and Rock, 1994]

“Uniform connectedness”: Any connected region that is defined by a uniform property will be perceived as an image region.

This uniform property can be a single color (item), or it can be a common feature shared by many elements (region defined by items).

We perceive connected 2D regions, not just groups of items.
Psychophysics of Contour Grouping

e.g. Contour popout? [Field 1993]
To group or not? [Zucker (many papers)]

No

Yes
Bayesian model of contour grouping

(Elder and Goldberg 2003)

Given two image edges \((x_i, y_i, \theta_i), (x_j, y_j, \theta_j)\), what is the probability that they are on the same contour?
In natural images, edges that do belong to the same object boundary have highly correlated $\theta_{ij}$ angles.
Use statistics of natural images edges to model conditional probabilities $p(I_i \mid S)$:

$I_i$ is an image property (strength of a Gestalt cue),

e.g. (1) proximity, (2) good continuation, (3) luminance similarity

$S$ is a binary variable indicating if two edges are on the same contour.

Given two image edges $(x_i, y_i, \theta_i), (x_j, y_j, \theta_j)$, the posterior probability that they are on the same contour is:

$$p(S \mid I_1 I_2 I_3) \approx p(I_1 \mid S) p(I_2 \mid S) p(I_3 \mid S) p(S)$$
\[ p(S \mid I_1 I_2 I_3) \approx p(I_1 \mid S) p(I_2 \mid S) p(I_3 \mid S) \ p(S) \]
Next topic: PO & “mid-level vision”

Given a 2D image, assume the visual system can form contours and regions e.g. using principles just discussed.

Next, how to interpret these contours and regions in terms of 3D scene properties e.g. depth, material, illumination?
Low reflectance, high illumination

Low reflectance, low illumination

High reflectance, low illumination

Reflectance image

Illuminance image

[images from Adelson 2000]
Reflectance edge

Illumination edge

(same local images)

“Junctions”
X junctions often indicate shadows or transparent layers.

[Kingdom 2003]
T junctions indicate occlusions

X junctions indicate transparency (or shadows)
Recall White’s illusion.

Short white bars on the left are either painted onto the black bars or they belong to a transparent foreground rectangle. The short dark grey bars on the right belong to a background rectangle that is occluded by the long black bars.
The four disks in the left image have the same intensity as the four in the right, but the percept of the surface colors are very different.

Perceptual organization: the visual system seems to use a multilayered fog model to explain the intensities.

[Anderson & Winowar 2005]
From PO to Object Recognition
Why is object recognition difficult?

• object can be at any position, size, pose

• illumination can vary

• object can be partly occluded or camouflaged

• objects can be classified or recognized at many levels:
  
  Animal ... Dog ... Beagle ... my dog

  Vehicle ... Car ... Volvo XC 70 (2006) ... my car

  ...
Classic theory of recognition (Marr, 1982)

1. Image (2D)
2. Lines, edges, texture, shading (2D)
3. Local surface depth, slant & tilt ("2.5 D")
4. Volumes & surfaces: cylinders, spheres, ...

Match

3D object models stored in memory
‘2.5D Sketch” (Marr 1982)
3D object models stored in memory

![Diagram showing 3D object models](image-url)
Hierarchy of 3D Models

<table>
<thead>
<tr>
<th>cylinder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>limb</th>
<th>quadruped</th>
<th>biped</th>
<th>bird</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>thick-limb</th>
<th>cow</th>
<th>human</th>
<th>ostrich</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>thin-limb</th>
<th>giraffe</th>
<th>ape</th>
<th>dove</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Q: What’s wrong with Marr’s theory?

• Computing an accurate 2.5D sketch is difficult.

• Unclear how to go from a 2.5 sketch to a 3D model.

• Often we can recognize objects from very limited image information.

• Even if we can perceive some 3D shapes accurately (from stereo, etc), this does not imply that 3D shape is used for recognition.
Last 20 years of Object Recognition: Machine Learning

models in memory: large collections of 2D image patches + classifiers (parameters in a neural network)
Example: Human vs Computer Vision Experiment

[Ullman 2016]

How much can you shrink or crop images before recognition rates fall?
The cropping and resolution size where human performance fell was quite different from where current machine learning method fell. (Authors concluded that computer vision models do not predict human vision performance.)