Where are we?

**Image Formation**
- Human vision
- Cameras
- Geometric Camera models
- Camera Calibration
- Radiometry
- Color

**Early Vision (one image)**
- Linear Filters
- Edge Detection
- Texture
- Motion

**Early Vision (Multiple images)**
- Geometry of Multiple images
- Stereo

**Mid-Level Vision:**
- Segmentation
  - By clustering
  - By model fitting
  - Probabilistic
  - Tracking

**High-Level Vision:**
- Model-based vision
- Appearance-based vision
Outlines

• Mid-level vision
• What is segmentation
• Perceptual Grouping
• Segmentation by clustering

Mid-level vision

• Vision as an inference problem:
  – Some observation/measurements (images)
  – A model
  – Objective: what caused this measurement?
• What distinguishes vision from other inference problems?
  – A lot of data.
  – We don’t know which of these data may be useful to solve the inference problem and which may not.
    • Which pixels are useful and which are not?
    • Which edges are useful and which are not?
    • Which texture features are useful and which are not?
Why do these tokens belong together?

It is difficult to tell whether a pixel (token) lies on a surface by simply looking at the pixel.

Segmentation

- Can we achieve a compact and suggestive representation of the interesting image data that emphasizes the properties that make it interesting
  - Segmentation
  - Grouping
  - Perceptual organization
  - Fitting
- What is interesting and what is not depends on the application
General ideas

- **tokens**
  - whatever we need to group (pixels, points, surface elements, etc., etc.)
- **top down segmentation**
  - tokens belong together because they lie on the same object
- **bottom up segmentation**
  - tokens belong together because they are locally coherent

- **Grouping (or clustering)**
  - collect together tokens that “belong together”
- **Fitting**
  - associate a model with tokens
  - issues
    - which model?
    - which token goes to which element?
    - how many elements in the model?

Segmentation

Different problems – same problem: segmentation

- Summarizing a video: segment a video into shots, find coherent segments in the video, find key frames…
- Finding machine parts: finding lines, circles,…
- Finding people: find body segments, find human motion patterns
- Finding buildings from aerial imagery: find polygonal regions, line segments…
- Searching a collection of images: find coherent color, texture regions, shape…
- …
Segmentation

Segmentation is a big topic
We will look into:

• Segmentation by clustering: Forming image segments:
  – How to decompose the image into “superpixels” image regions that are coherent in color and texture
  – Shape of the region is not that important while segmenting

• Segmentation by model fitting:
  – Fitting lines and curves to edge points:
  – Which points belong to which line, how many lines ?
  – What about more complicated models, e.g. fitting a deformable contour!

Segmentation as Clustering

• Objective: Which components of a data set naturally belong together
• This is a clustering problem which can be done in two ways:
  • Partitioning – Decomposition:
    – Starting from a large data set how to partition it into pieces given some notion of association between data items
      • Decompose an image into regions that have coherent color and texture
      • Decompose a video sequence into shots
  • Grouping
    – Collect sets of data item that make sense together given our notion of association
      • Collect together edge segments that seems to belong to a line
• Question: what is our notion of association ?
One view of segmentation is that it determines which component of the image form the figure and which form the ground.

What is the figure and the background in this image?

Grouping and Gestalt

- Gestalt: German for form, whole, group
- Laws of Organization in Perceptual Forms (Gestalt school of psychology) Max Wertheimer 1912-1923

“there are contexts in which what is happening in the whole cannot be deduced from the characteristics of the separate pieces, but conversely; what happens to a part of the whole is, in clearcut cases, determined by the laws of the inner structure of its whole”

Muller-Layer effect:
This effect arises from some property of the relationships that form the whole rather than from the properties of each separate segment.
Grouping and Gestalt

- Can we write down a series of rules by which image elements would be associated together and interpreted as a group?
- What are the factors that makes a set of elements to be grouped?
- Human vision uses these factors in some way.

**Not Grouped**

- **Proximity**:
  Tokens that are nearby tend to be grouped together.
- **Similarity**:
  Similar tokens tend to be grouped together.
- **Common Fate**:
  Tokens that have coherent motion tend to be grouped together.

**Common Region**:
Tokens that lie inside the same closed region tend to be grouped together.
**Parallelism:** Parallel curves or tokens tend to be grouped together

**Symmetry:** Curves that lead to symmetric groups are grouped together

**Continuity:** Tokens that lead to continuous curves tend to be grouped

**Closure:** Tokens or curves that tend to lead to closed curves tend to be grouped together.

Familiar configuration: tokens that, when grouped, lead to a familiar object tend to be grouped
Occlusion appears to be a very important cue in grouping
Illusory contours: tokens are grouped together because they provide a cue to the presence of an occluding object.
These rules function as explanation only
Very hard to form algorithms
When one rule applied and when another?

Image Segmentation as Clustering

• Cluster together (pixels, tokens, etc.) that belong together
• Pixels may belong together because they have the similar intensity, color, texture, and they are nearby.
• Framework:
  – (Representation) For each pixel extract a feature vector describing:
    • Intensity, color, texture (filter response)
    • Spatial location
  – (Clustering) Cluster the feature vectors
  – Replace each pixel by its cluster representation.
Image Segmentation as Clustering

- Progress:
- 1970s: Hierarchical Clustering
- 1980s: Markov Random Fields (MRF)
- 1990s:
  - Graph theoretic clustering – Graph cuts
  - Mean shift clustering (2000+)

Segmentation as clustering

Hierarchical Clustering:
- Agglomerative clustering – clustering by merging – bottom-up
  - Each data point is assumed to be a cluster
  - Recursively merge clusters
  - Algorithm:
    - Make each point a separate cluster
    - Until the clustering is satisfactory
    - Merge the two clusters with the smallest inter-cluster distance

- Divisive clustering – clustering by splitting – top-down
  - The entire data set is regarded as a cluster
  - Recursively split clusters
  - Algorithm:
    - Construct a single cluster containing all points
    - Until the clustering is satisfactory
    - Split the cluster that yields the two components with the largest inter-cluster distance
Segmentation as clustering

- Two main issues:
  - What is a good inter-cluster distance
    - single-link clustering: distance between the closest elements -> extended clusters
    - complete-link clustering: the maximum distance between elements -> rounded clusters
    - group-average clustering: Average distance between elements – rounded clusters
  - How many clusters are there (model selection)
  - Dendrograms
    - yield a picture of output as clustering process continues
K-Means

- Choose a fixed number of clusters $K$
- Each cluster has a center (mean) $\mu_i$
- Choose
  - cluster centers and
  - point-cluster allocations to minimize error
- can’t do this by search, because there are too many possible allocations.

Algorithm:

Repeat until centers are unchanged:
- fix cluster centers; allocate points to closest cluster
- fix allocation; compute cluster centers

$x$ could be any set of features for which we can compute a distance (careful about scaling)

$$\sum_{i \in \text{clusters}} \left( \sum_{j \in \text{elements of } i\text{th cluster}} \left\| x_j - \mu_i \right\|^2 \right)$$

K-means clustering using intensity alone and color alone
K=5 segmented image is labeled with cluster means
K-means using color alone, 11 segments

K-means using color alone, 11 segments.
K-means using color and position, 20 segments

Sources

- Slides by
  - D. Forsyth @ Berkeley