



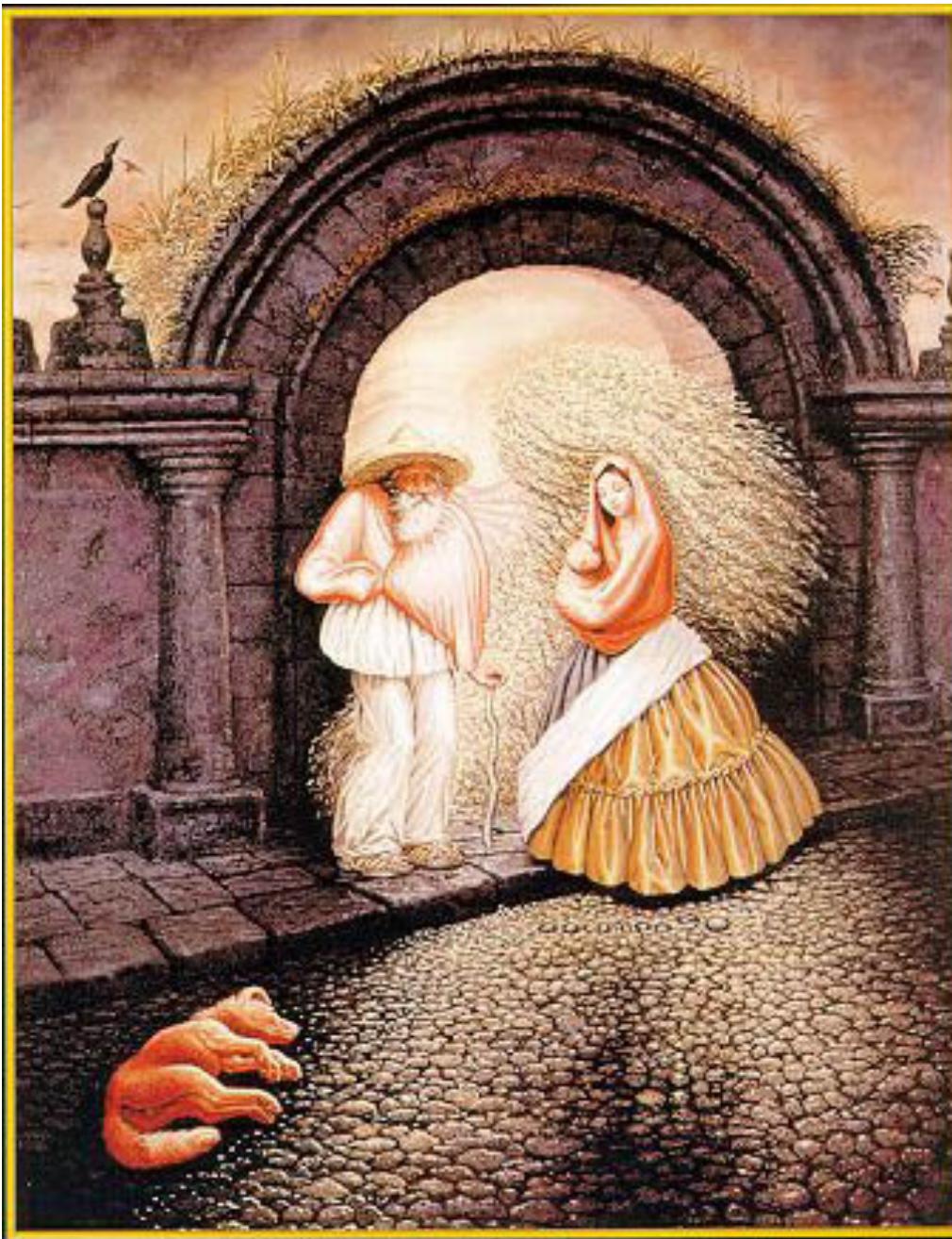
# Lecture 12: Clustering and Segmentation

Professor Fei-Fei Li  
Stanford Vision Lab

# What we will learn today

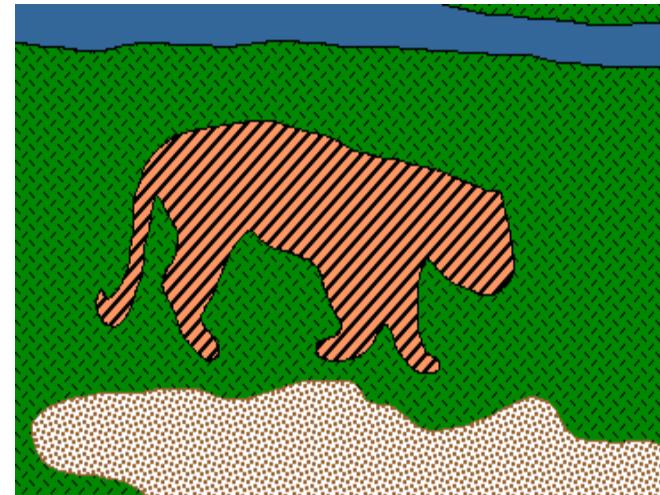
- Introduction to segmentation and clustering
- Gestalt theory for perceptual grouping
- Agglomerative clustering

**Reading:** [FP] Chapters: 14.2, 14.4



# Image Segmentation

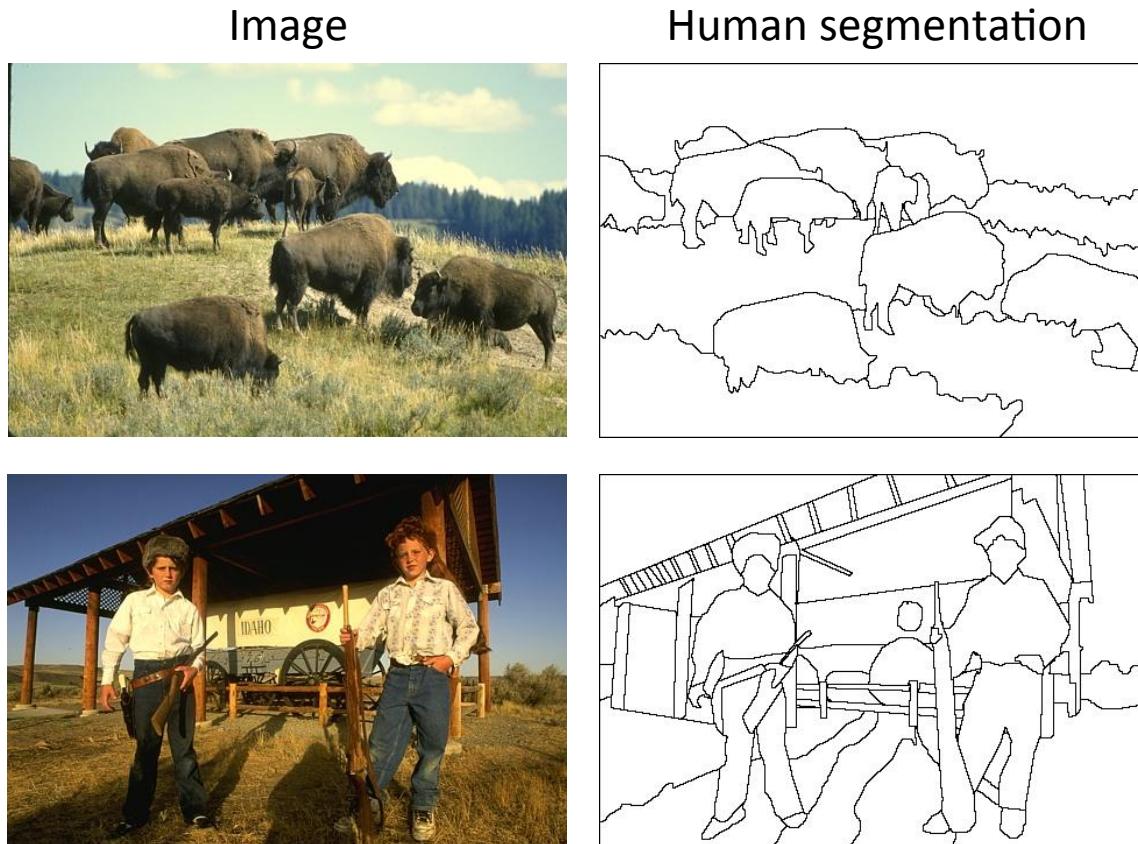
- Goal: identify groups of pixels that go together



Slide credit: Steve Seitz, Kristen Grauman

# The Goals of Segmentation

- Separate image into coherent “objects”

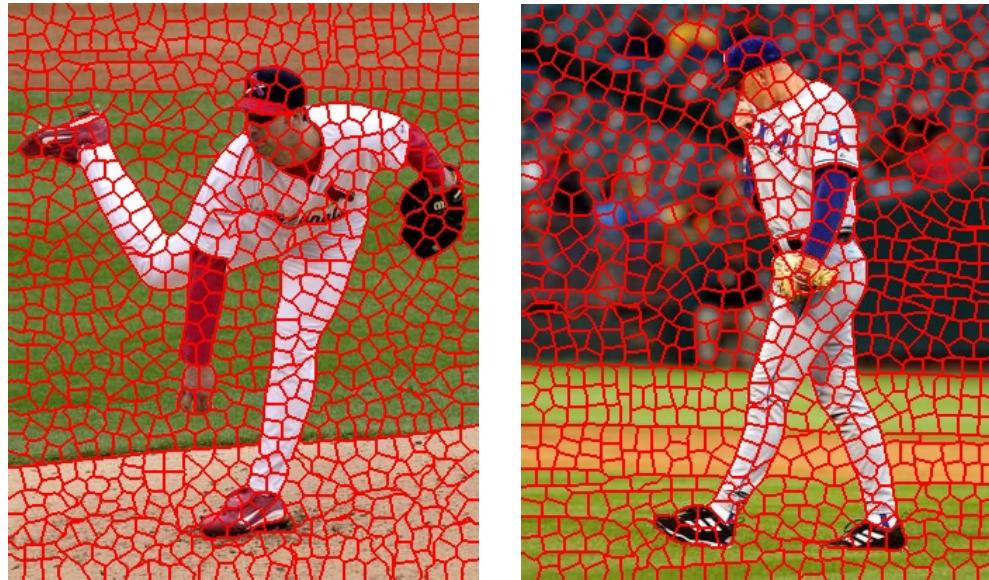


Slide credit: Svetlana Lazebnik

# The Goals of Segmentation

- Separate image into coherent “objects”
- Group together similar-looking pixels for efficiency of further processing

“superpixels”



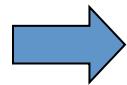
X. Ren and J. Malik. [Learning a classification model for segmentation](#). ICCV 2003.

# Segmentation for feature support

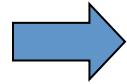


Slide: Derek Hoiem

# Segmentation for efficiency



[Felzenszwalb and Huttenlocher 2004]



[Hoiem et al. 2005, Mori 2005]

[Shi and Malik 2001]

Slide: Derek Hoiem

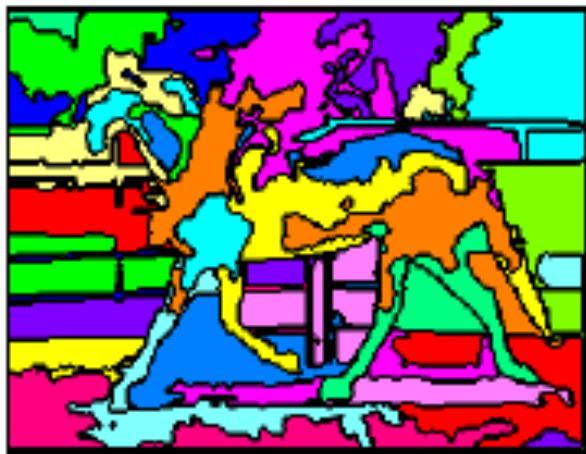
# Segmentation as a result



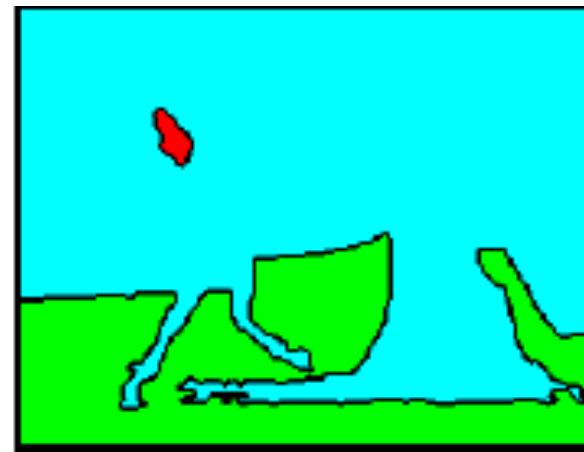
Rother et al. 2004



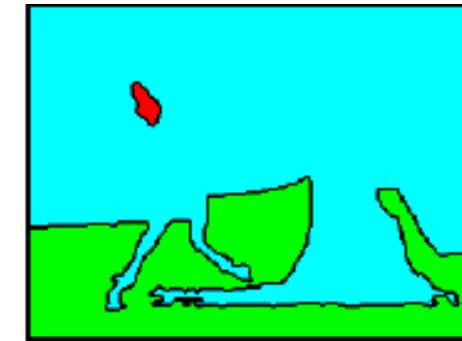
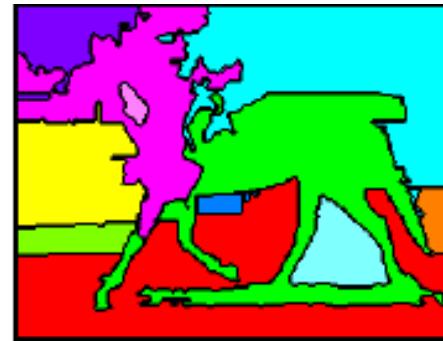
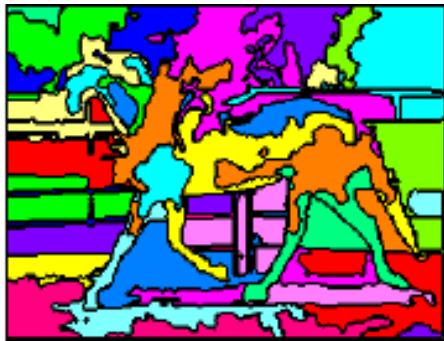
# Types of segmentations



Oversegmentation



Undersegmentation



Multiple Segmentations

One way to think about  
“segmentation” is Clustering

Clustering: group together similar points and represent them with a single token

Key Challenges:

- 1) What makes two points/images/patches similar?
- 2) How do we compute an overall grouping from pairwise similarities?

Slide: Derek Hoiem

# Why do we cluster?

- **Summarizing data**
  - Look at large amounts of data
  - Patch-based compression or denoising
  - Represent a large continuous vector with the cluster number
- **Counting**
  - Histograms of texture, color, SIFT vectors
- **Segmentation**
  - Separate the image into different regions
- **Prediction**
  - Images in the same cluster may have the same labels

Slide: Derek Hoiem

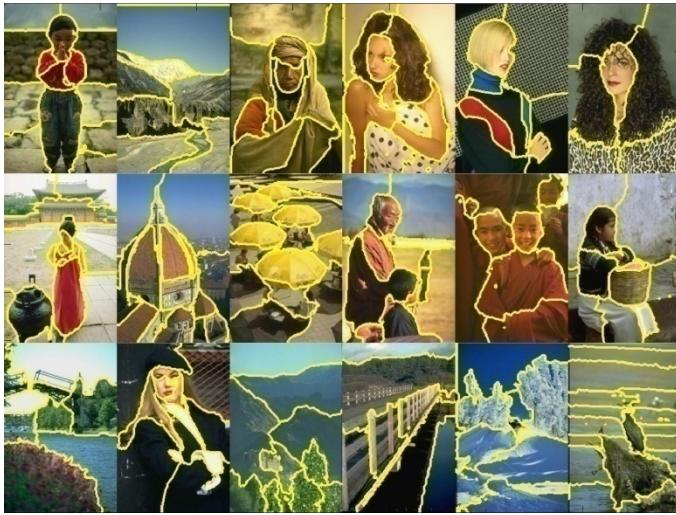
# How do we cluster?

- Agglomerative clustering
  - Start with each point as its own cluster and iteratively merge the closest clusters
- K-means (**next lecture**)
  - Iteratively re-assign points to the nearest cluster center
- Mean-shift clustering (**next lecture**)
  - Estimate modes of pdf
- Spectral clustering (CS231a, winter quarter)
  - Split the nodes in a graph based on assigned links with similarity weights

# General ideas

- **Tokens**
    - whatever we need to group (pixels, points, surface elements, etc., etc.)
  - **Bottom up clustering**
    - tokens belong together because they are locally coherent
  - **Top down clustering**
    - tokens belong together because they lie on the same visual entity (object, scene...)
- > These two are not mutually exclusive

# Examples of Grouping in Vision



Determining image regions

*What things should  
be grouped?*

*What cues  
indicate groups?*



Grouping video frames into shots

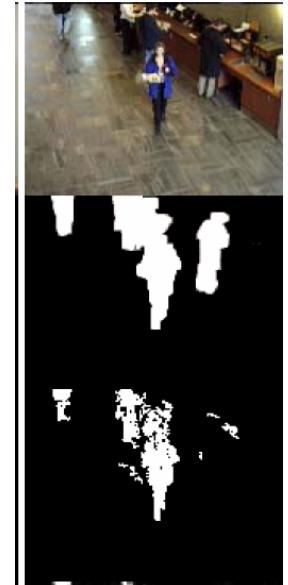
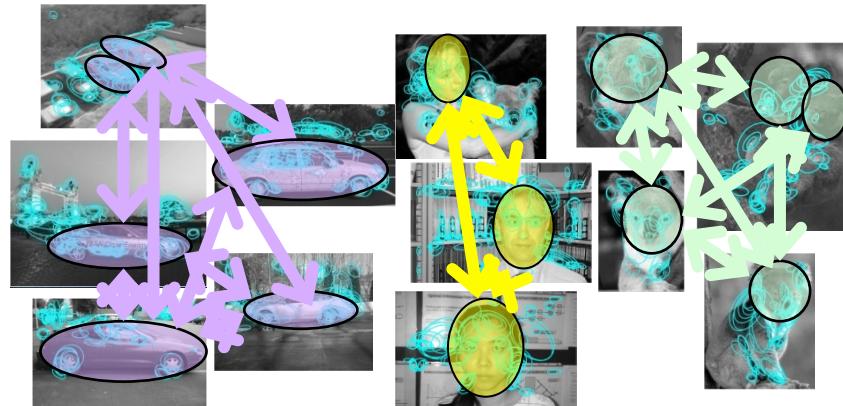


Figure-ground



Object-level grouping

Slide credit: Kristen Grauman

# Similarity



Slide credit: Kristen Grauman

# Symmetry



Slide credit: Kristen Grauman

# Common Fate



Image credit: Arthus-Bertrand (via F. Durand)

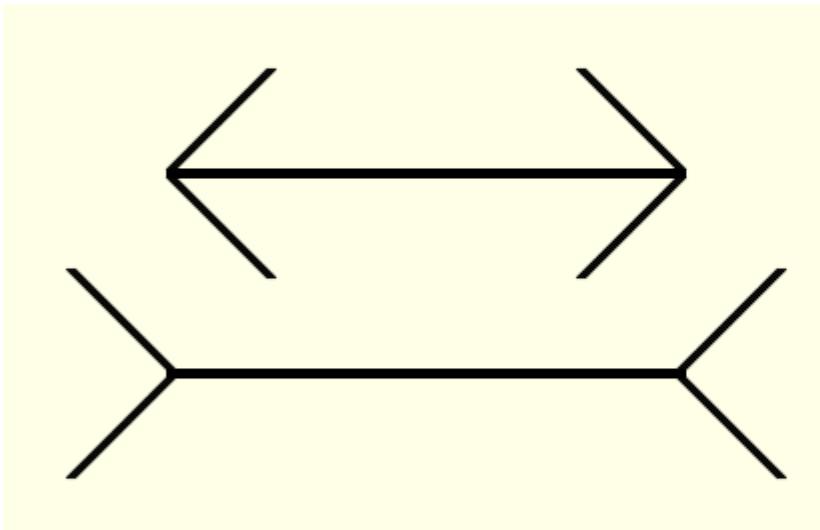
Slide credit: Kristen Grauman

# Proximity



Slide credit: Kristen Grauman

# Muller-Lyer Illusion



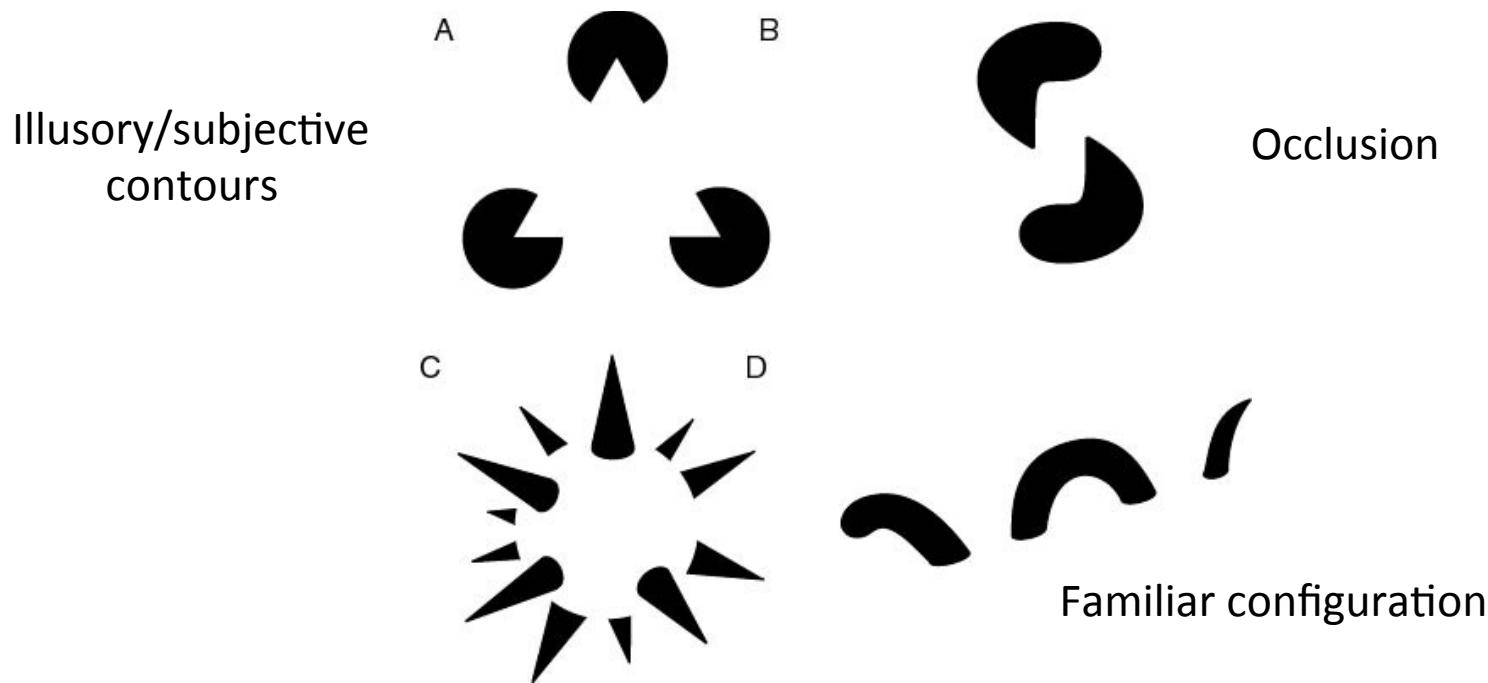
- What makes the bottom line look longer than the top line?

# What we will learn today

- Introduction to segmentation and clustering
- **Gestalt theory for perceptual grouping**
- Agglomerative clustering

# The Gestalt School

- Grouping is key to visual perception
- Elements in a collection can have properties that result from relationships
  - “The whole is greater than the sum of its parts”



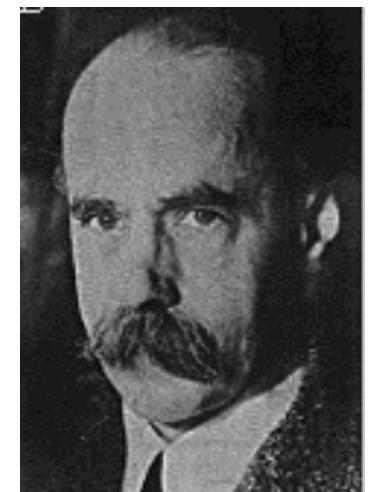
[http://en.wikipedia.org/wiki/Gestalt\\_psychology](http://en.wikipedia.org/wiki/Gestalt_psychology)

# Gestalt Theory

- Gestalt: whole or group
  - Whole is greater than sum of its parts
  - Relationships among parts can yield new properties/features
- Psychologists identified series of factors that predispose set of elements to be grouped (by human visual system)

*"I stand at the window and see a house, trees, sky.  
Theoretically I might say there were 327 brightnesses  
and nuances of colour. Do I have "327"? No. I have sky, house,  
and trees."*

Max Wertheimer  
(1880-1943)



Untersuchungen zur Lehre von der Gestalt,  
*Psychologische Forschung*, Vol. 4, pp. 301-350, 1923  
<http://psy.ed.asu.edu/~classics/Wertheimer/Forms/forms.htm>

# Gestalt Factors



Not grouped



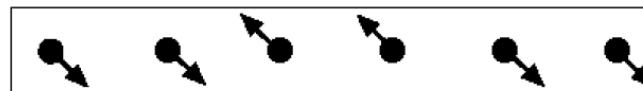
Proximity



Similarity



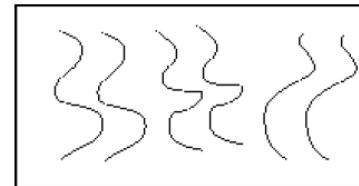
Similarity



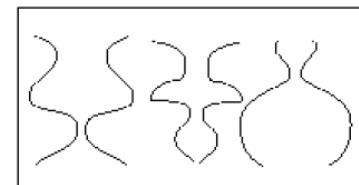
Common Fate



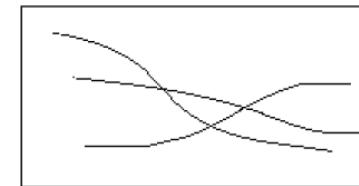
Common Region



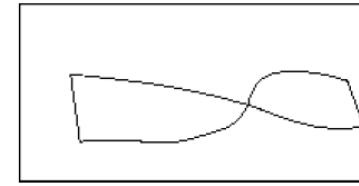
Parallelism



Symmetry



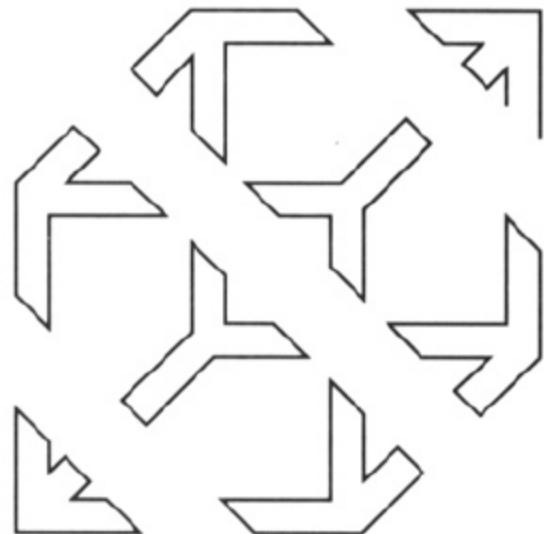
Continuity



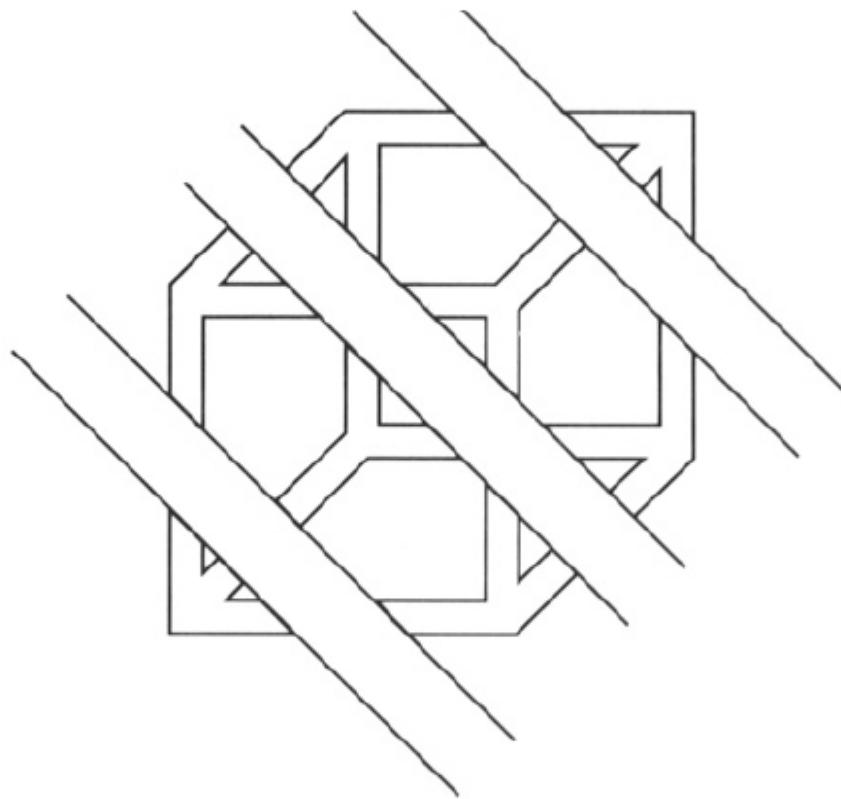
Closure

- These factors make intuitive sense, but are very difficult to translate into algorithms.

# Continuity through Occlusion Cues



# Continuity through Occlusion Cues



Continuity, explanation by occlusion

# Continuity through Occlusion Cues

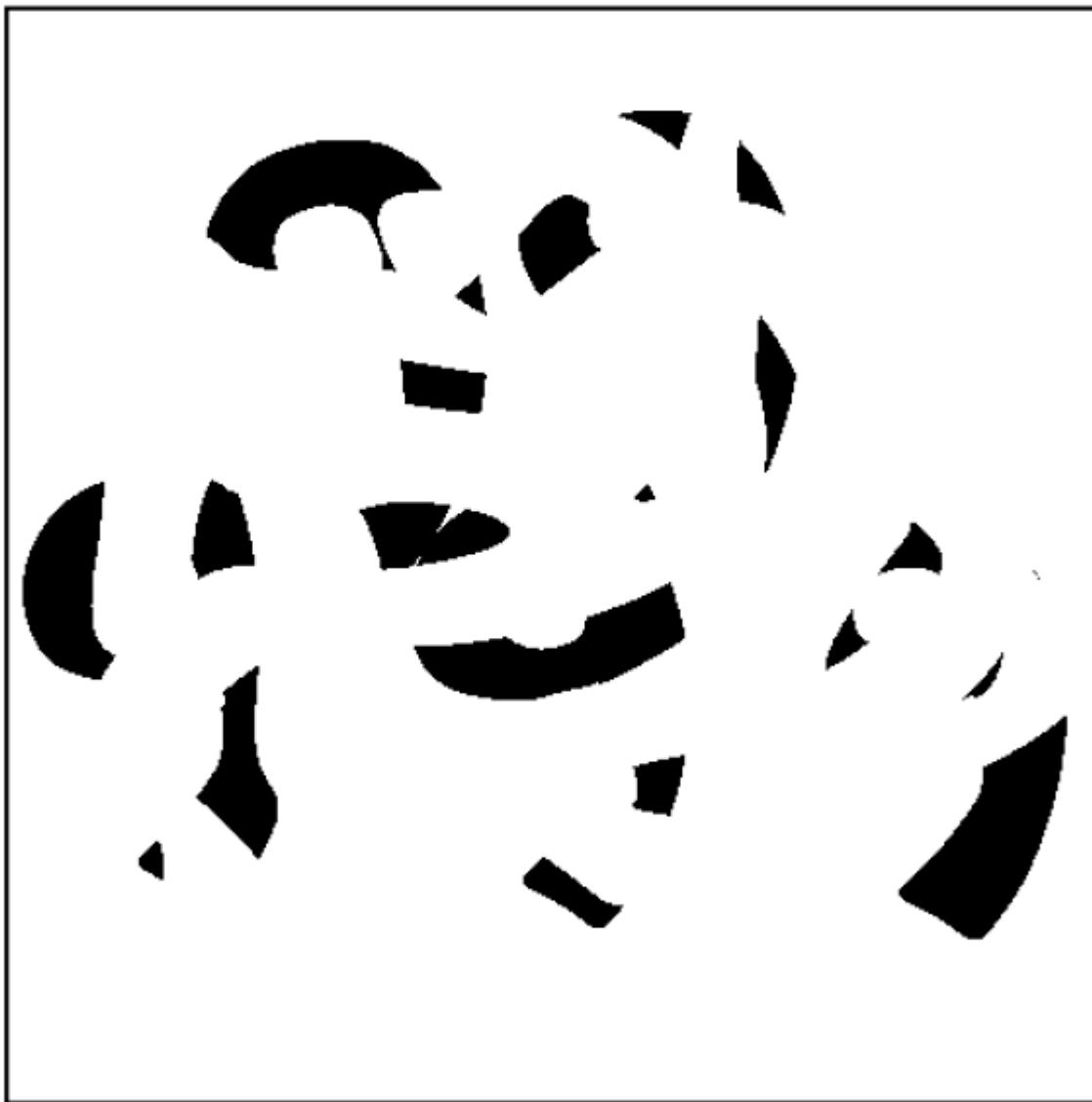


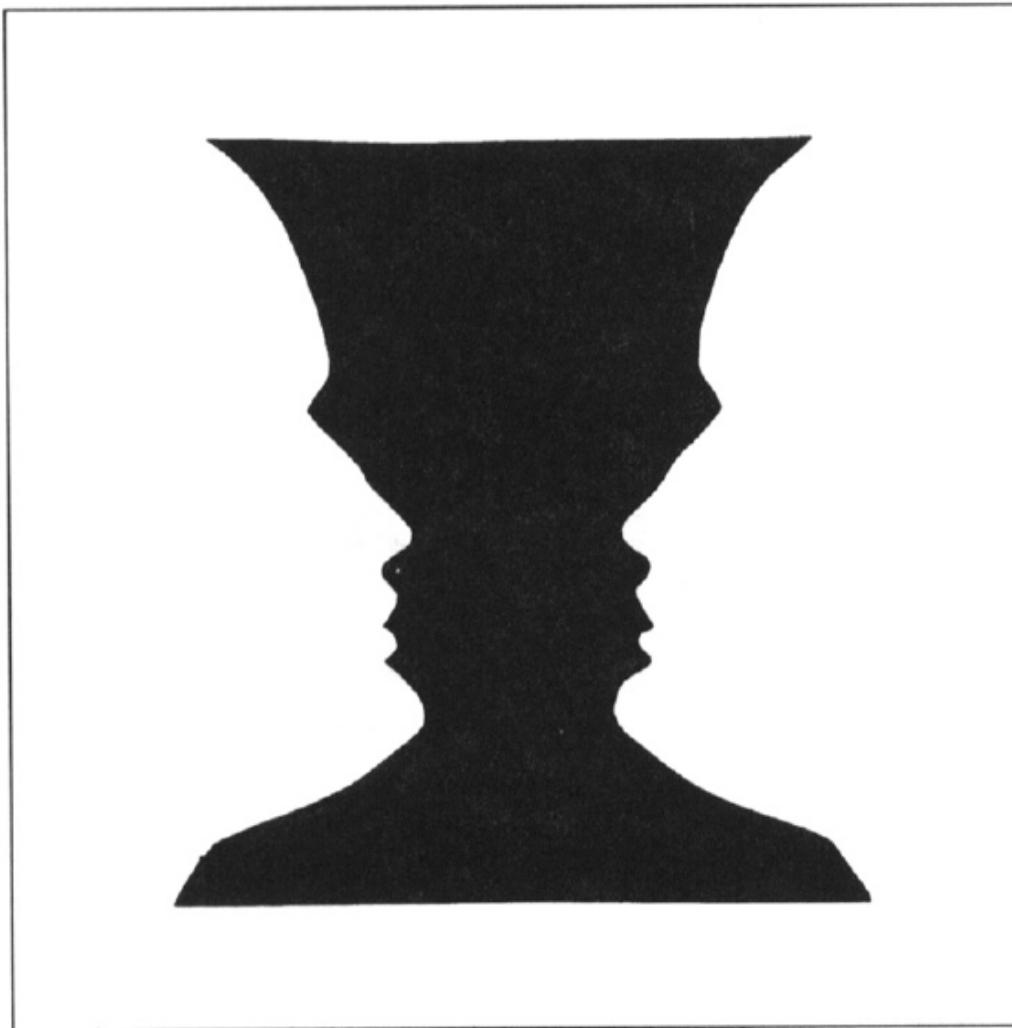
Image source: Forsyth & Ponce

# Continuity through Occlusion Cues



Image source: Forsyth & Ponce

# Figure-Ground Discrimination



# The Ultimate Gestalt?



# What we will learn today

- Introduction to segmentation and clustering
- Gestalt theory for perceptual grouping
- Agglomerative clustering

technical  
note

# Clustering: distance measure

Clustering is an unsupervised learning method. Given items  $x_1, \dots, x_n \in \mathbb{R}^D$ , the goal is to group them into clusters. We need a pairwise distance/similarity function between items, and sometimes the desired number of clusters.

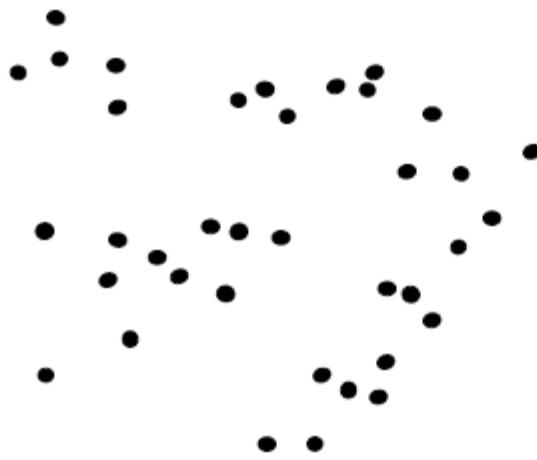
When data (e.g. images, objects, documents) are represented by feature vectors, a commonly used similarity measure is the *cosine similarity*. Let  $x, x'$  be two data vectors. There is angle  $\theta$  between the two vectors  $x, x'$ . The cosine similarity is defined as

$$\begin{aligned} sim(x, x') &= \cos(\theta) \\ &= \frac{x^\top x'}{\|x\| \cdot \|x'\|} \\ &= \frac{x^\top x'}{\sqrt{x^\top x} \sqrt{x'^\top x'}}. \end{aligned}$$

In contrast, Euclidean distance measure would be

$$sim(x, x') = x^\top x'$$

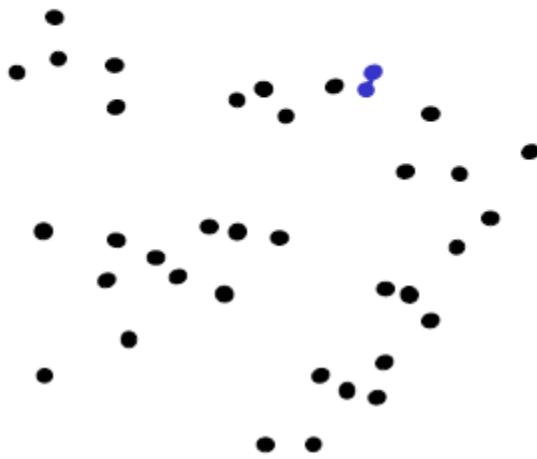
# Agglomerative clustering



1. Say "Every point is its own cluster"

Slide credit: Andrew Moore

# Agglomerative clustering

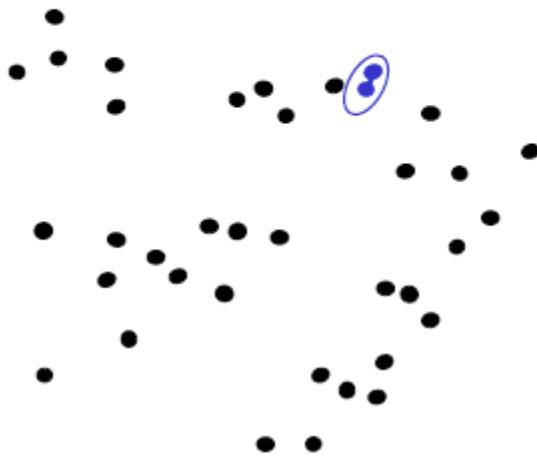


1. Say "Every point is its own cluster"
2. Find "most similar" pair of clusters



Slide credit: Andrew Moore

# Agglomerative clustering

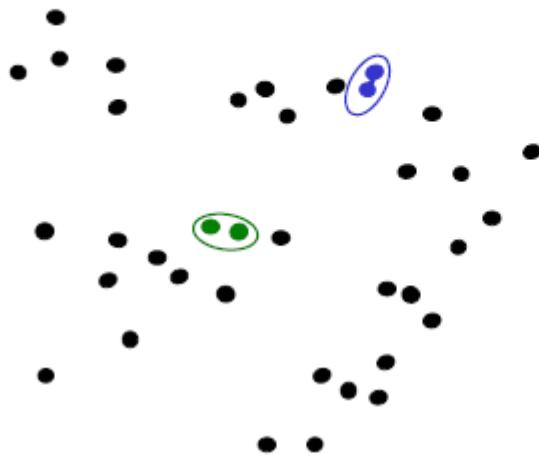


1. Say "Every point is its own cluster"
2. Find "most similar" pair of clusters
3. Merge it into a parent cluster



Slide credit: Andrew Moore

# Agglomerative clustering

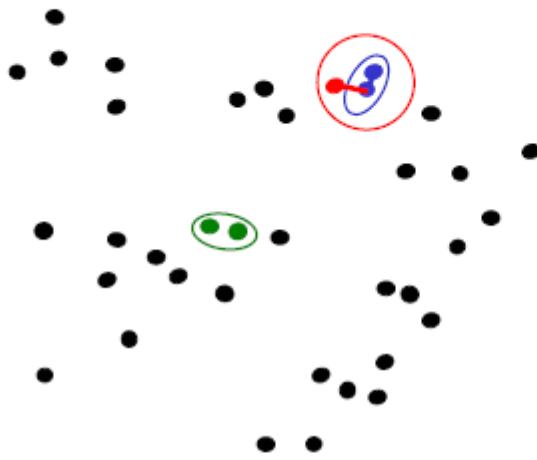


1. Say "Every point is its own cluster"
2. Find "most similar" pair of clusters
3. Merge it into a parent cluster
4. Repeat

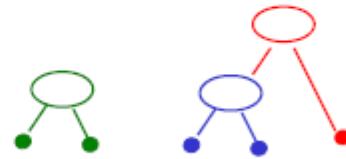


Slide credit: Andrew Moore

# Agglomerative clustering



1. Say "Every point is its own cluster"
2. Find "most similar" pair of clusters
3. Merge it into a parent cluster
4. Repeat

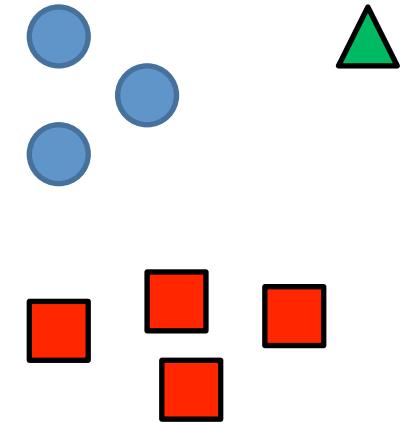


Slide credit: Andrew Moore

# Agglomerative clustering

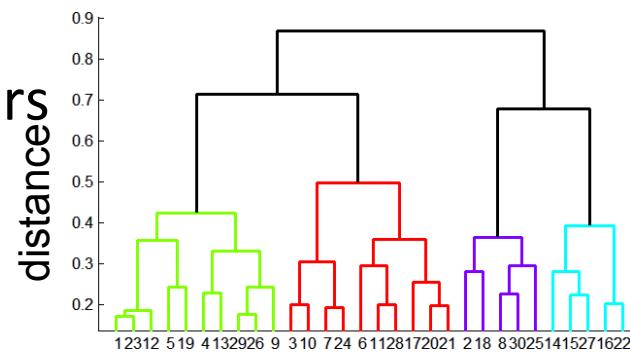
How to define cluster similarity?

- Average distance between points,
- maximum distance
- minimum distance
- Distance between means or medoids



How many clusters?

- Clustering creates a dendrogram (a tree)
- Threshold based on max number of clusters or based on distance between merges



technical  
note

# Agglomerative Hierarchical Clustering

1. Initially each item  $x_1, \dots, x_n$  is in its own cluster  $C_1, \dots, C_n$ .
2. Repeat until there is only one cluster left:
  3. Merge the nearest clusters, say  $C_i$  and  $C_j$ .

Different ways to define the “nearest clusters”:

- $d(C_i, C_j) = \min_{x \in C_i, x' \in C_j} d(x, x')$ . This is known as *single-linkage*. It is equivalent to the minimum spanning tree algorithm. One can set a threshold and stop clustering once the distance between clusters is above the threshold. Single-linkage tends to produce long and skinny clusters.
- $d(C_i, C_j) = \max_{x \in C_i, x' \in C_j} d(x, x')$ . This is known as *complete-linkage*. Clusters tend to be compact and roughly equal in diameter.
- $d(C_i, C_j) = \frac{\sum_{x \in C_i, x' \in C_j} d(x, x')}{|C_i| \cdot |C_j|}$ . This is the average distance between items. Somewhere between single-linkage and complete-linkage.
- and a million other ways you can think of ...

# Conclusions: Agglomerative Clustering

## Good

- Simple to implement, widespread application
- Clusters have adaptive shapes
- Provides a hierarchy of clusters

## Bad

- May have imbalanced clusters
- Still have to choose number of clusters or threshold
- Need to use an “ultrametric” to get a meaningful hierarchy

# What we have learned today?

- Introduction to segmentation and clustering
- Gestalt theory for perceptual grouping
- Agglomerative clustering