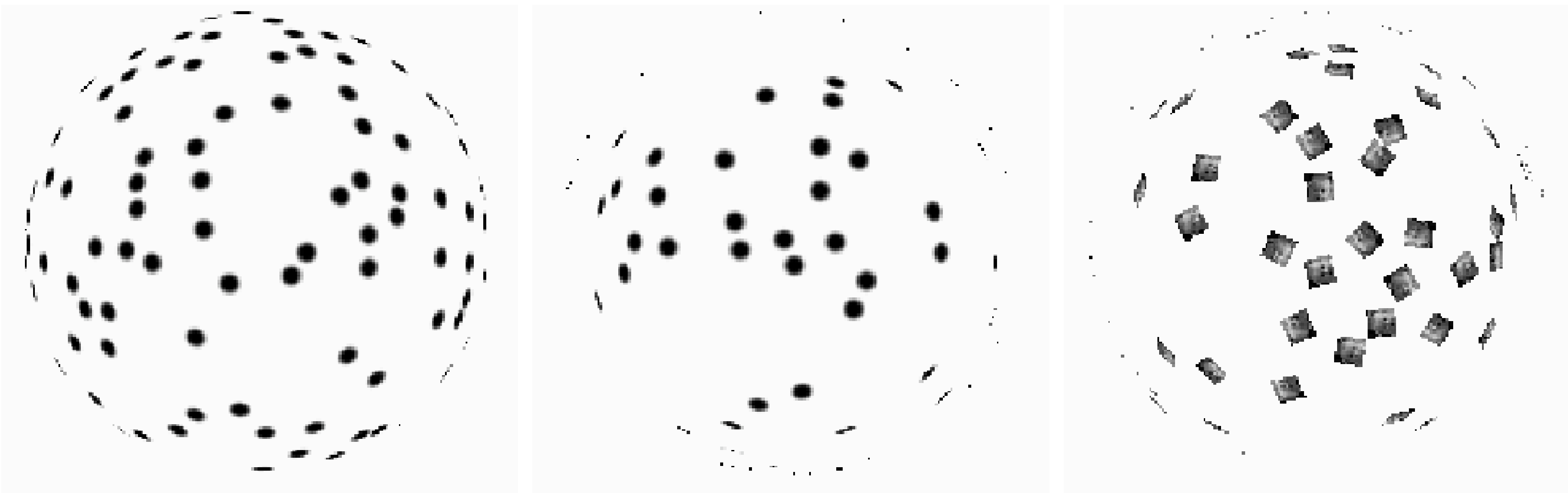
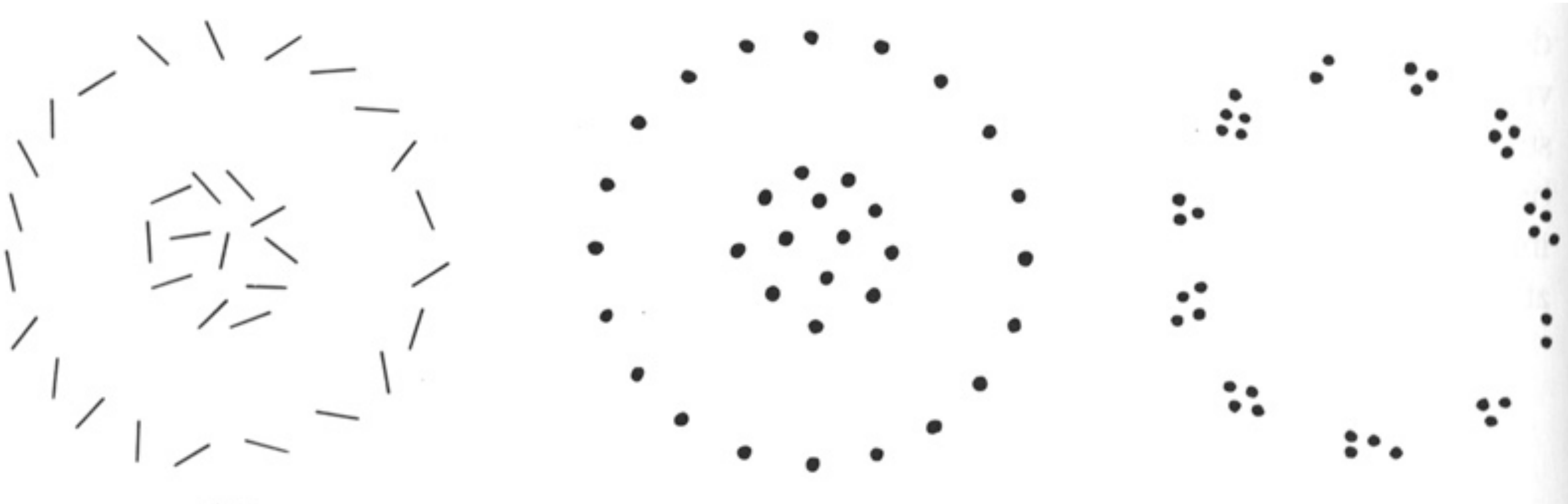
Segmentation and Grouping

- Motivation: not all information is evidence
- Obtain a compact representation
  - from an image/motion sequence/set of tokens
- Should support application
  - Broad theory is absent at present
- 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?

General ideas

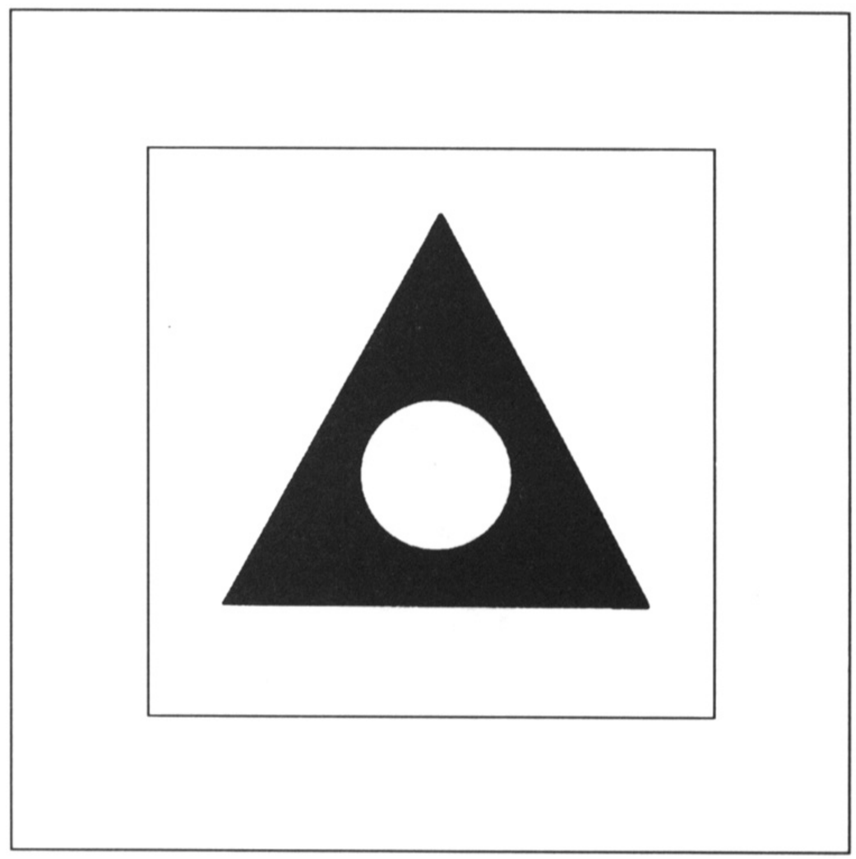
- tokens
  - whatever we need to group (pixels, points, surface elements, etc., etc.)
- top down segmentation
  - tokens belong together because they lie on the same object
- bottom up segmentation
  - tokens belong together because they are locally coherent
- These two are not mutually exclusive
  - e.g. symmetries, etc.

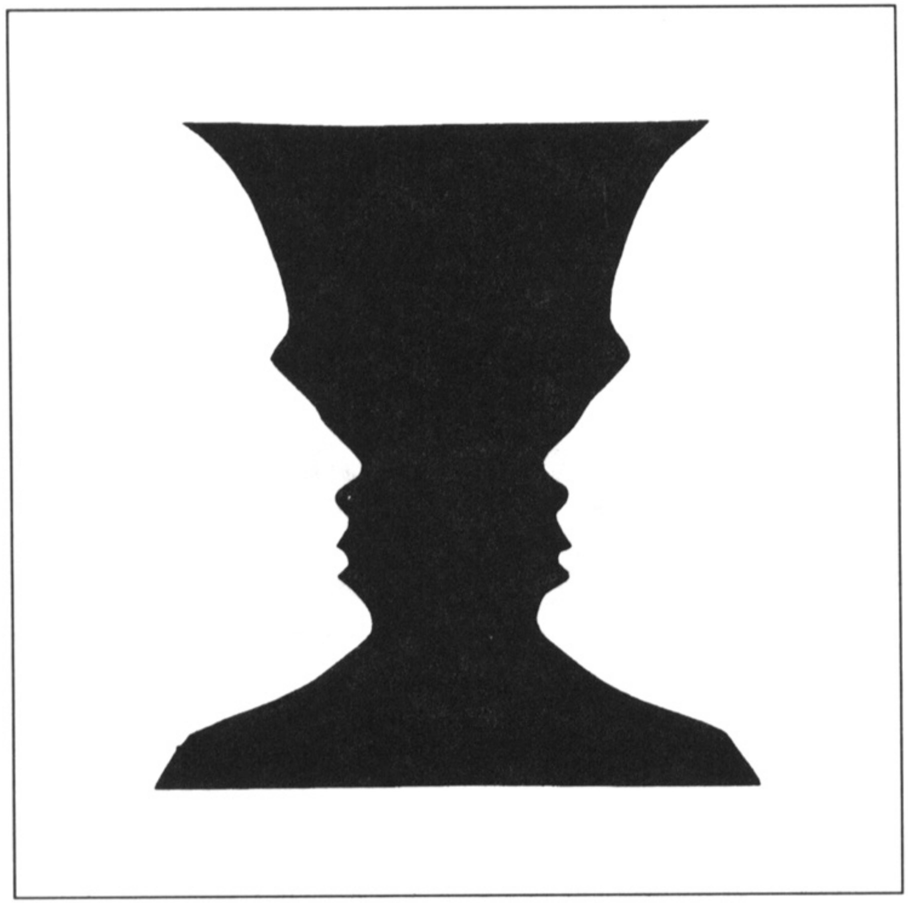


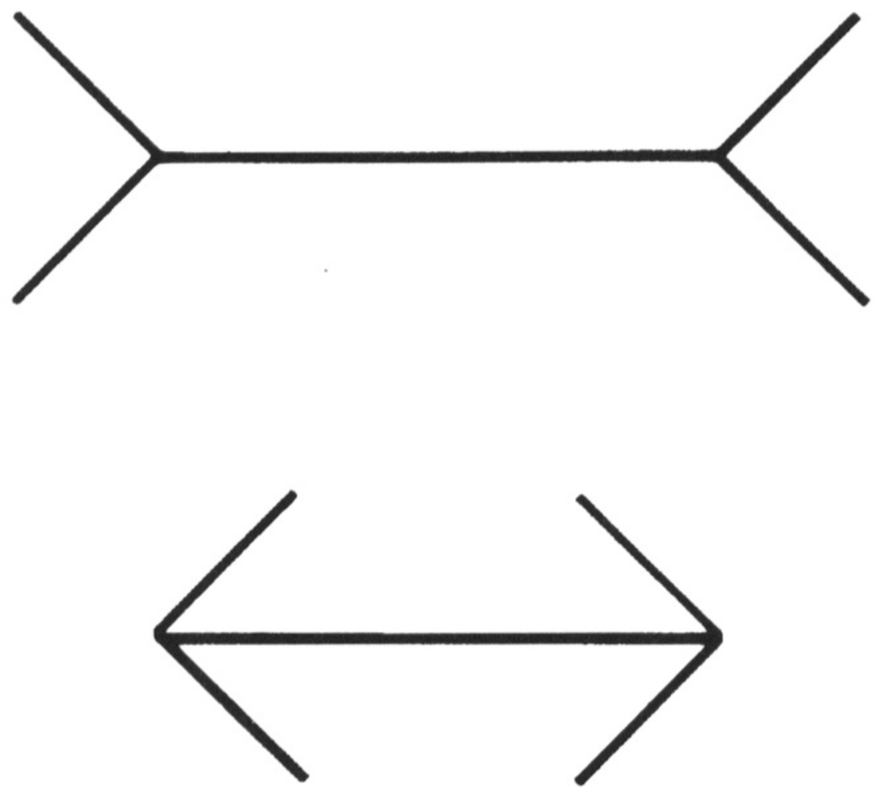


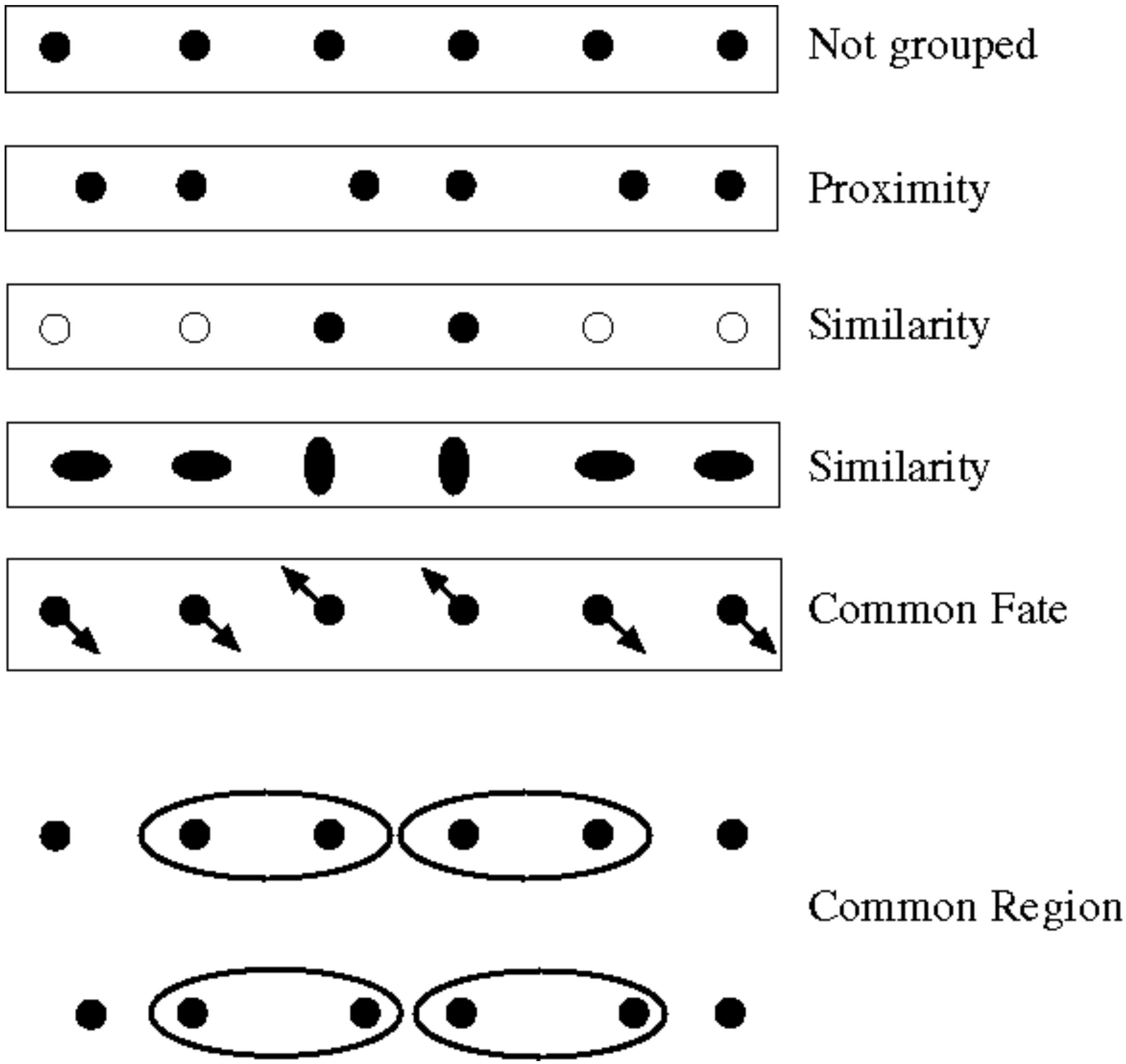
Basic ideas of grouping in humans

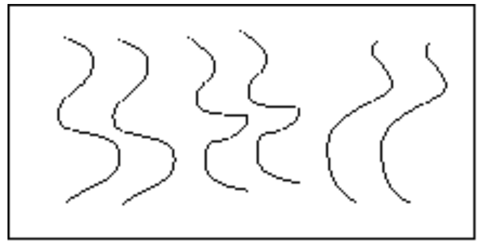
- Figure-ground discrimination
  - grouping can be seen in terms of allocating some elements to a figure, some to ground
  - impoverished theory
- Gestalt properties
  - elements in a collection of elements can have properties that result from relationships (Muller Lyer effect)
  - gestaltqualitat
  - A series of factors affect whether elements should be grouped together
  - Gestalt factors



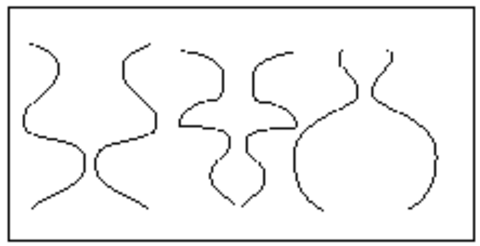




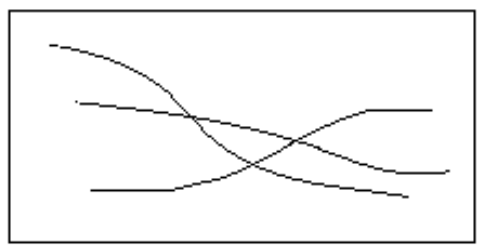

Not grouped
Proximity
Similarity
Similarity
Common Fate
Common Region



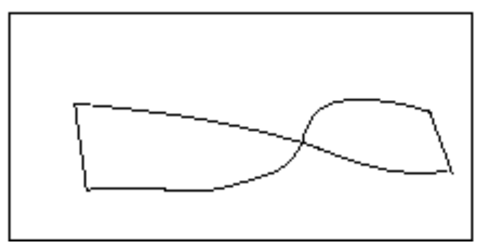
Parallelism



Symmetry



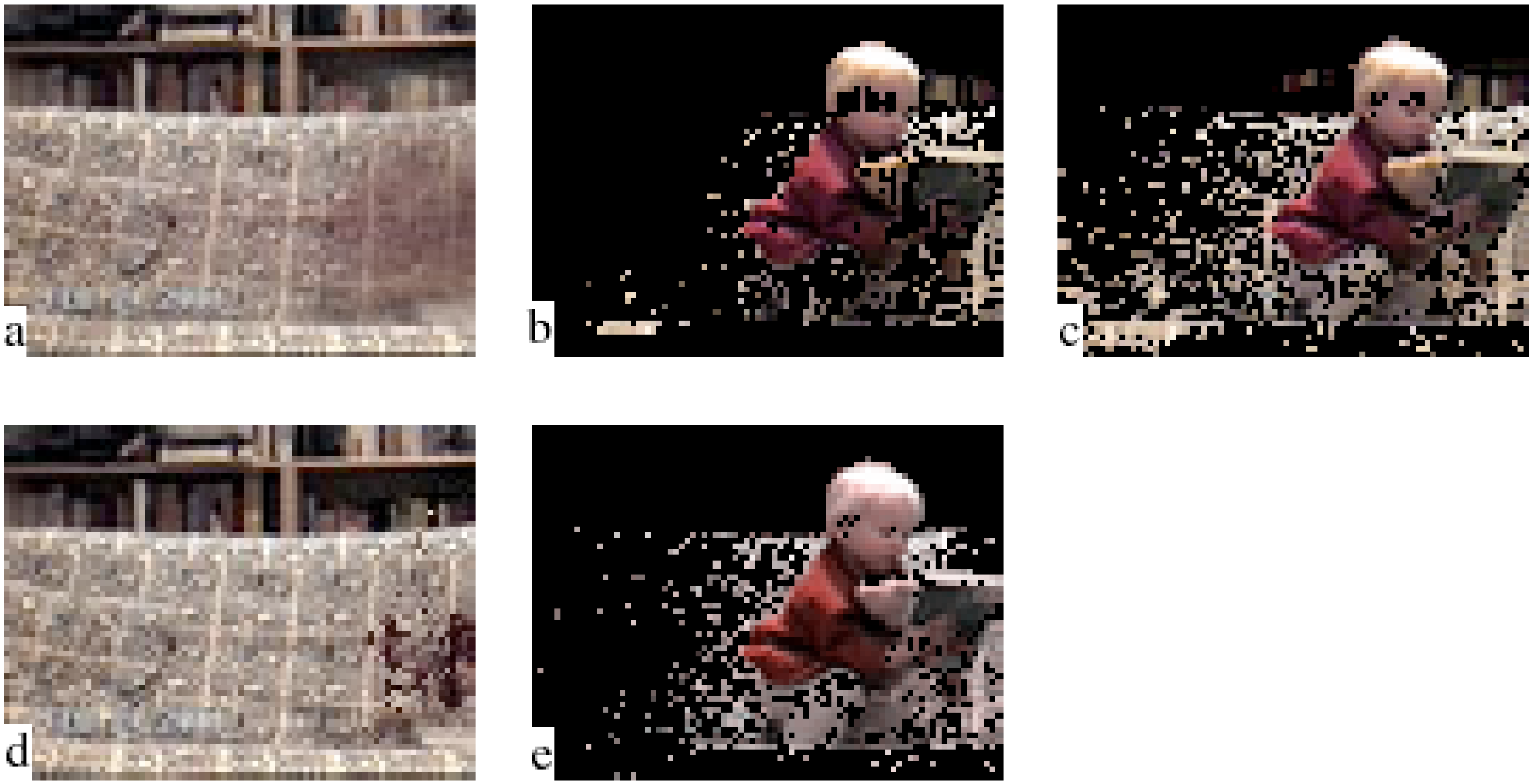
Continuity

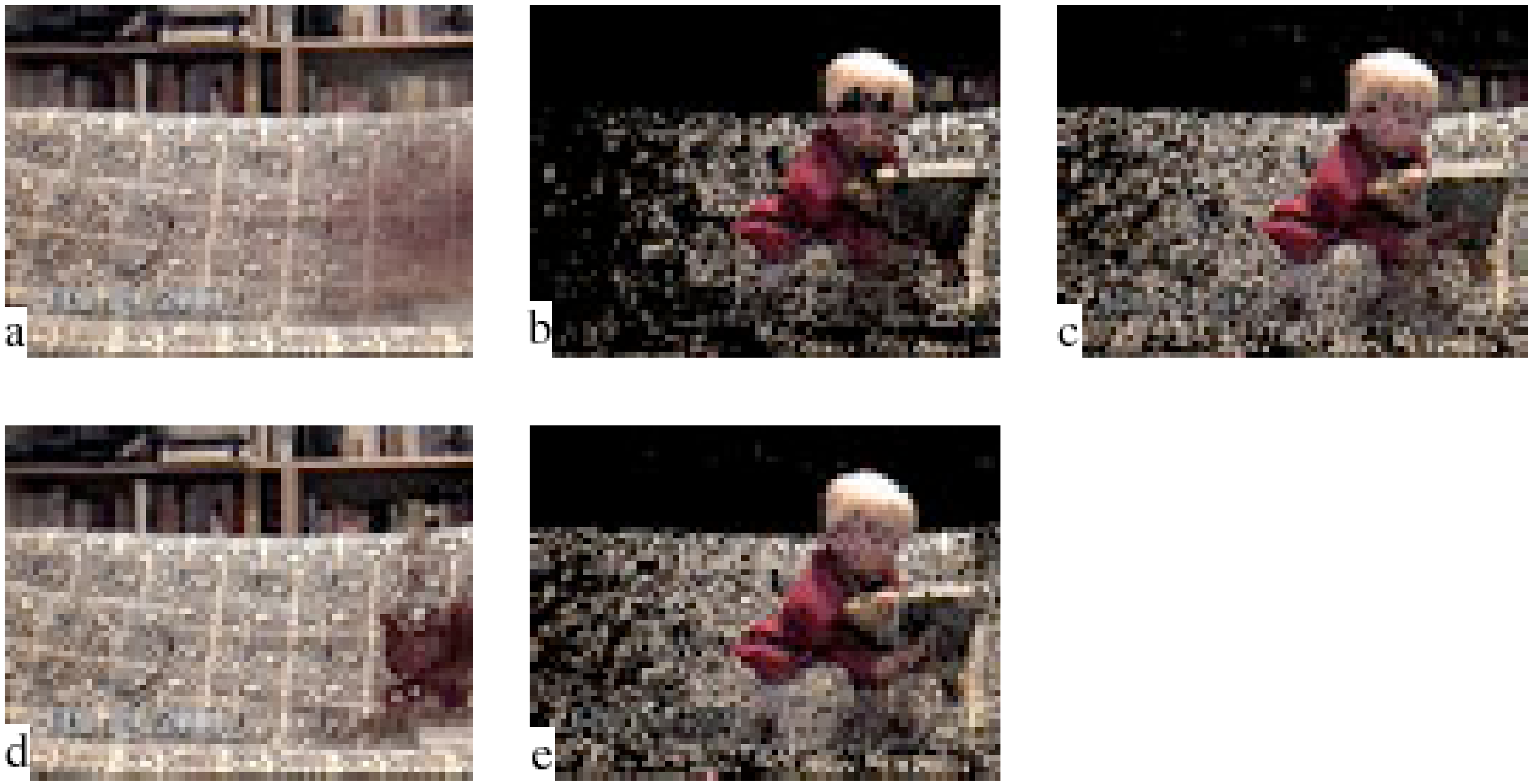


Closure

Technique: Background Subtraction

- If we know what the background looks like, it is easy to identify “interesting bits”
- Applications
  - Person in an office
  - Tracking cars on a road
  - surveillance
- Approach:
  - use a moving average to estimate background image
  - subtract from current frame
  - large absolute values are interesting pixels
  - trick: use morphological operations to clean up pixels



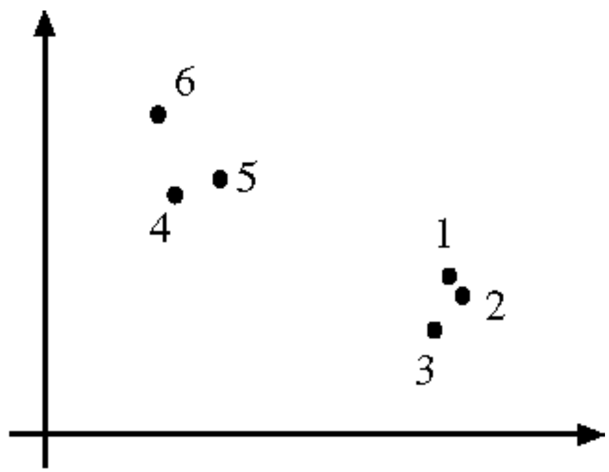

a
b
c
d
e

Technique: Shot Boundary Detection

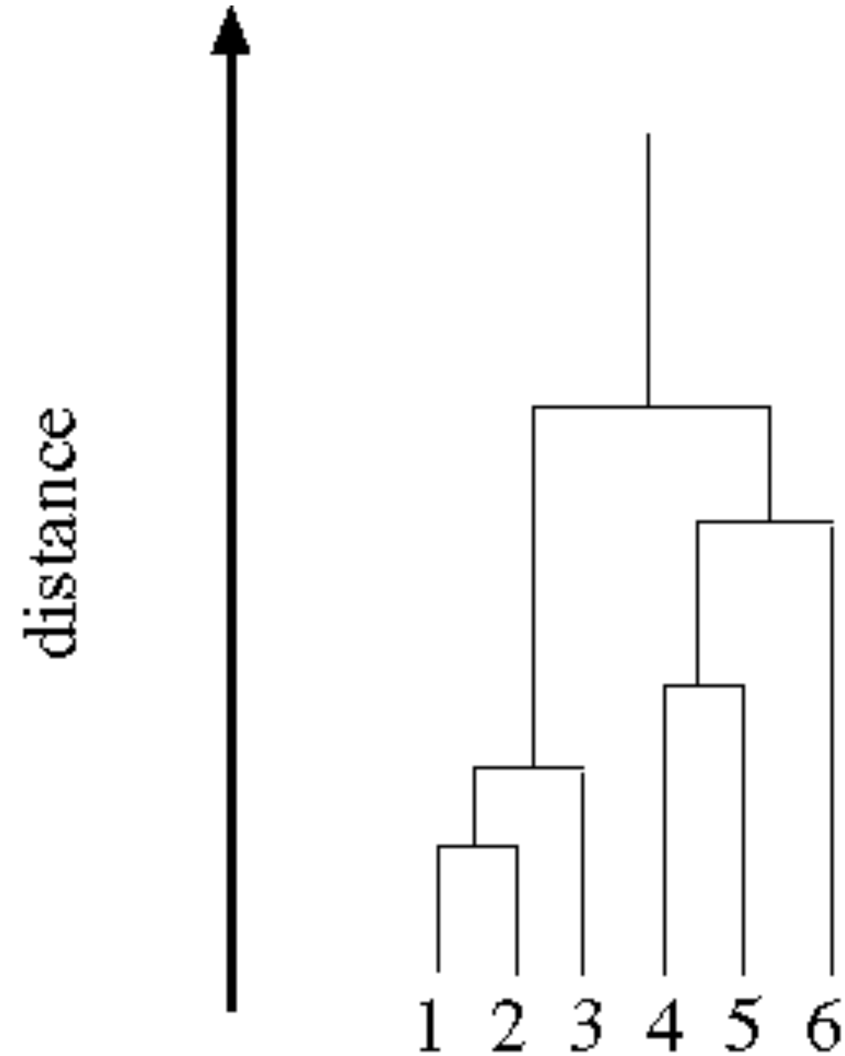
- Find the shots in a sequence of video
  - shot boundaries usually cause big differences between succeeding frames
- Strategy:
  - compute interframe distances
  - declare a boundary where these are big
- Possible distances
  - frame differences; histogram differences; block comparisons; edge differences
- Applications:
  - representation for movies, or video sequences
    - find shot boundaries
    - obtain "most representative" frame
      - supports search

Segmentation as clustering

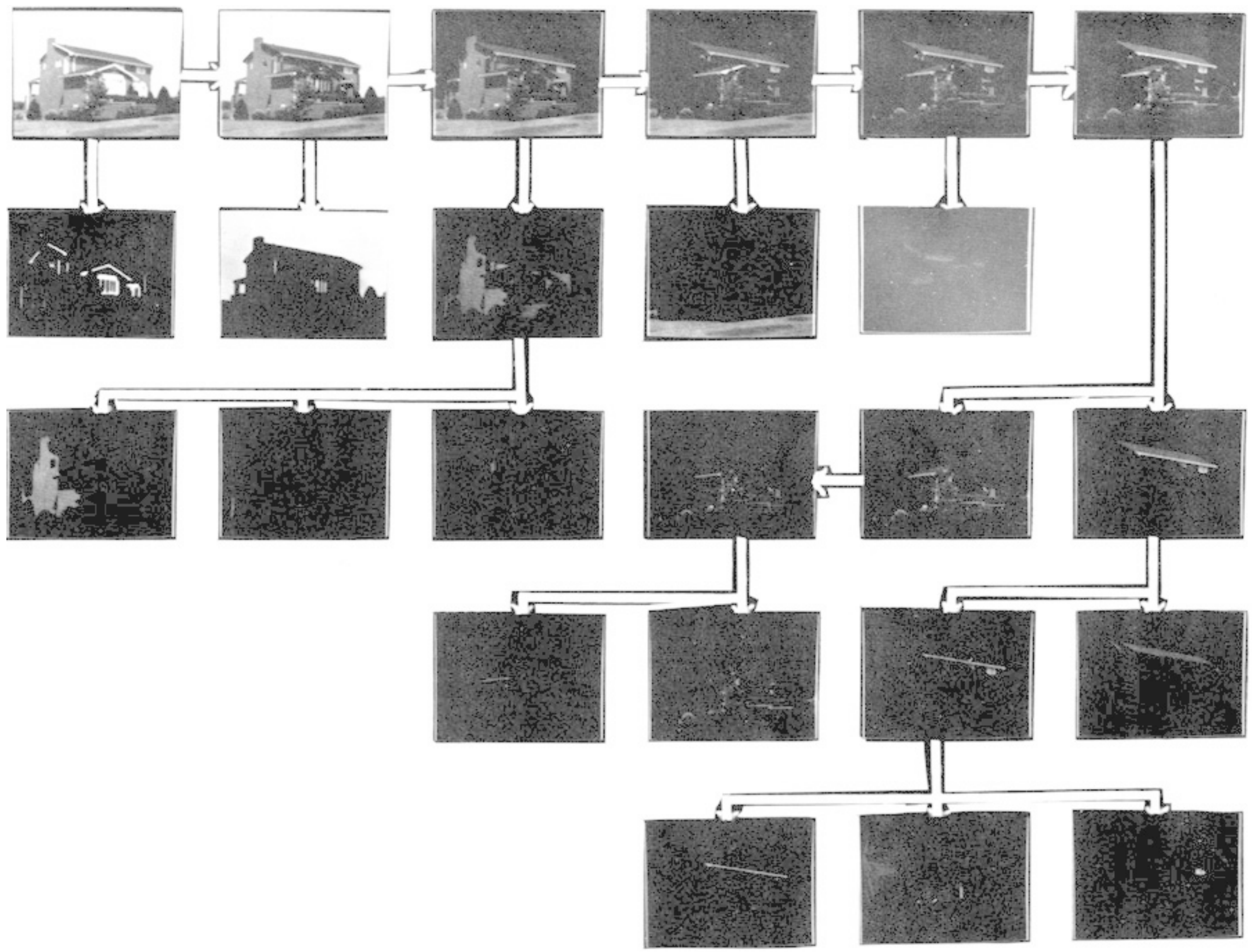
- Cluster together (pixels, tokens, etc.) that belong together
- Agglomerative clustering
    - attach closest to cluster it is closest to
    - repeat
- Divisive clustering
    - split cluster along best boundary
    - repeat
- Point-Cluster distance
    - single-link clustering
    - complete-link clustering
    - group-average clustering
- Dendrograms
    - yield a picture of output as clustering process continues



Dendrograms can inform choice of clusters



From Ohlander et al, 1978



From Ohlander et al, 1978

K-Means

- Choose a fixed number of clusters
- Choose cluster centers and point-cluster allocations to minimize error

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

- can't do this by search
  - there are too many possible allocations.
- Algorithm
  - fix cluster centers; allocate points to closest cluster
  - fix allocation; compute best cluster centers
  - x could be any set of features for which we can compute a distance (careful about scaling)

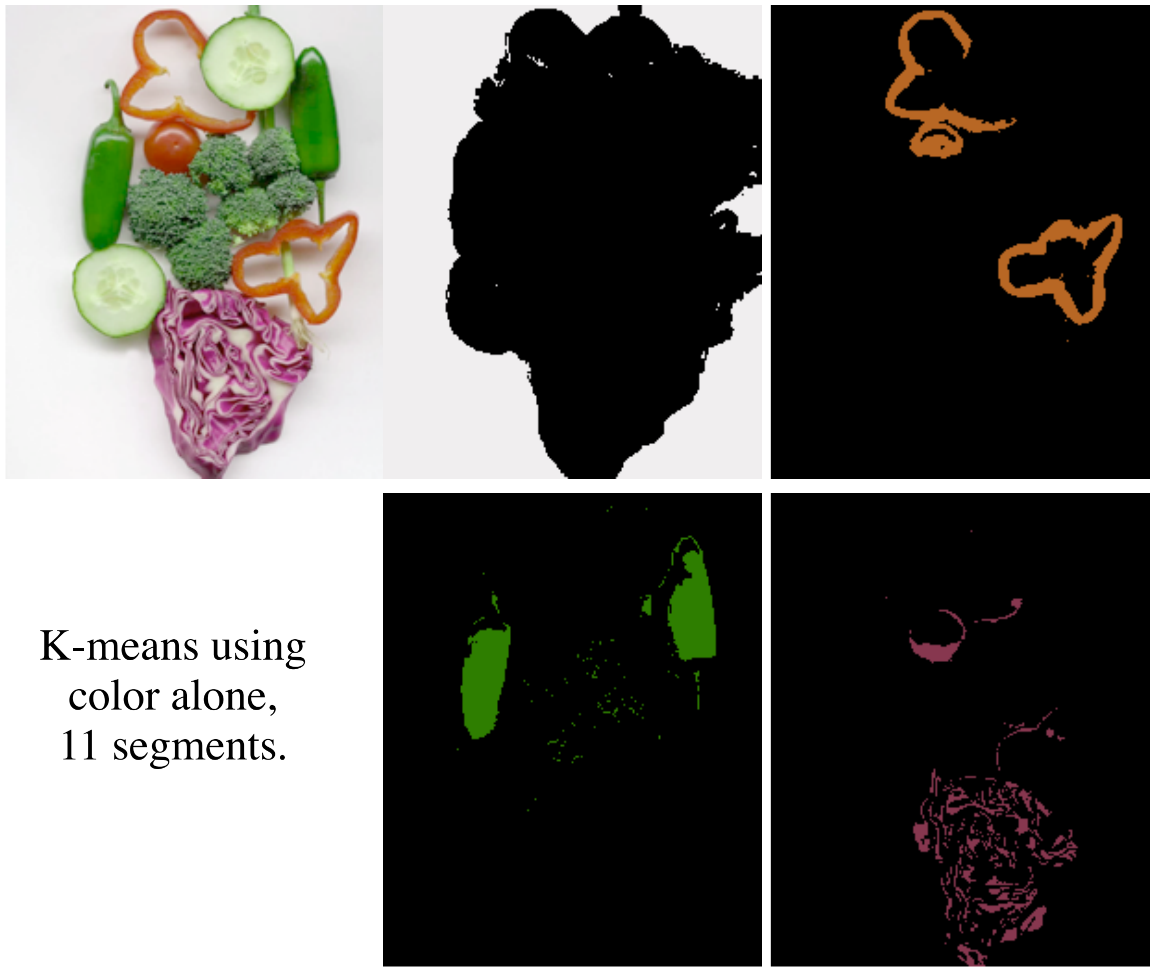
Image | Clusters on intensity | Clusters on color

K-means clustering using intensity alone and color alone

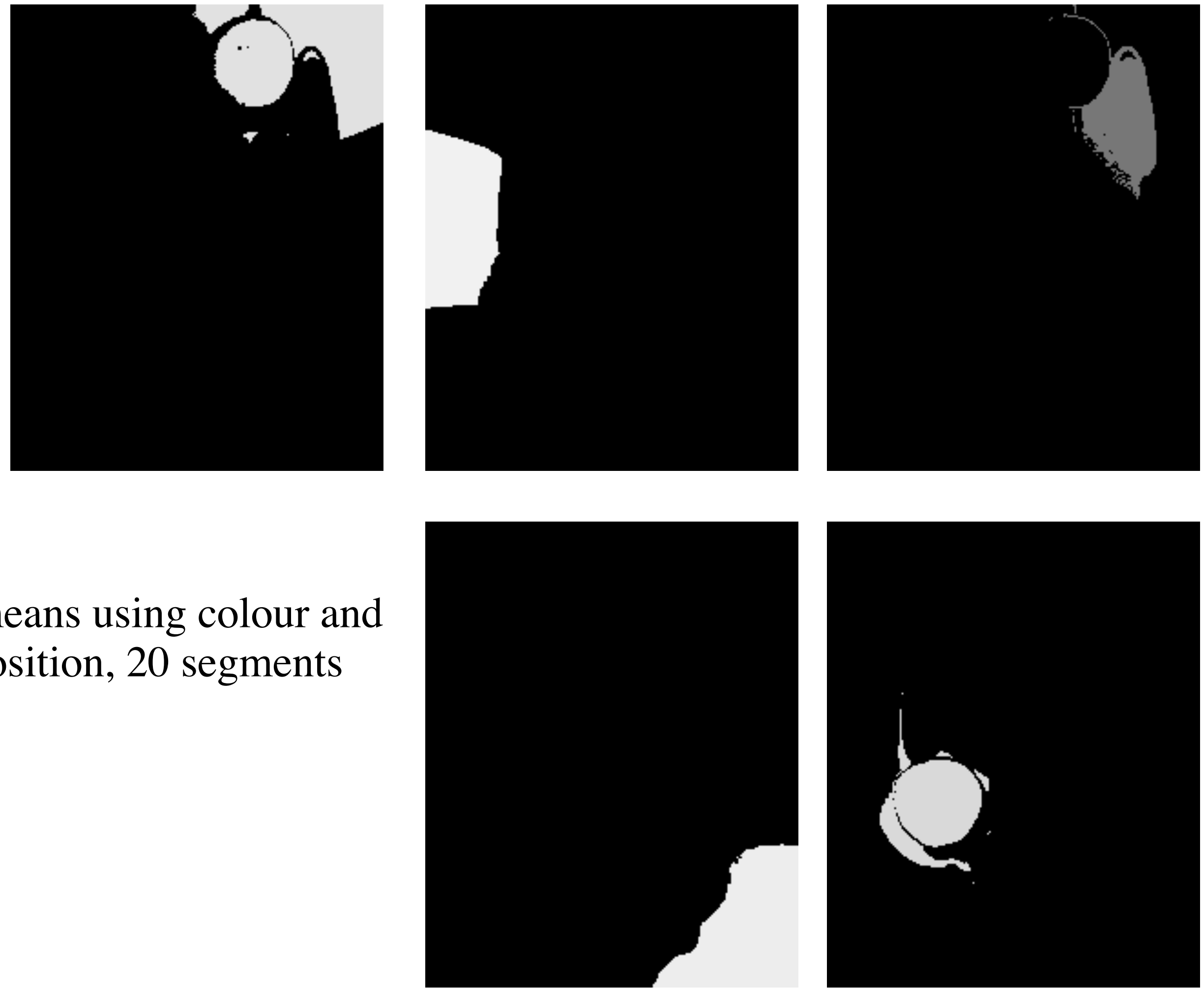
Image

Clusters on color

K-means using color alone, 11 segments



K-means using
color alone,
11 segments.



K-means using colour and position, 20 segments