Segmentation and Clustering

Segmentation and Clustering

- Segmentation:
  Divide image
  into regions
  of similar contents

- Clustering:
  Aggregate pixels
  into regions
  of similar contents

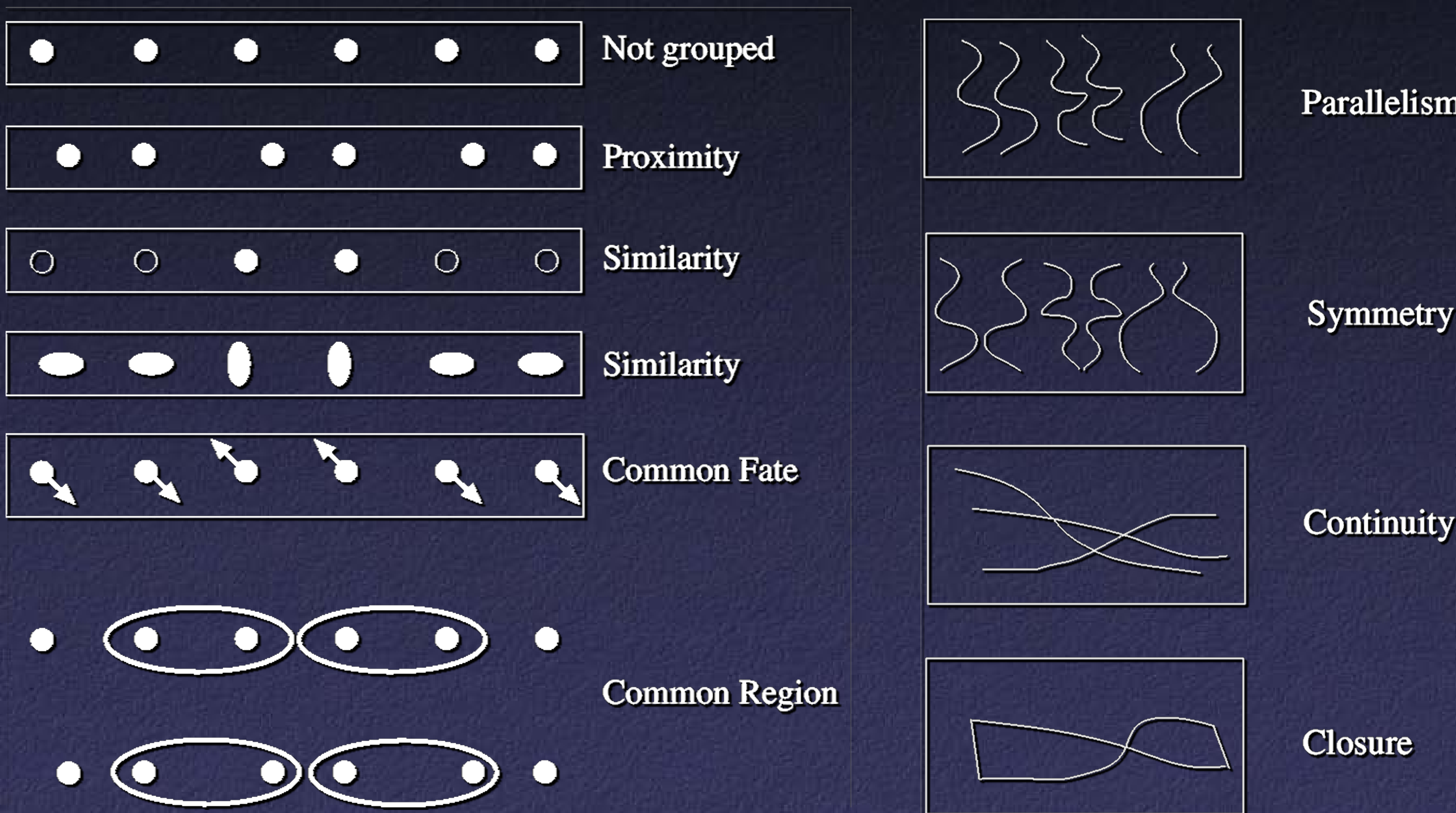
But Wait!

- We speak of segmenting foreground from background
- Segmenting out skin colors
- Segmenting out the moving person
- How do these relate to “similar regions”?

Segmentation and Clustering

- Defining regions
  - Should they be compact? Smooth boundary?
- Defining similarity
  - Color, texture, motion, …
- Defining similarity of regions
  - Minimum distance, mean, maximum

Grouping Cues



Segmentation and Clustering Applications

“Intelligent scissors”

Foreground / background segmentation

Finding skin-colored regions

Finding the moving objects

Finding the cars in a video sequence

Semantics

Segmentation and Clustering Applications

"Intelligent scissors"

Foreground / background segmentation

Finding skin-colored regions

Finding the moving objects

Finding the cars in a video sequence

Statistics

Templates

Clustering Based on Color

- Let's make a few concrete choices:
  - Arbitrary regions
  - Similarity based on color only
  - Similarity of regions = distance between mean colors

Simple Agglomerative Clustering

- Start with each pixel in its own cluster
- Iterate:
  - Find pair of clusters with smallest inter-cluster distance
  - Merge
- Stopping threshold
- "Superpixels": stop clustering early, pass result to more complex algorithms

Simple Divisive Clustering

- Start with whole image in one cluster
- Iterate:
  - Find cluster with largest intra-cluster variation
  - Split into two pieces that yield largest inter-cluster distance
- Stopping threshold

Difficulties with Simple Clustering

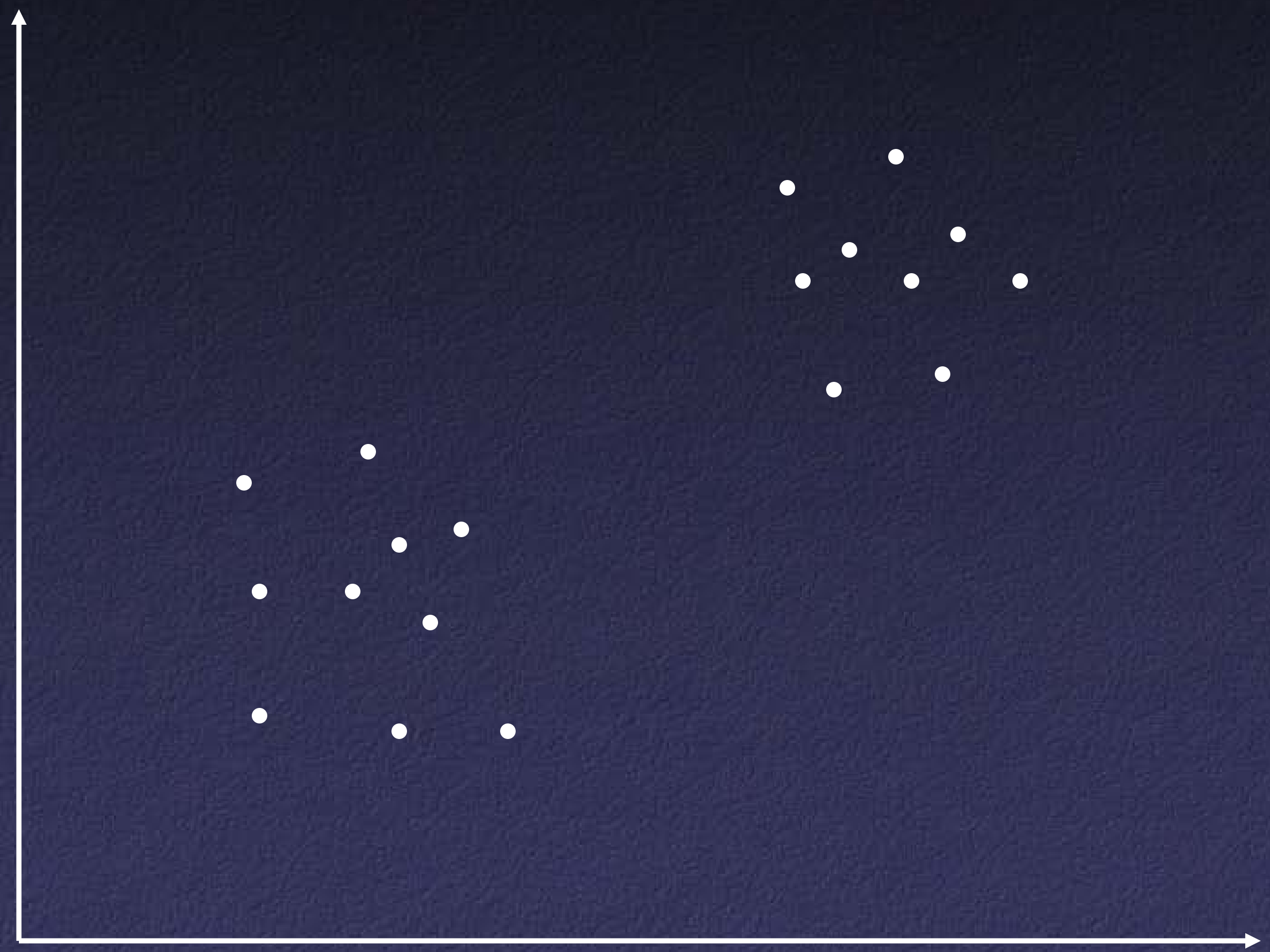
- Many possibilities at each iteration
- Computing distance between clusters or optimal split expensive
- Heuristics to speed this up:
  - For agglomerative clustering, approximate each cluster by average for distance computations
  - For divisive clustering, use summary (histogram) of a region to compute split

k-means Clustering

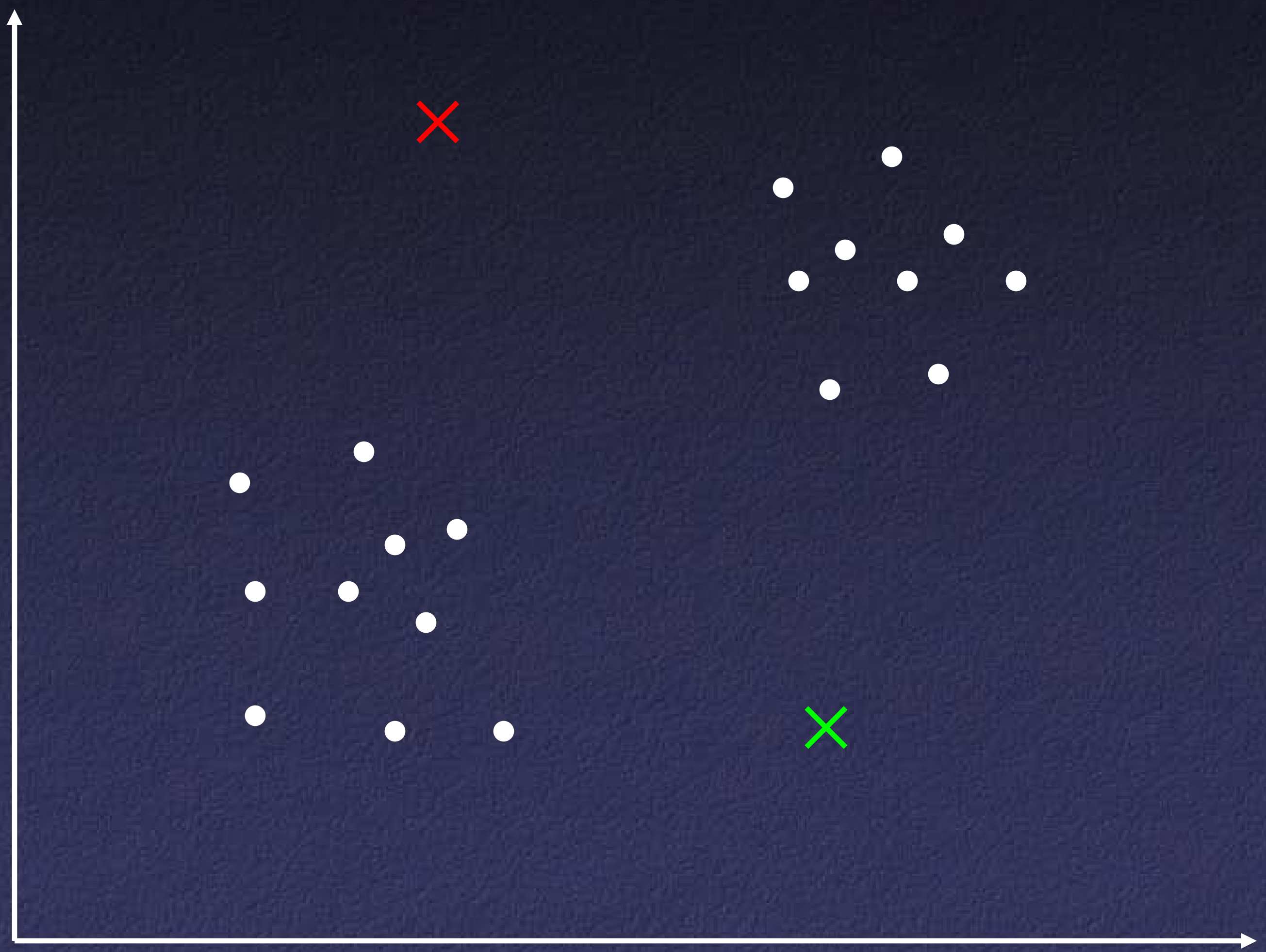
- Instead of merging or splitting, start out with the clusters and move them around

1. Pick number of clusters k
2. Randomly scatter k "cluster centers" in color space
3. Repeat:
    a. Assign each data point to its closest cluster center
    b. Move each cluster center to the mean of the points assigned to it

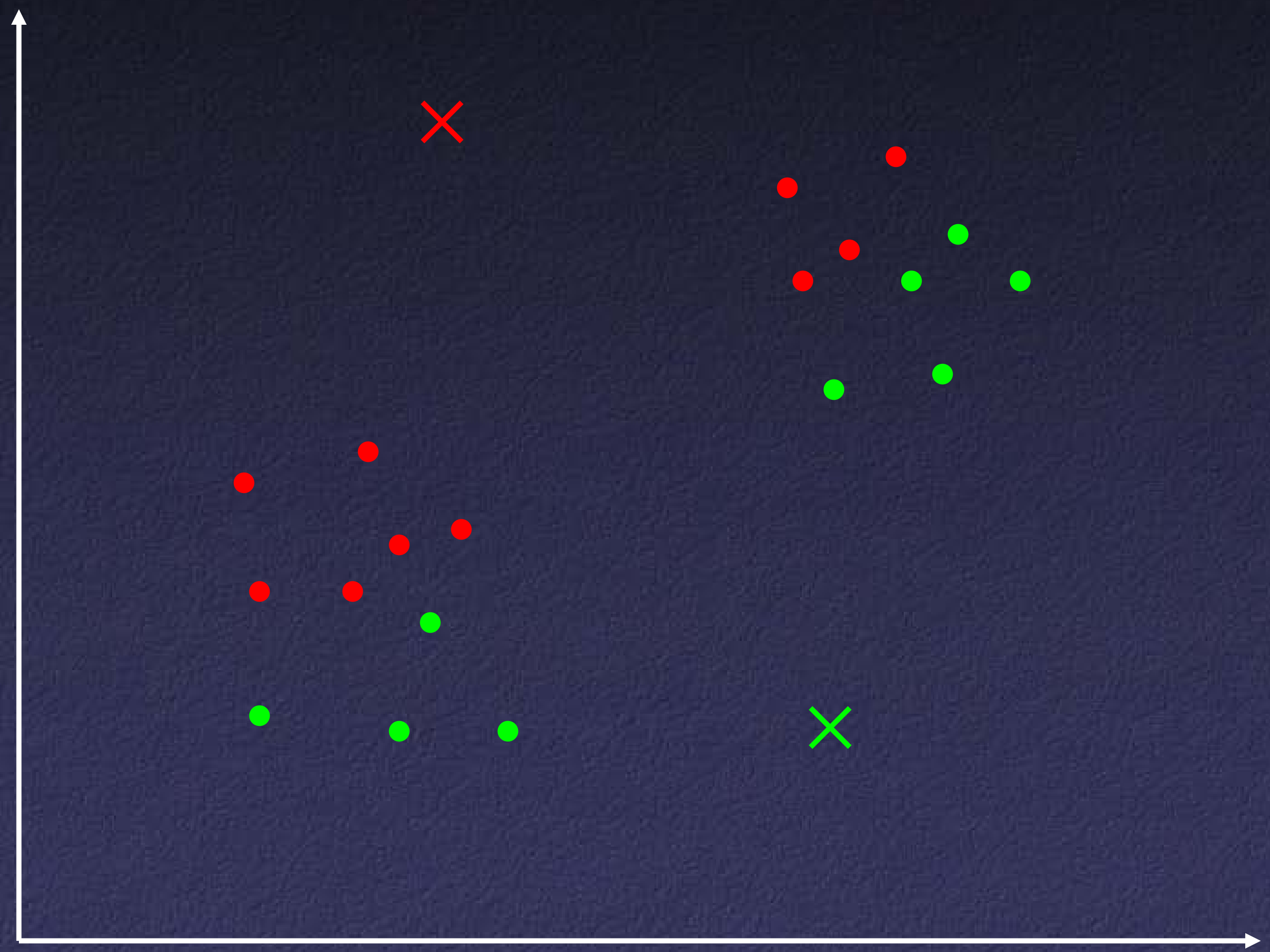
k-means Clustering



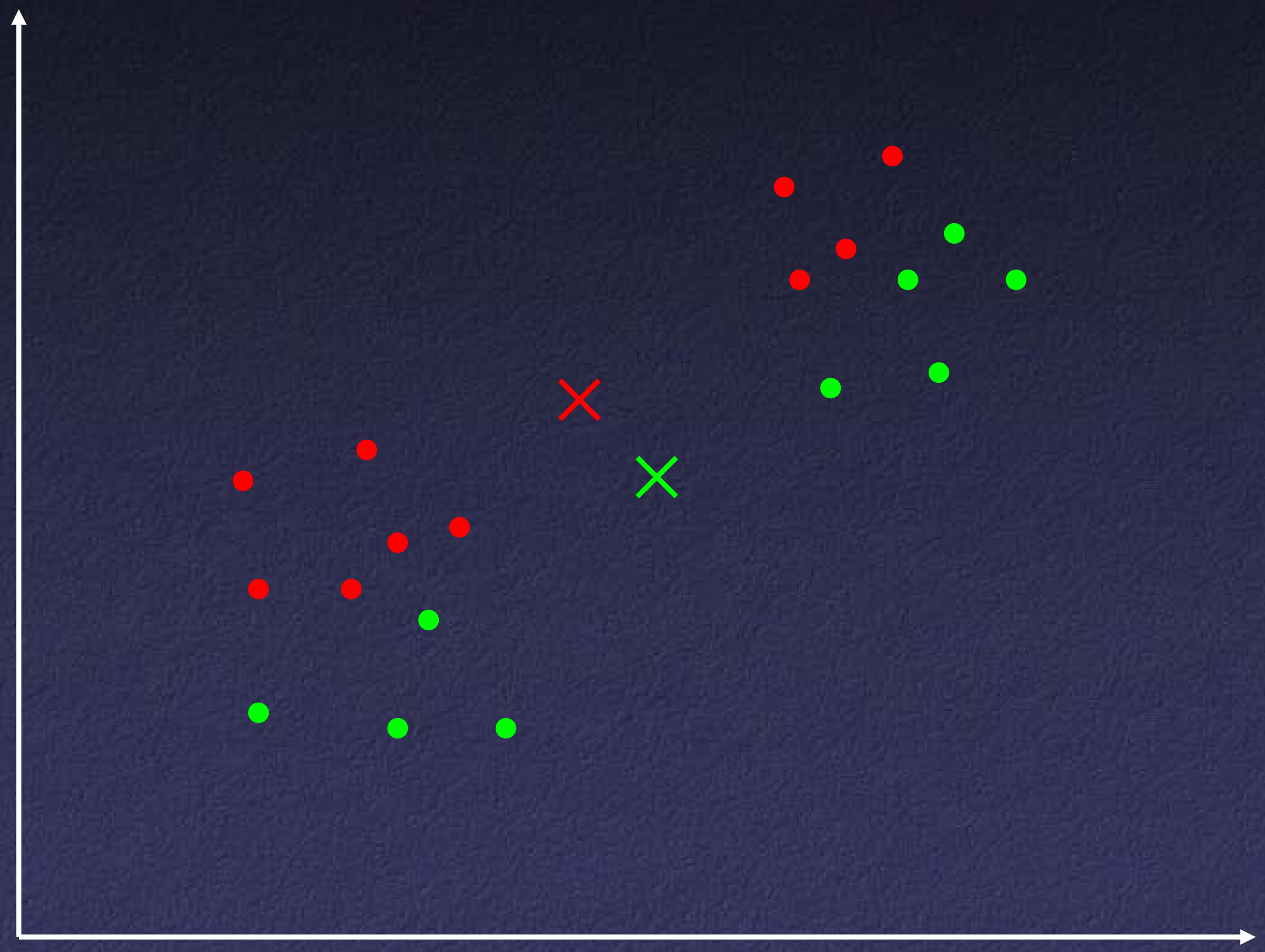
k-means Clustering



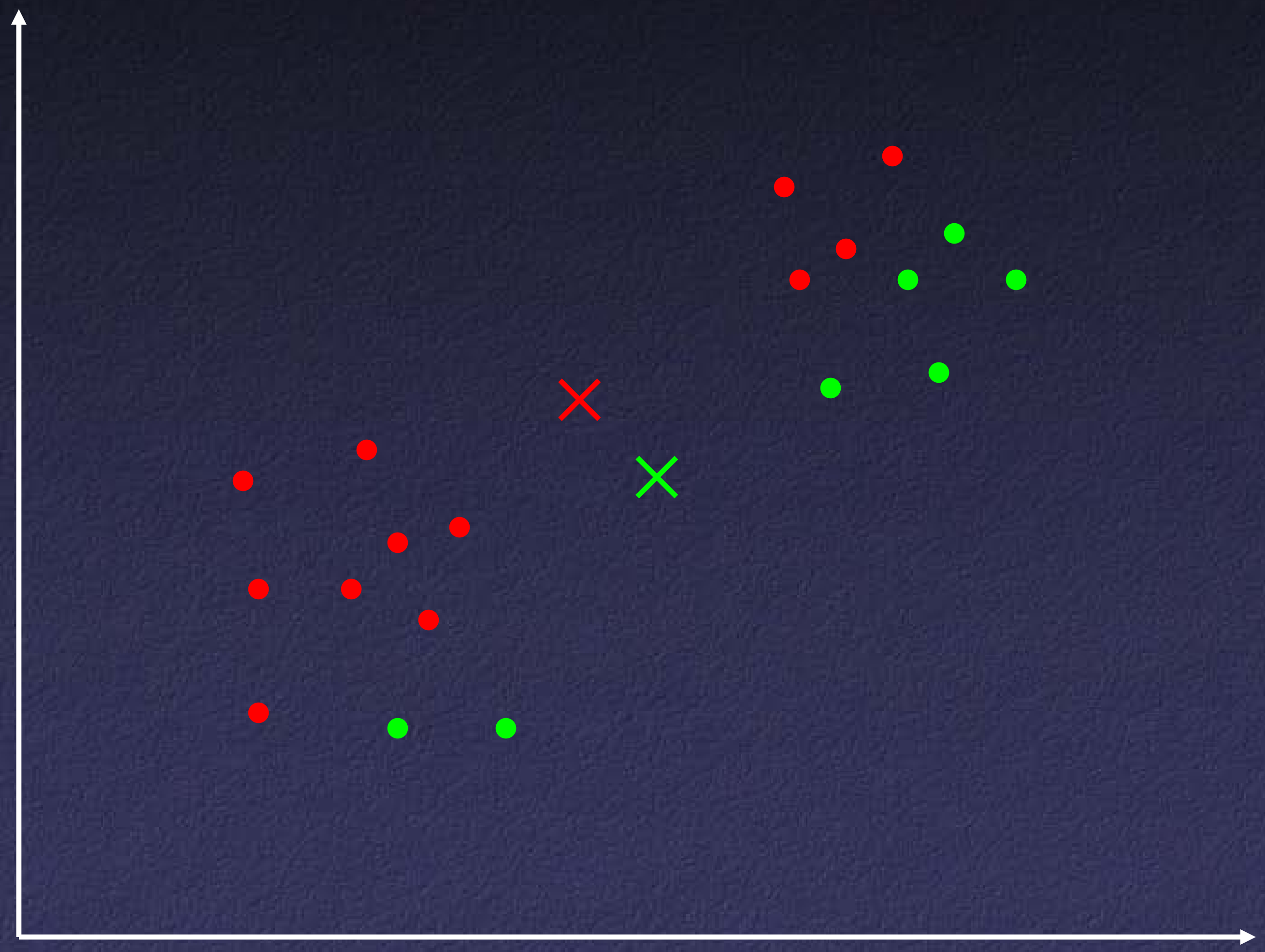
k-means Clustering



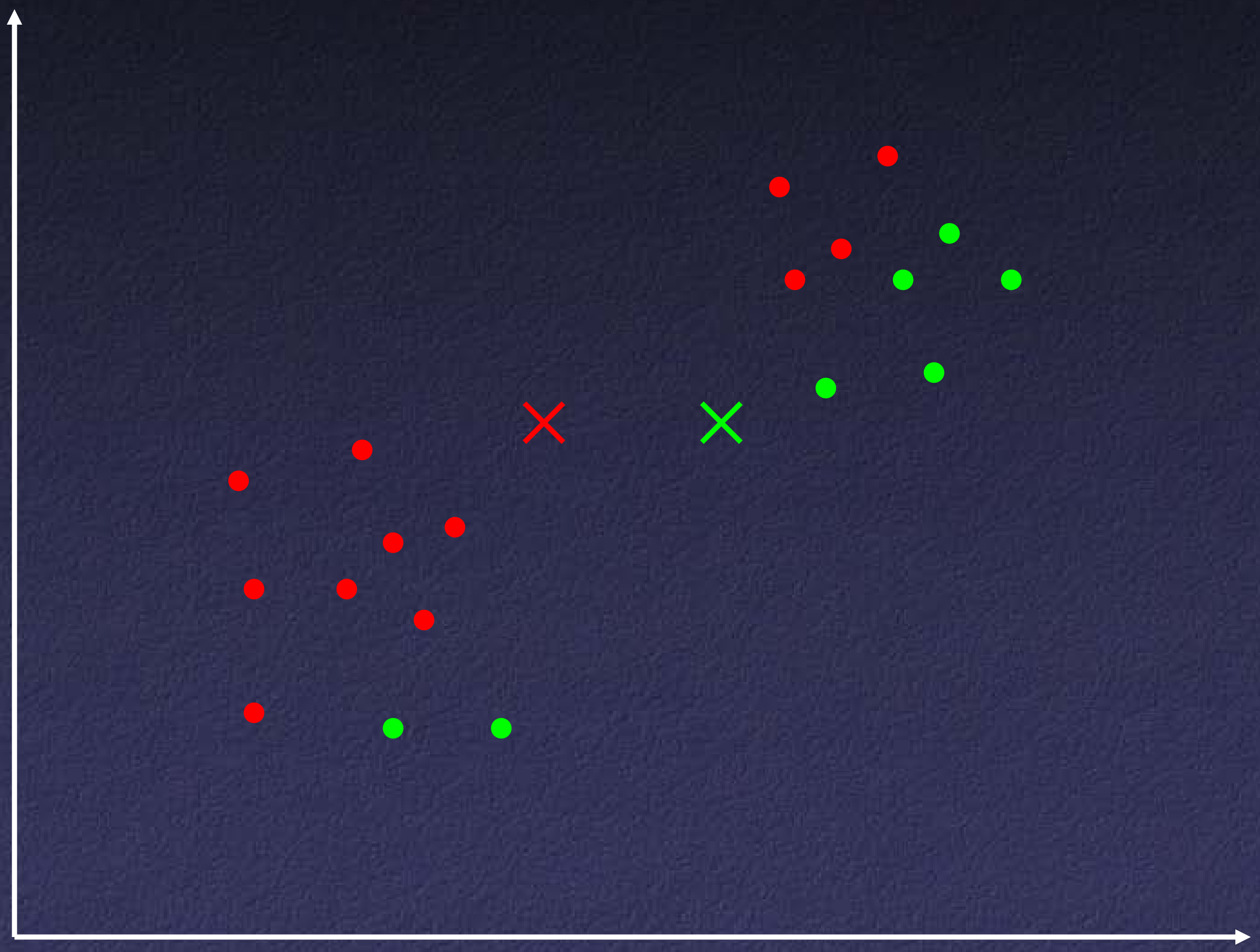
k-means Clustering



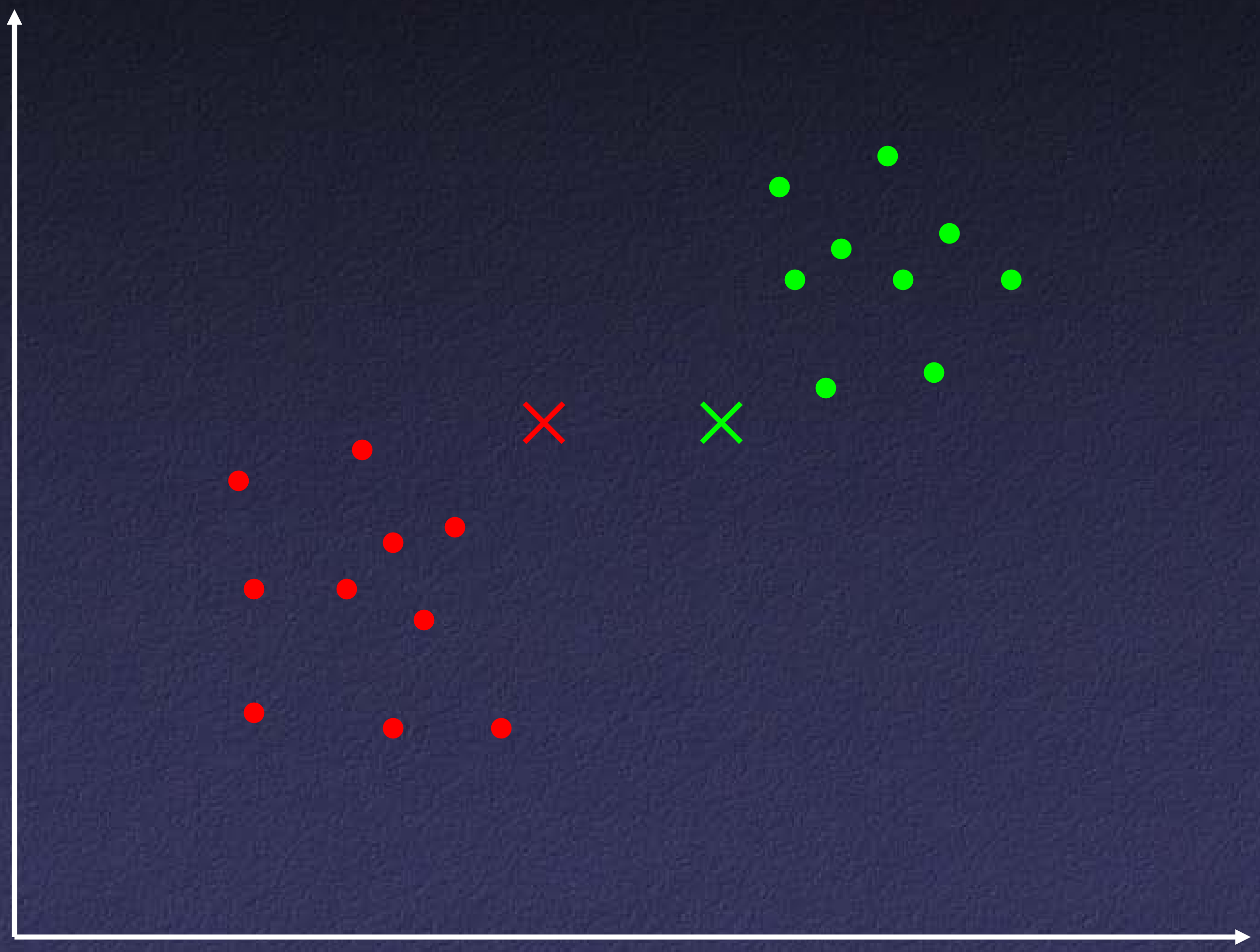
k-means Clustering



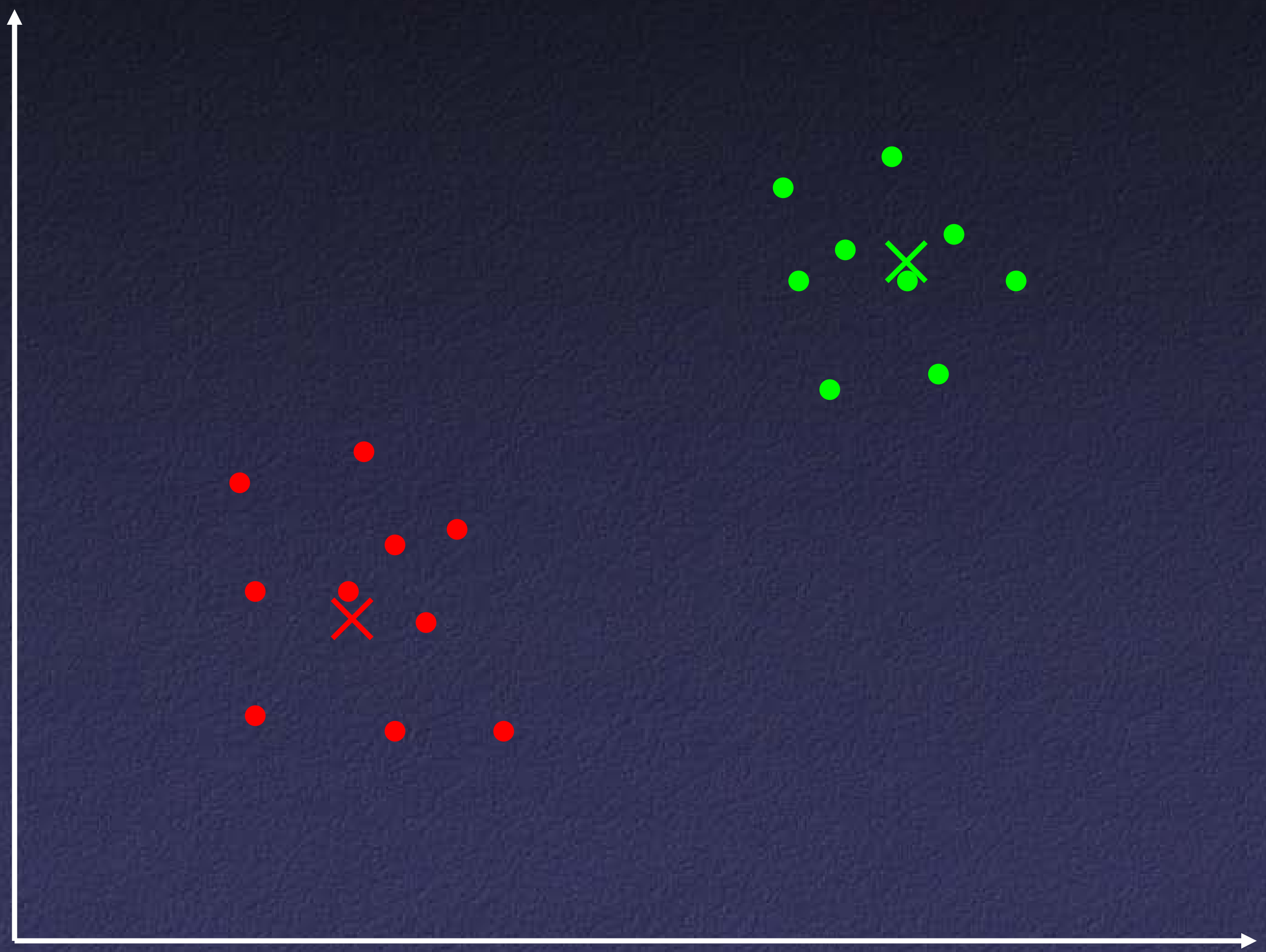
k-means Clustering



k-means Clustering



k-means Clustering



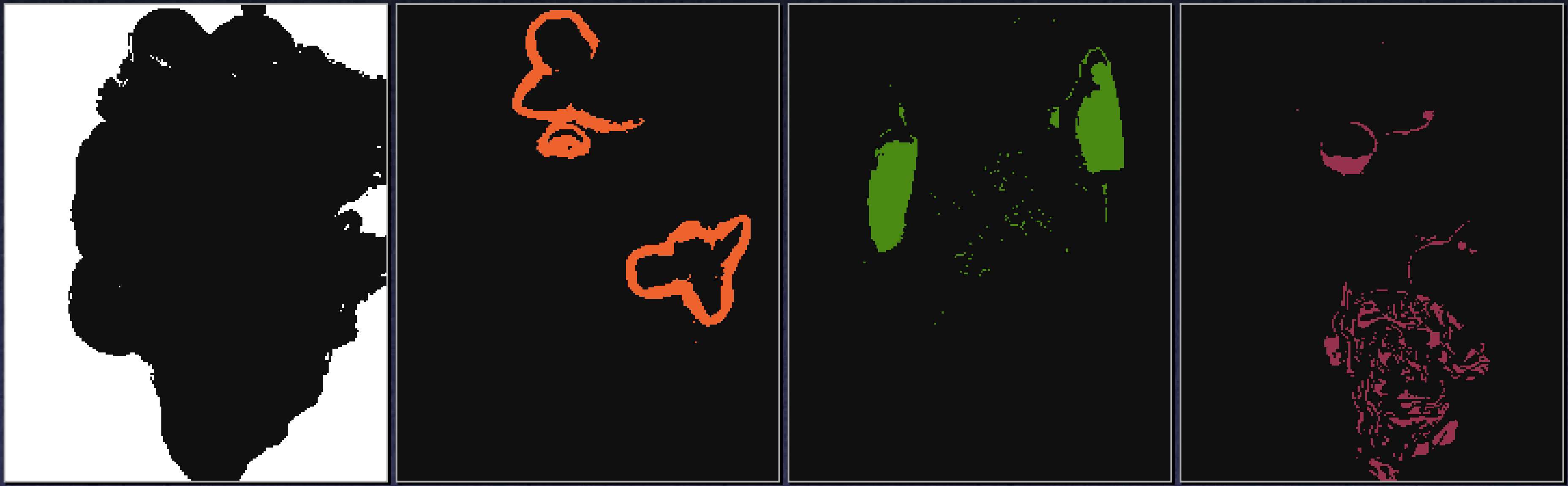
Results of Clustering

Original Image

k-means, $k=5$

k-means, $k=11$

Results of Clustering



Sample clusters with *k*-means clustering based on color

Other Distance Measures

- Suppose we want to have compact regions
- New feature space: 5D
  (2 spatial coordinates, 3 color components)
- Points close in this space are close both in color and in actual proximity

Results of Clustering

Sample clusters with *k*-means clustering
based on color and distance

Other Distance Measures

- Problem with simple Euclidean distance: what if coordinates range from 0-1000 but colors only range from 0-255?
  - Depending on how things are scaled, gives different weight to different kinds of data
- Weighted Euclidean distance: adjust weights to emphasize different dimensions

$$\|x - y\|^2 = \sum c_i (x_i - y_i)^2$$

Mahalanobis Distance

- Automatically assign weights based on actual variation in the data

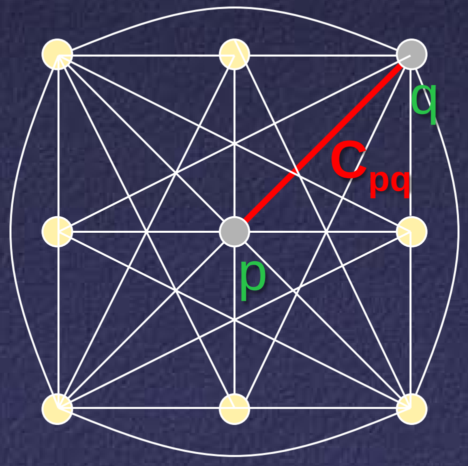
$$\left\|\vec{x}-\vec{y}\right\|^{2}=\left(\vec{x}-\vec{y}\right)^{\mathrm{T}}\mathbf{C}^{-1}(\vec{x}-\vec{y})$$

  where C is covariance matrix of all points

- Gives each dimension “equal” weight
- Also accounts for correlations between different dimensions

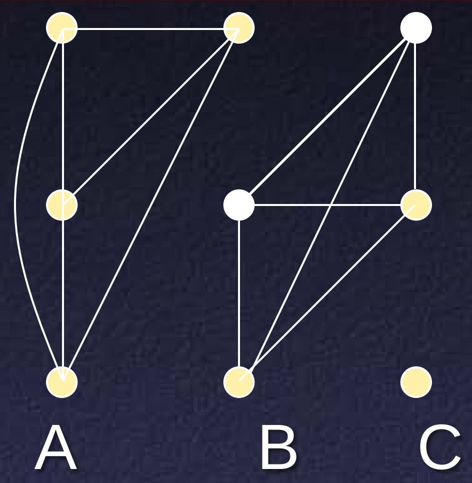
Segmentation Based on Graph Cuts

- Create weighted graph:
  - Nodes = pixels in image
  - Edge between each pair of nodes
  - Edge weight = similarity (intensity, color, texture, etc.)



[Based on slide by S. Seitz]

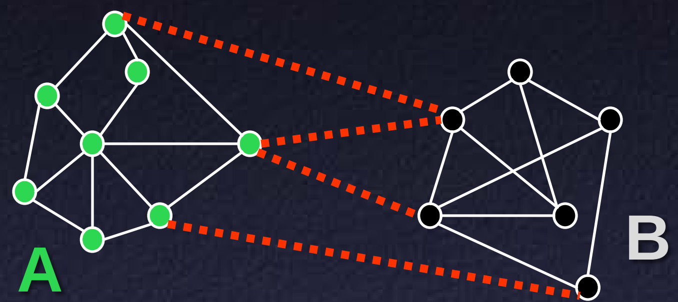
Segmentation Based on Graph Cuts



- Partition into disconnected segments
- Easiest to break links that have low cost (low similarity)
  - similar pixels should be in the same segments
  - dissimilar pixels should be in different segments

[Based on slide by S. Seitz]

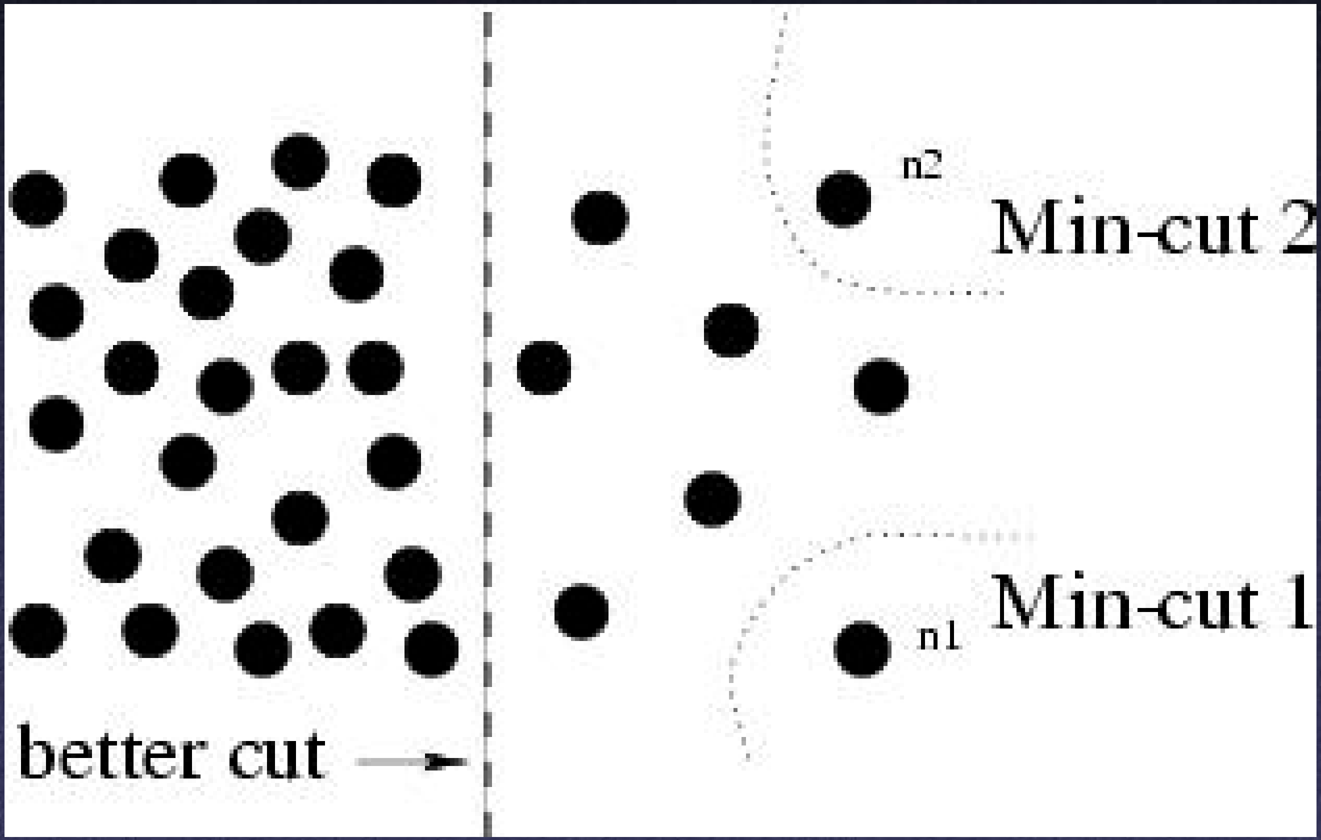
Cuts in a Graph



- Link Cut
  - set of links whose removal makes a graph disconnected
  - cost = sum of costs of all edges
- Min-cut
  - fast (polynomial-time) algorithm
  - gives segmentation

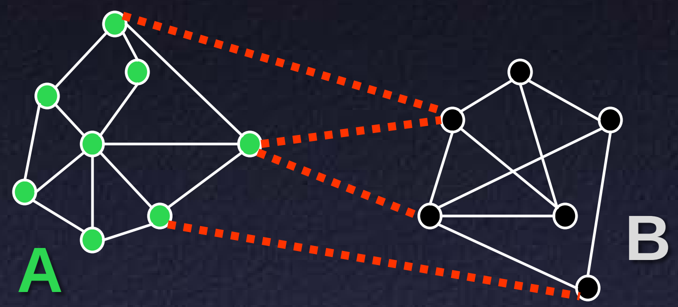
[Based on slide by S. Seitz]

But Min Cut Is Not Always the Best Cut...



[Based on slide by S. Seitz]

Cuts in a Graph



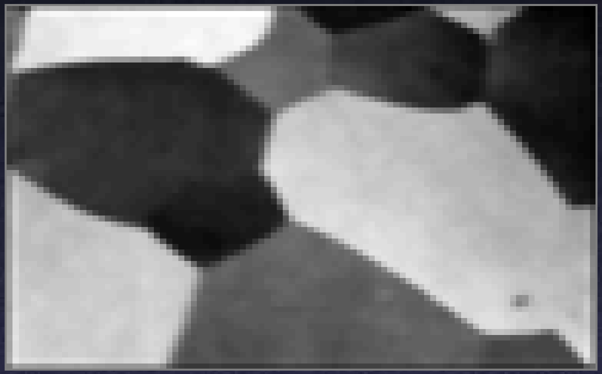
- Normalized Cut
  - removes penalty for large segments

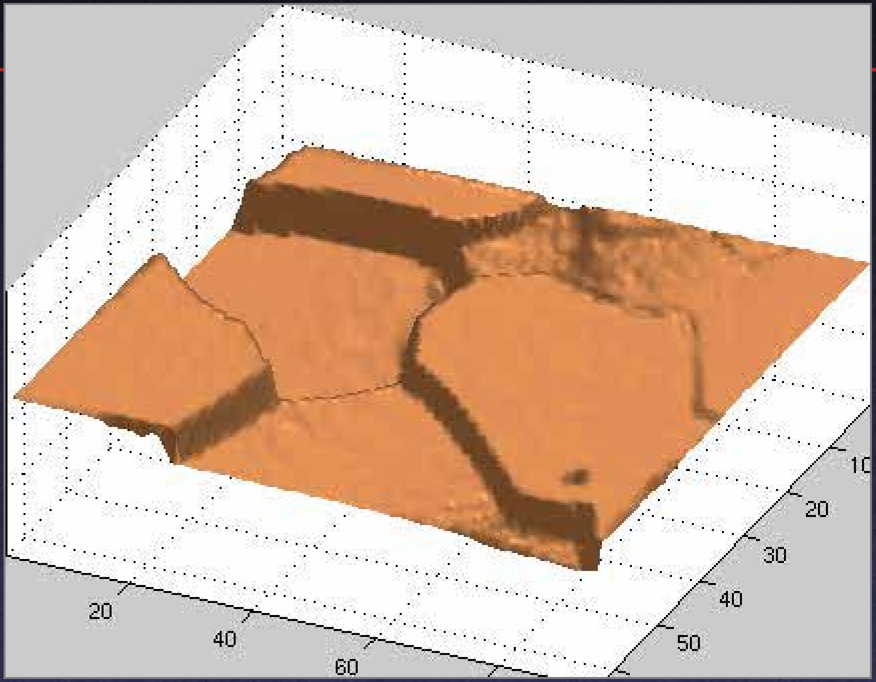
$$Ncut(A,B) = \frac{cut(A,B)}{volume(A)} + \frac{cut(A,B)}{volume(B)}$$

  - volume(A) = sum of costs of all edges that touch A
  - no fast exact algorithms…

[Based on slide by S. Seitz]

Interpretation as a Dynamical System

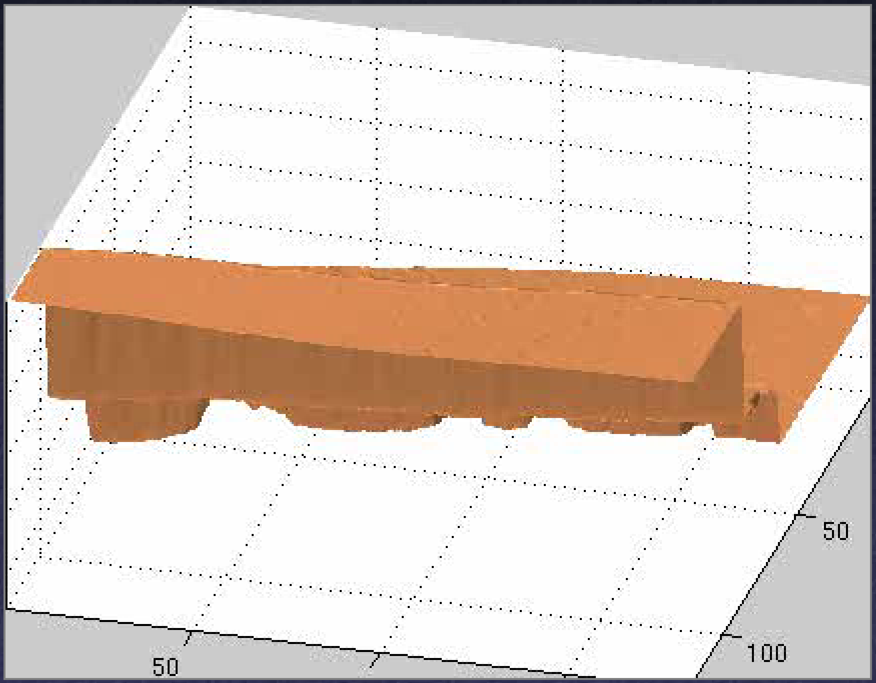




- Treat the links as springs and shake the system
  - elasticity proportional to cost
  - vibration "modes" correspond to segments
    - can compute these by solving a generalized eigenvector problem
    - for more details, see
      - J. Shi and J. Malik, Normalized Cuts and Image Segmentation, CVPR, 1997

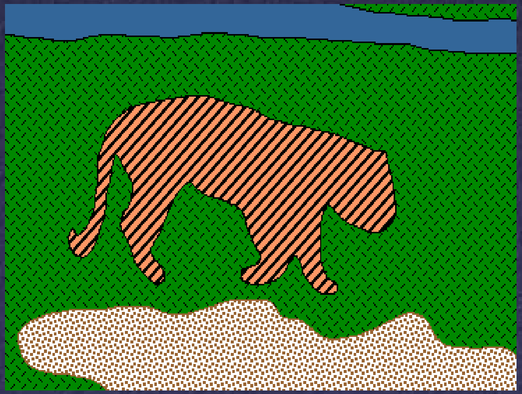
[Based on slide by S. Seitz]

Interpretation as a Dynamical System



[Based on slide by S. Seitz]

Designing Grouping Features



Low-level cues

- Brightness similarity
- Color similarity
- Texture similarity

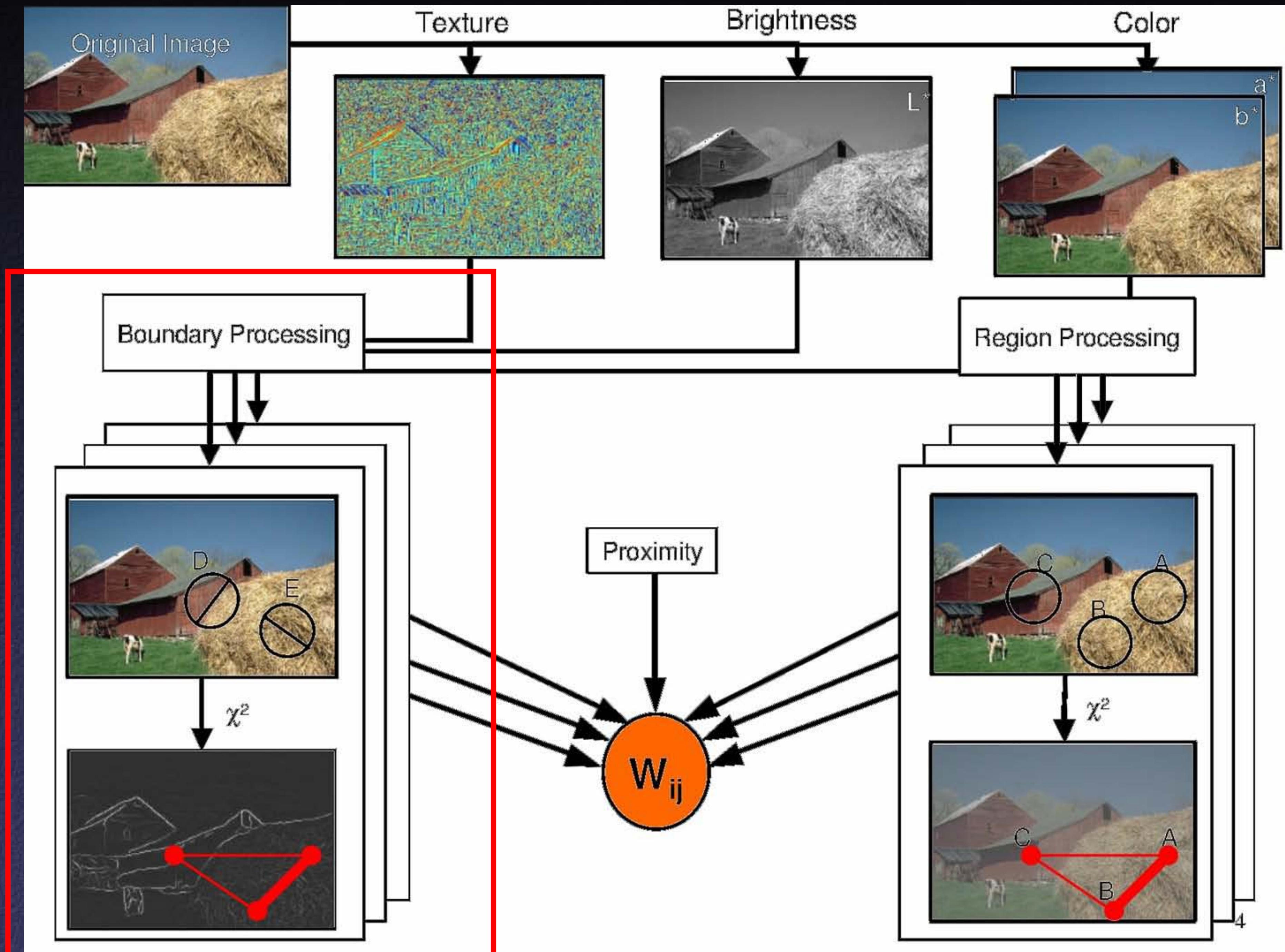
Mid-level cues

- Contour continuity
- Convexity
- Parallelism
- Symmetry

High-level cues

- Object knowledge
- Scene structure

[Based on slides by Xiaofeng Ren]



Original Image
Texture
Brightness
Color
L*
a*
b*
Boundary Processing
Region Processing
D
E
C
A
B
χ^2
Proximity
W_{ij}
χ^2
C
A
B
4

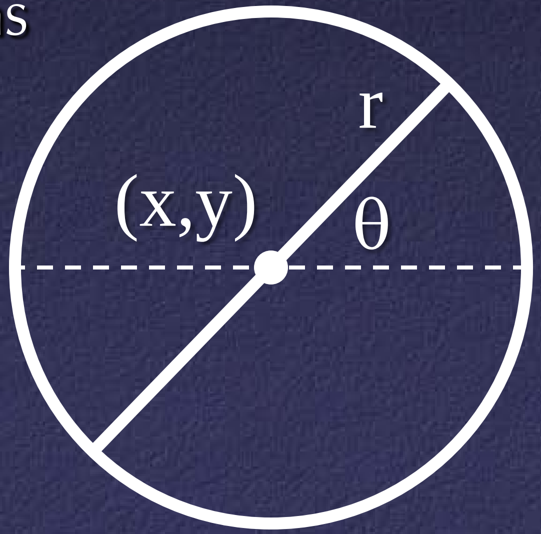
Brightness and Color Contrast

- 1976 CIE L*a*b* colorspace
- Brightness Gradient BG(x,y,r,θ)

  χ^2 difference in L* distribution
- Color Gradient CG(x,y,r,θ)

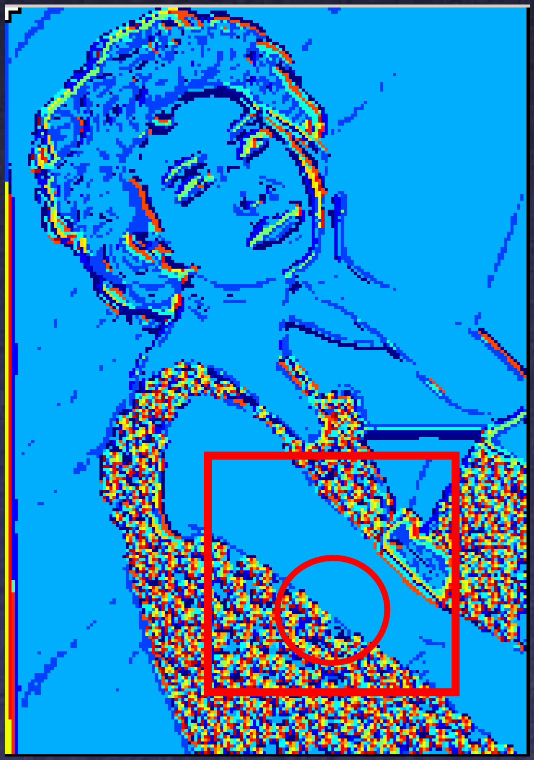
  χ^2 difference in a* and b* distributions

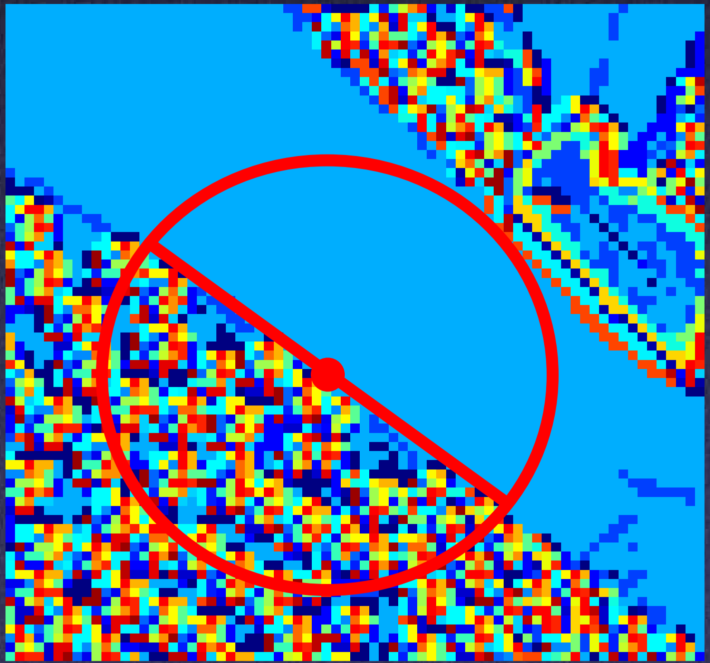
$$\chi^2(g,h) = \frac{1}{2}\sum_i \frac{(g_i - h_i)^2}{g_i + h_i}$$



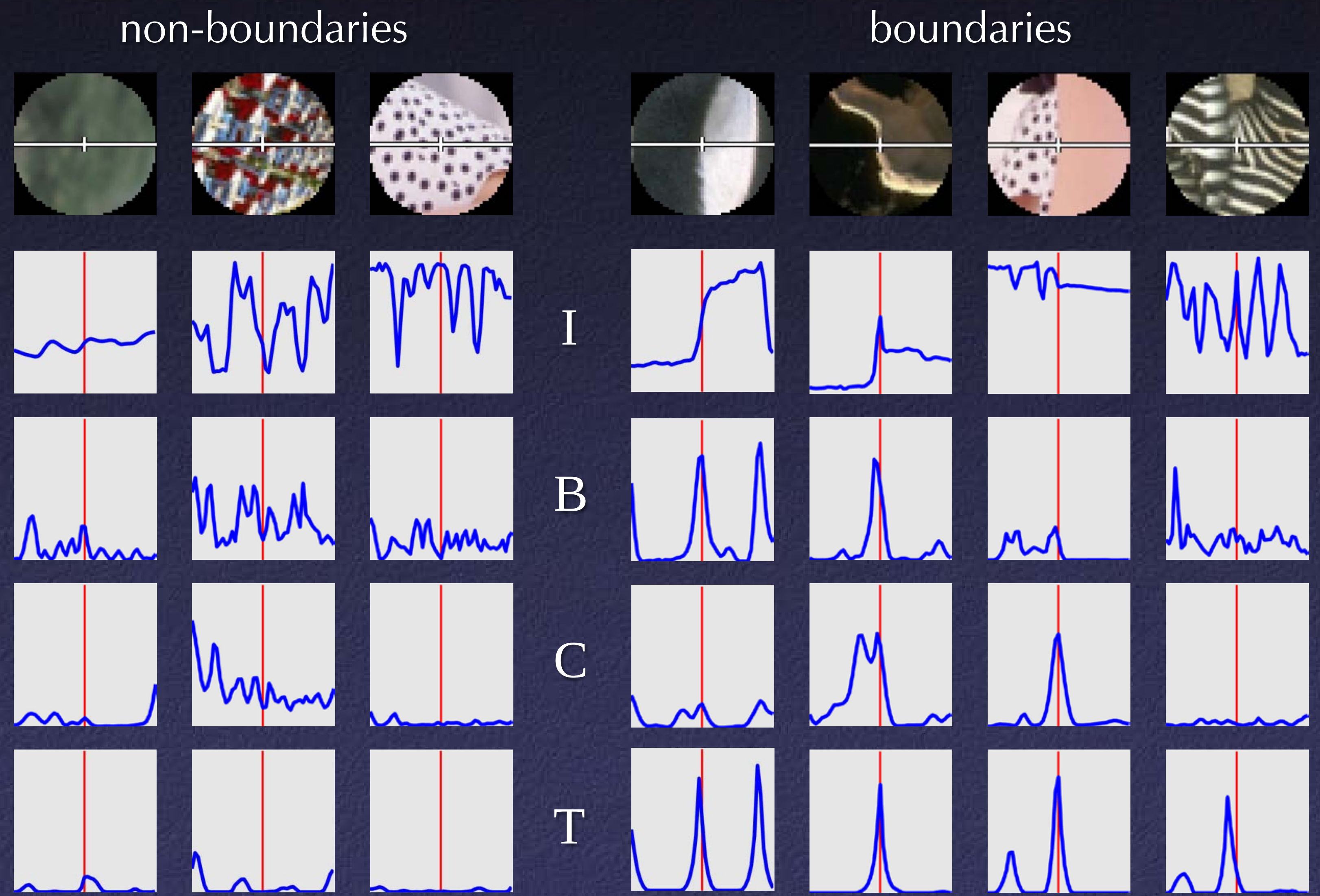
Texture Contrast

- Texture Gradient TG(x,y,r,θ)
  - χ^2 difference of texton histograms
  - Textons are vector-quantized filter outputs (through k-means)





Boundary Classification



Combining Cues

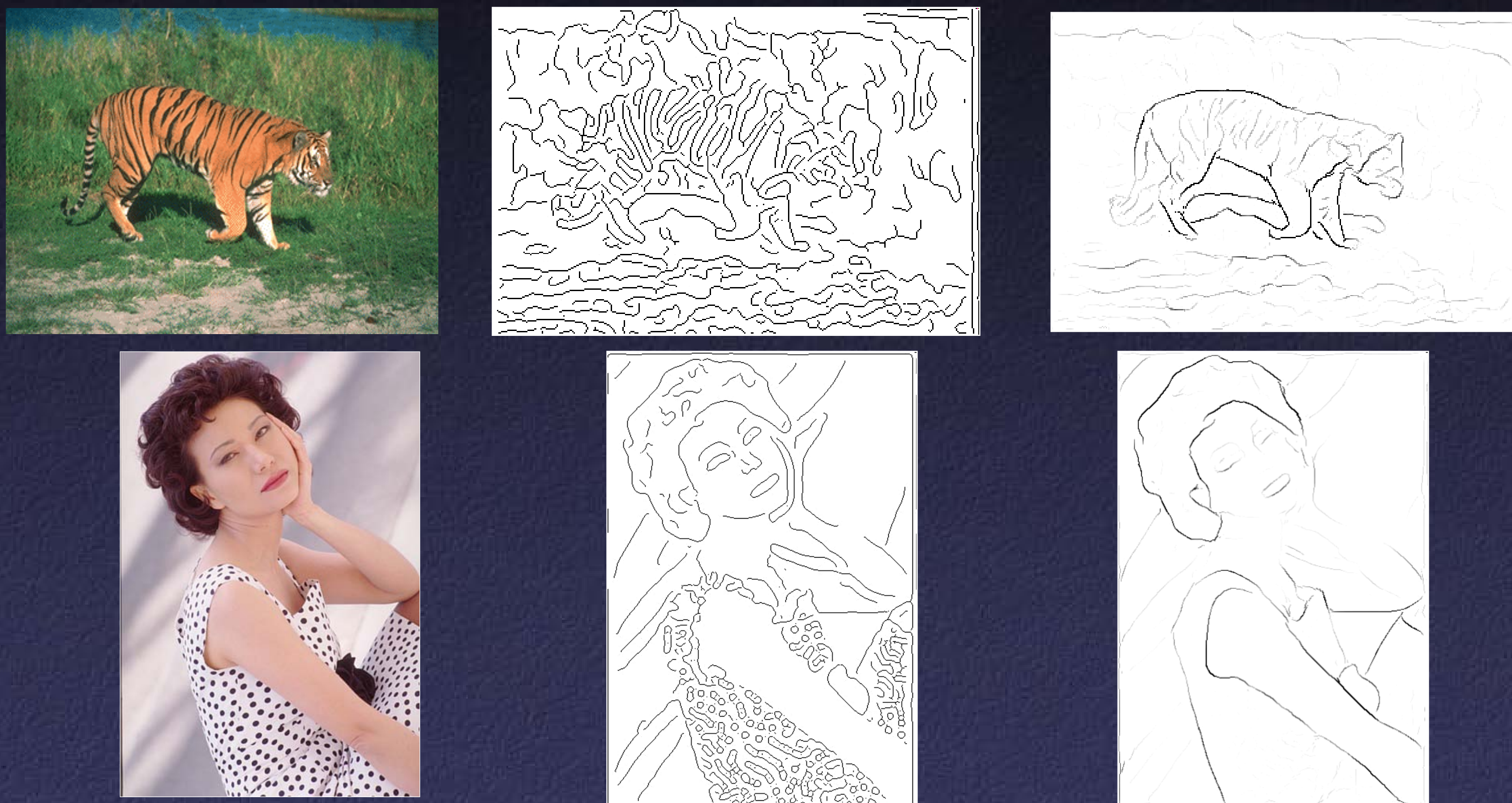
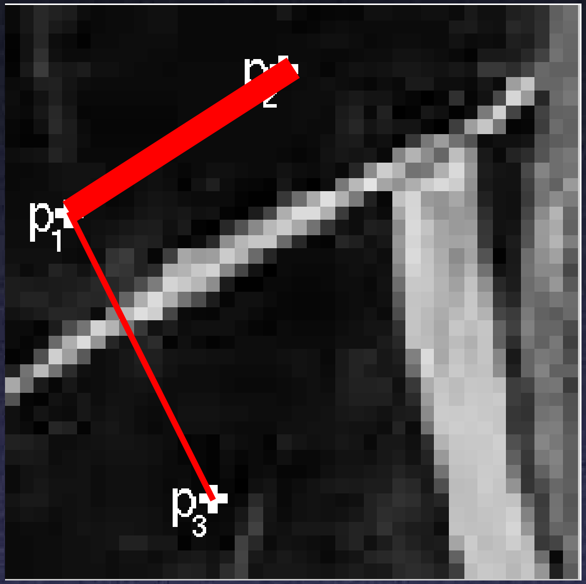


Image Canny Pb

[Martin, Fowlkes, Malik, *Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues*, PAMI 2004]

Affinity using Intervening Contour



W(p1,p2) >>W(p1,p3) as p1 and p2 are more likely to belong to the same region than are p1 and p3, which are separated by a strong boundary.