Computer Vision (CSE3010)

Dr. Susant Kumar Panigrahi
Assistant Professor
School of Electrical & Electronics Engineering
VIT BHOPAL

Module-1 Syllabus

Digital Image Formation And Low Level Processing:

- Overview and State-of-the-art, Fundamentals of Image Formation, Transformation: Orthogonal, Euclidean, Affine, Projective, Fourier Transform,
- Convolution and Filtering, Image Enhancement, Restoration, Histogram Processing.

Module-2 Syllabus

Depth Estimation And Multi-Camera Views:

Depth Estimation and Multi-Camera Views: Perspective, Binocular Stereopsis: Camera and Epipolar Geometry; Homography, Rectification, DLT, RANSAC, 3-D reconstruction framework; Auto-calibration. apparel.

Module-3 Syllabus

Feature Extraction And Image Segmentation:

- **Feature Extraction**: Edges - Canny, LOG, DOG; Line detectors (Hough Transform), Corners - Harris and Hessian Affine, Orientation Histogram, SIFT, SURF, HOG, GLOH, Scale-Space Analysis- Image Pyramids and Gaussian derivative filters, Gabor Filters and DWT.
- **Image Segmentation:** Region Growing, Edge Based approaches to segmentation, Graph-Cut, Mean-Shift, MRFs, Texture Segmentation; Object detection.

Module-4 Syllabus

Pattern Analysis And Motion Analysis:

- **Pattern Analysis**: Clustering: K-Means, K-Medoids, Mixture of Gaussians, Classification: Discriminant Function, Supervised, Un-supervised, Semi-supervised; Classifiers: Bayes, KNN, ANN models;
- **Dimensionality Reduction**: PCA, LDA, ICA; Non-parametric methods. Motion Analysis: Background Subtraction and Modelling, Optical Flow, KLT, Spatio-Temporal Analysis, Dynamic Stereo; Motion parameter estimation.

Module-5 Syllabus

Shape From X:

Light at Surfaces; Phong Model; Reflectance Map;

Albedo estimation; Photometric Stereo; Use of Surface Smoothness

Constraint; Shape from Texture, color, motion and edges.

Guest Lecture on Contemporary Topics

Text Books

1. Richard Szeliski, Computer Vision: Algorithms and Applications, Springer-Verlag London Limited 2011.
2. Computer Vision: A Modern Approach, D. A. Forsyth, J. Ponce, Pearson Education, 2003.

Reference Book(s):

1. R.C. Gonzalez and R.E. Woods, Digital Image Processing, Addison- Wesley, 1992.
2. Richard Hartley and Andrew Zisserman, Multiple View Geometry in Computer Vision, Second Edition, Cambridge University Press, March 2004.
3. K. Fukunaga; Introduction to Statistical Pattern Recognition, Second Edition, Academic Press, Morgan Kaufmann, 1990.

Required Tools/Software/IDLE:

1. Python/jupyter-notebook/google-colab
2. OpenCV
3. MATLAB

Indicative List of Experiments:

1. Implement image preprocessing and Edge
2. Implement camera calibration methods
3. Implement Projection
4. Determine depth map from Stereo pair
5. Construct 3D model from Stereo pair
6. Implement Segmentation methods
7. Construct 3D model from defocus image
8. Construct 3D model from Images
9. Implement optical flow method
10. Implement object detection and tracking from video
11. Face detection and Recognition
12. Object detection from dynamic Background for Surveillance
13. Content based video retrieval
14. Construct 3D model from single image

Computer Vision

Unit – 04

Pattern Analysis: Segmentation and Clustering

Standing on the shoulder of Giants: Ref: Few Slides borrowed from:

1. Prof. Shree Nayar, *First Principles of Computer Vision is a lecture series.*
2. Prof. Mubarak Shah, *Computer Vision Video Lectures.*

Image Segmentation

- ✓ Image segmentation is the process of partitioning a digital image into multiple image segments, also known as image regions or image objects (sets of pixels).

- ✓ The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze.

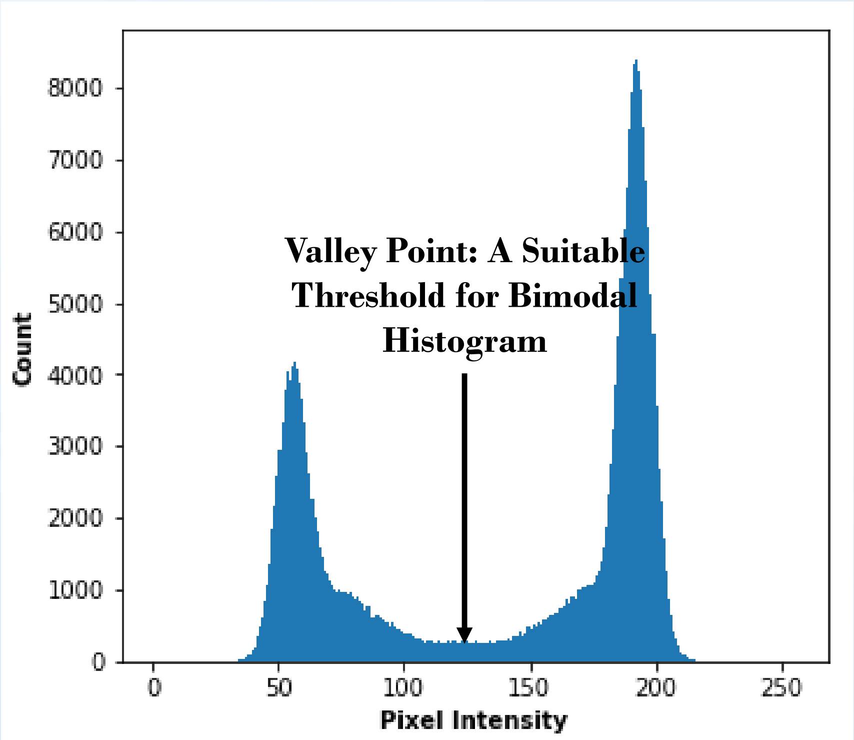
- ✓ More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.

Binary Segmentation: The Simplest Approach

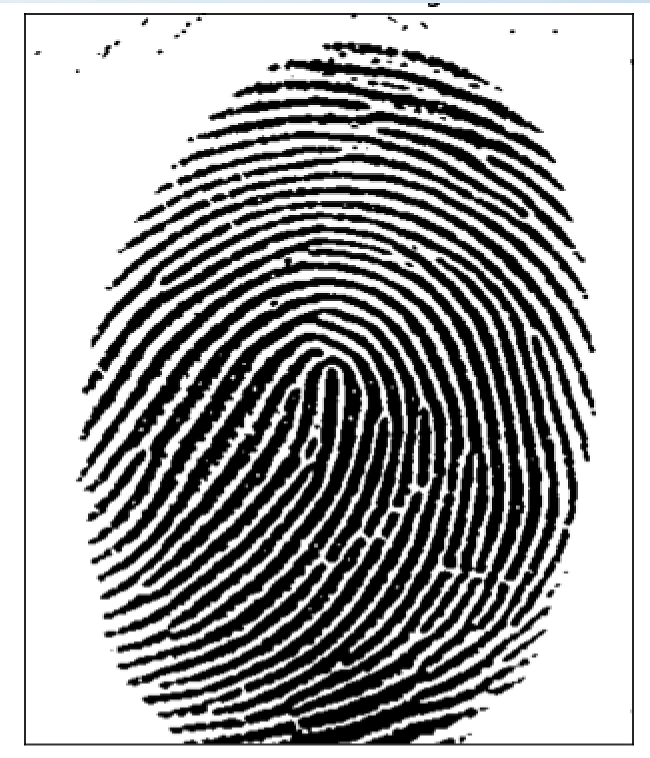
- Segments the image into two regions : Object and Background.
- Mathematically, an Image $f(x, y)$ can be thresholded at T as;

$$g(x,y) = \begin{cases} 1, & if\ f(x,y) > T \\ 0, & if\ f(x,y) \leq T \end{cases}$$

Grayscale Image



Histogram of Corresponding Image



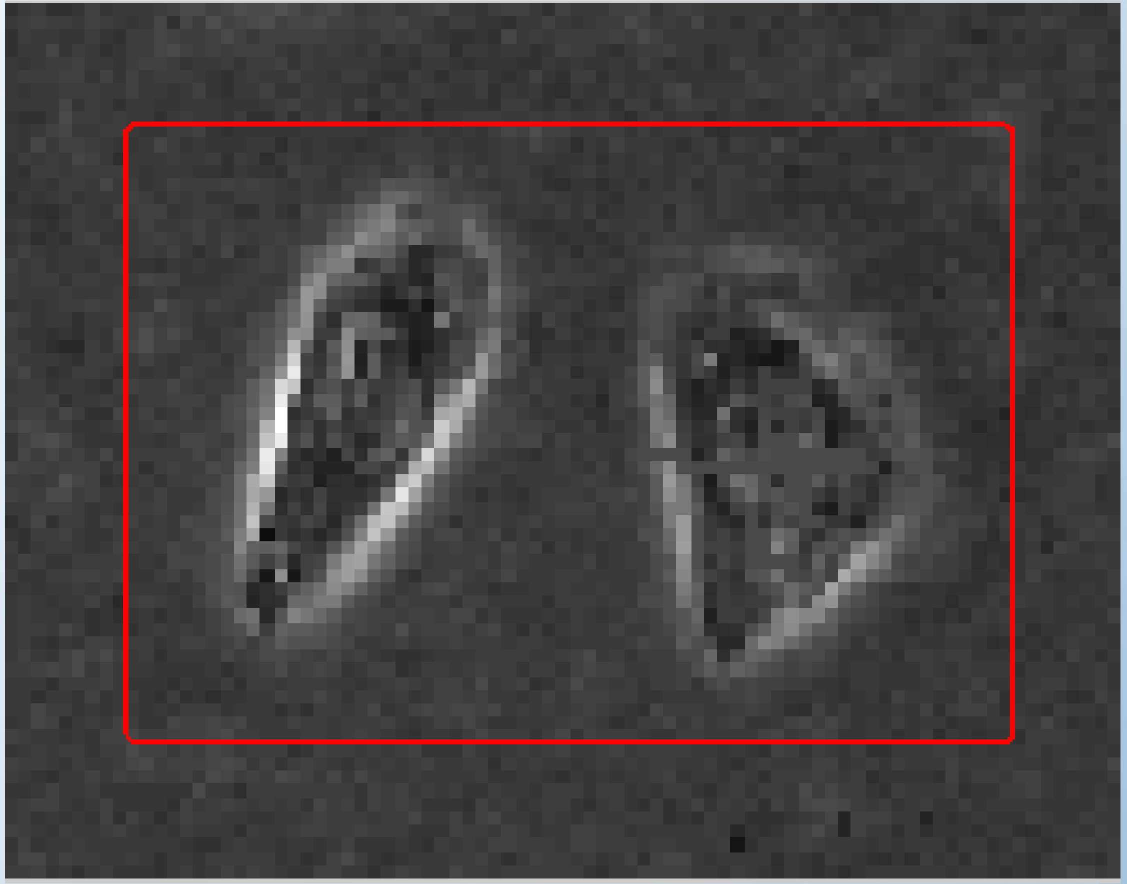
Thesholded Image
[Binary Image]

Active Contour Detection

Given:
Approximate boundary (contour) around the object.

Active Contour:
Evolves (moves) the contour to fit exact object boundary.

- ✓ The boundary segments the object from the rest of the boundary.
- ✓ Require manual annotation of initial object boundary



Segmentation is Hard! Natural Scenes

Image Segmentation

- Group pixels with similar visual characteristics.

Topics:

1. Segmentations by Human [CO3]
2. Segmentation as Clustering [CO4]
3. K-Mean Clustering [CO4]
4. Mean Shift Segmentation [CO4]
5. Graph-Cut Segmentation [CO3]

Segmentation as Grouping by Human Perception

- ✓ The dog is not recognized through its parts but perceived as whole.
- ✓ We see as a whole then create/segment the boundary of the object

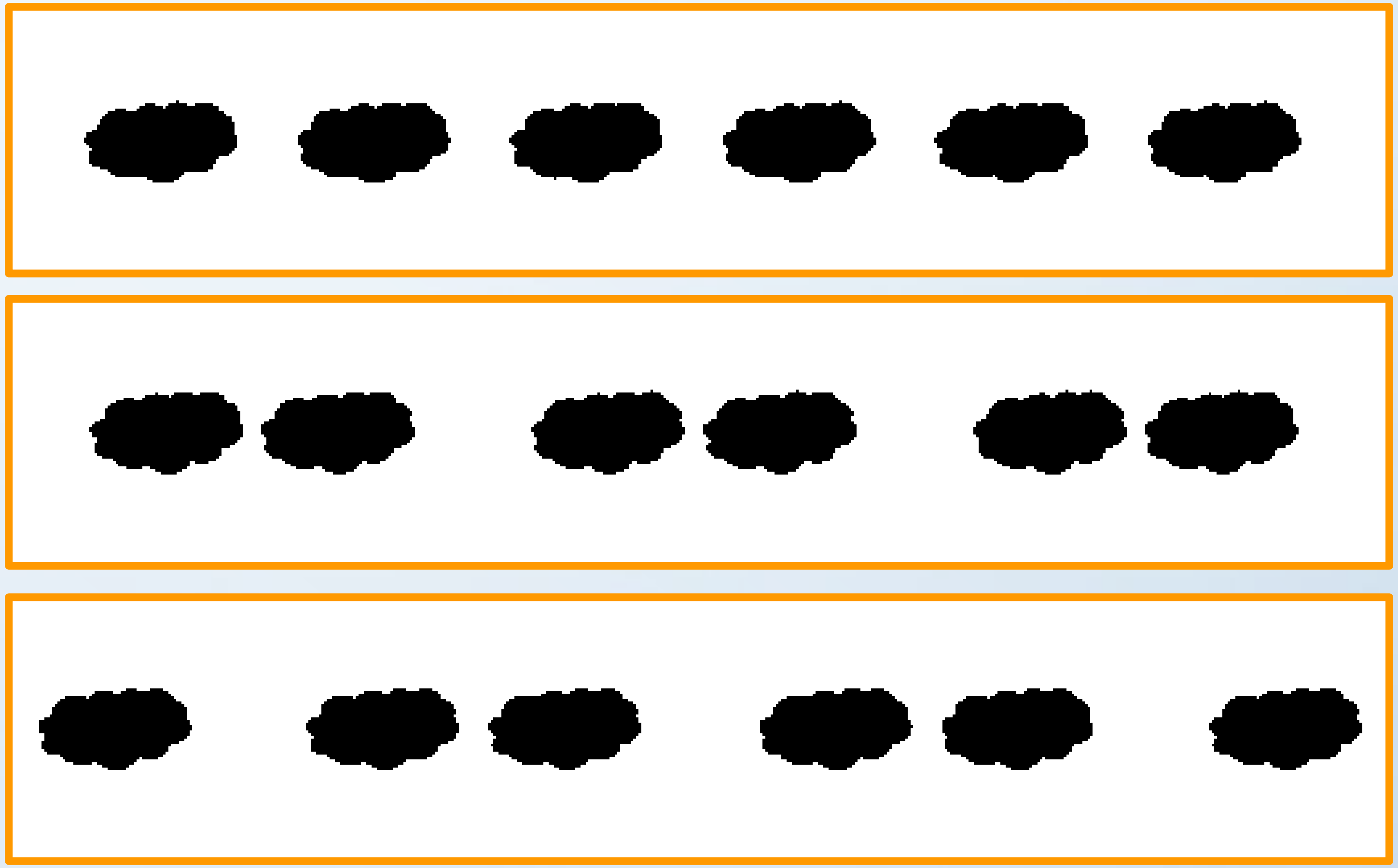
Gestalt Psychology

Gestalt a form structure, arrangement, or pattern of physical, biological, or psychological phenomena

- Something such as a structure or experience that, when considered as a whole, has qualities that are more than the total of all its parts

Principles of Segmentation by Human Visual System

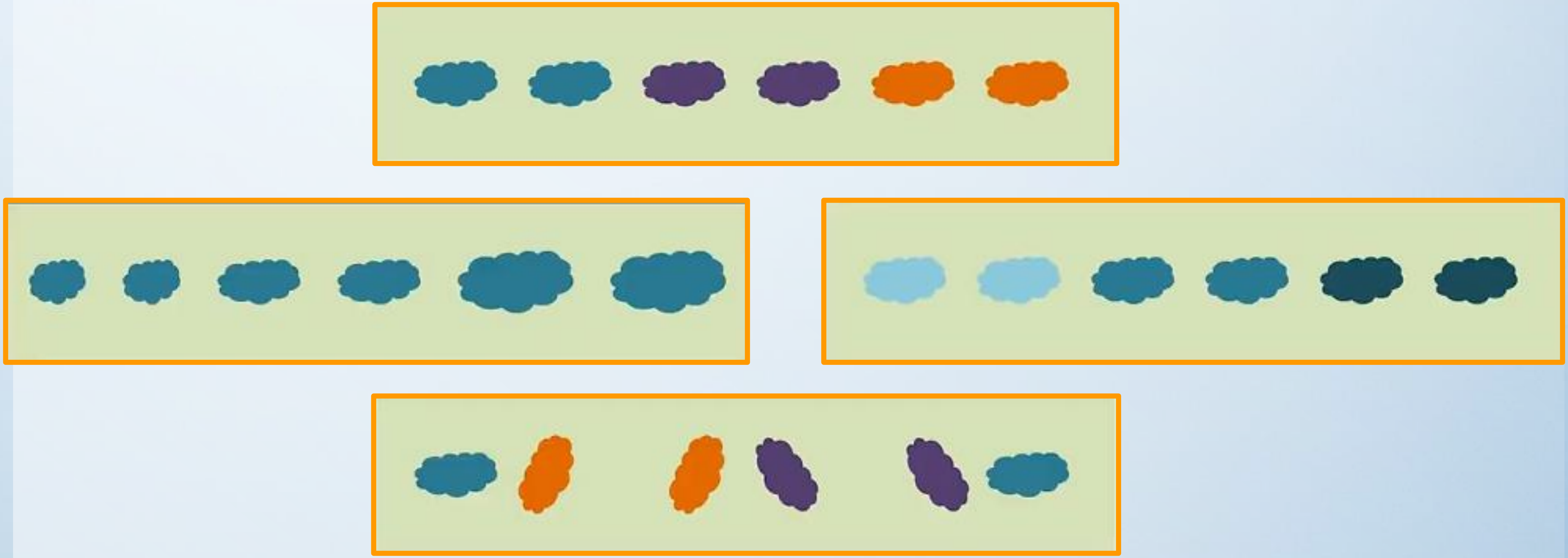
Proximity Principle



Objects that are nearby tend to be grouped together

Principles of Segmentation by Human Visual System

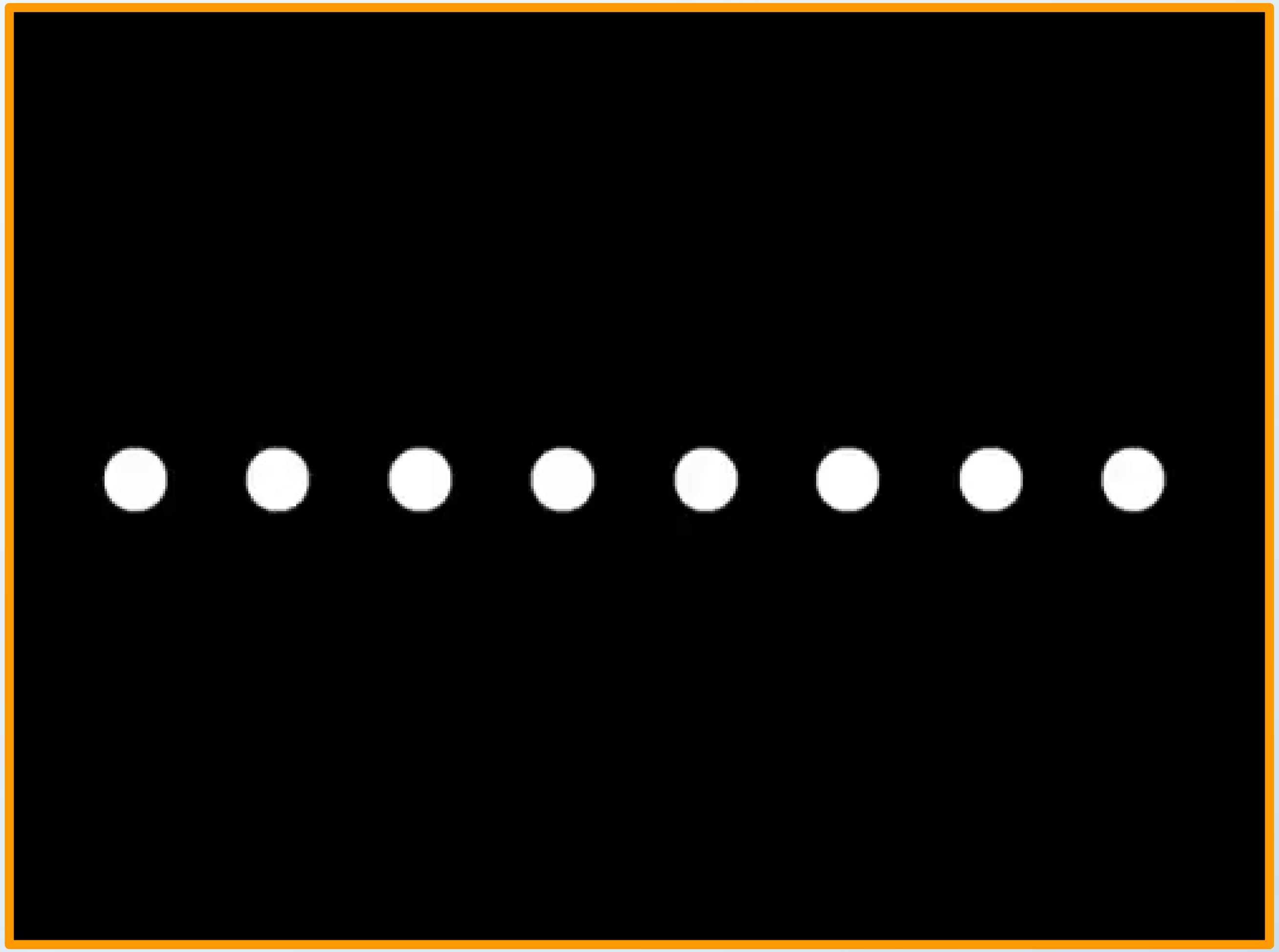
Similarity Principle



Similar objects are grouped together

Principles of Segmentation by Human Visual System

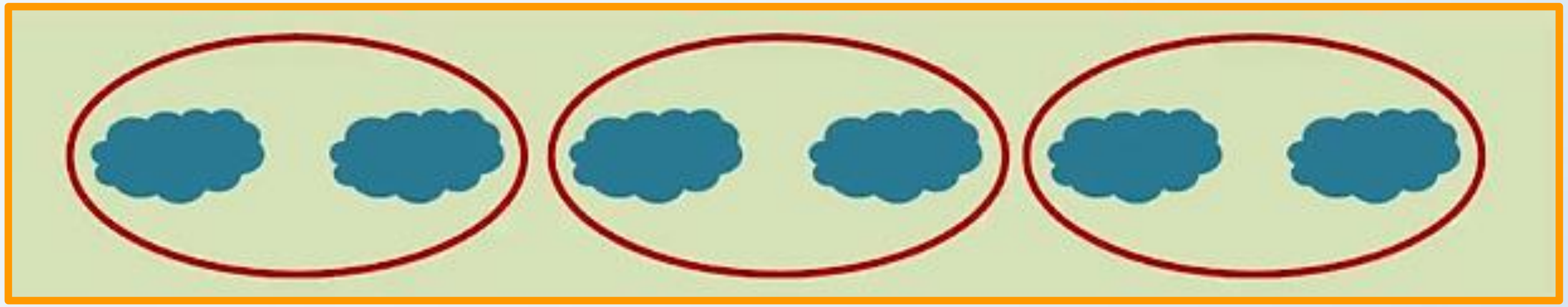
Common Fate

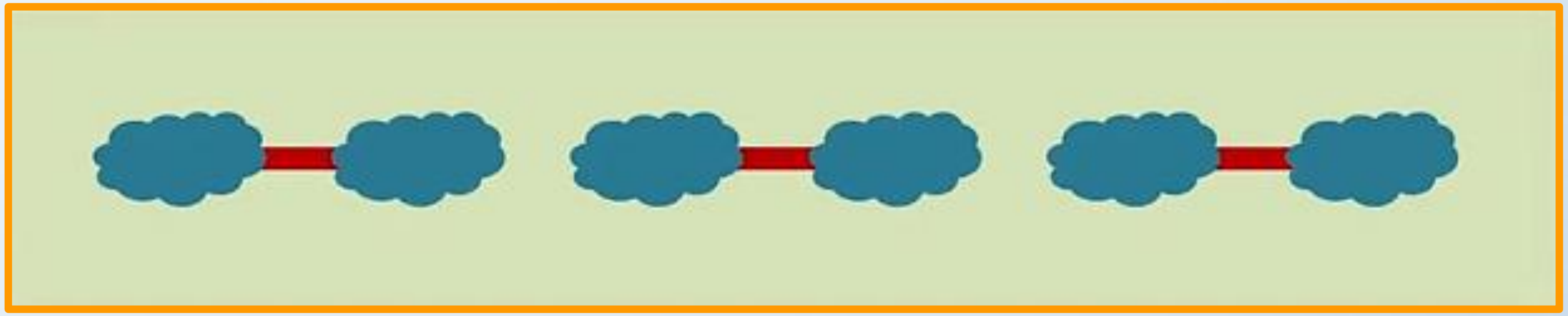


Objects with similar motion or change in appearance are grouped together

Principles of Segmentation by Human Visual System

Common Region or Connectivity

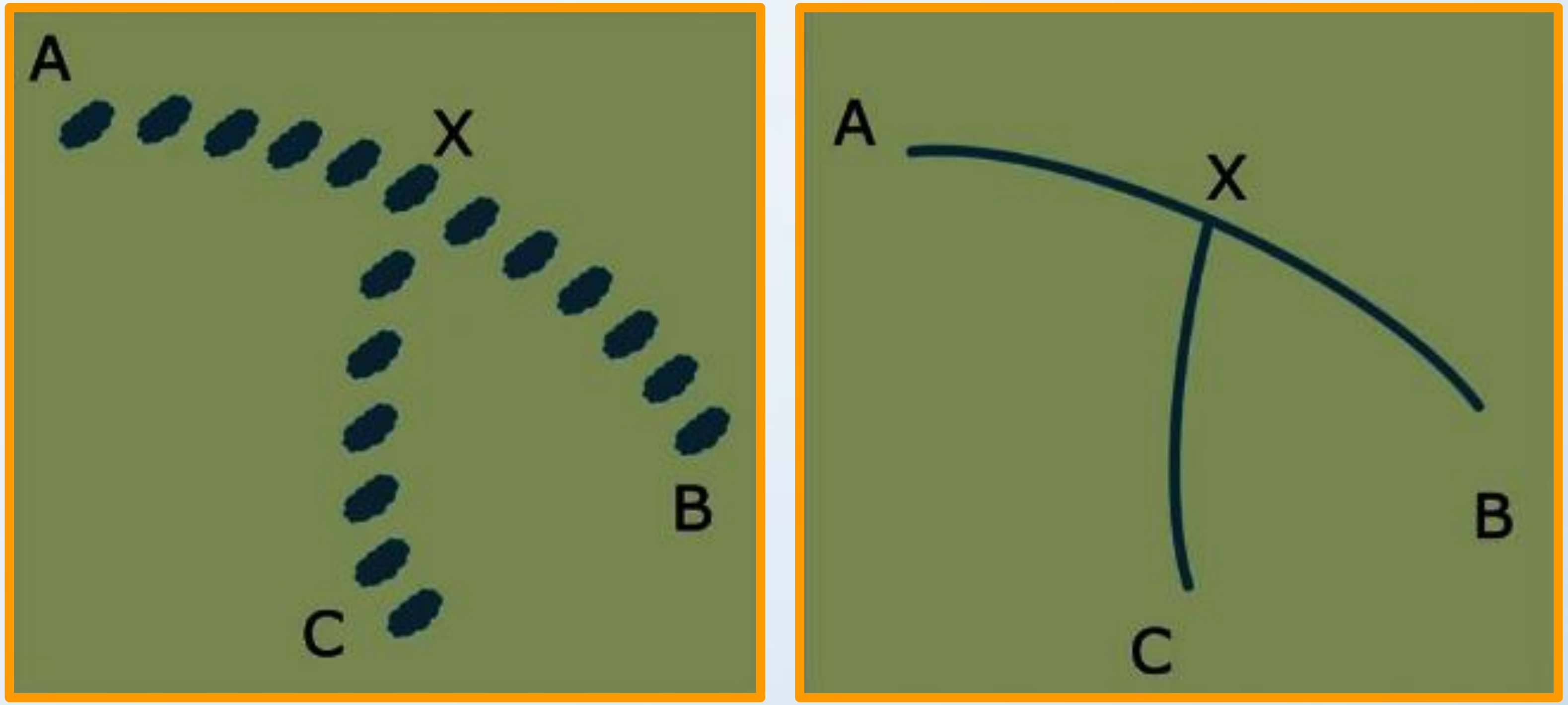




Connected objects are grouped together

Principles of Segmentation by Human Visual System

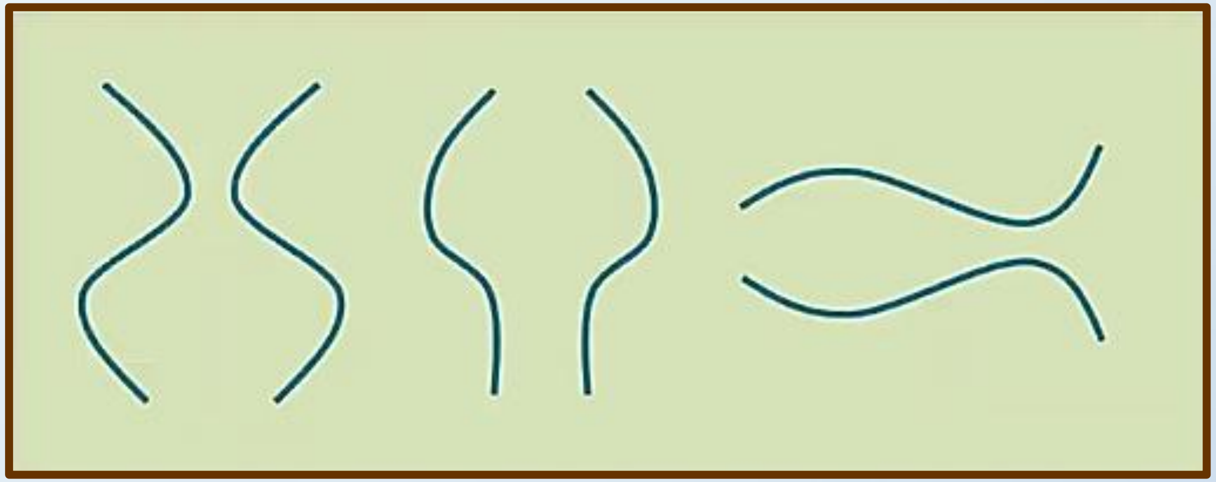
Continuity



Features on a continuous curve are grouped together

Principles of Segmentation by Human Visual System

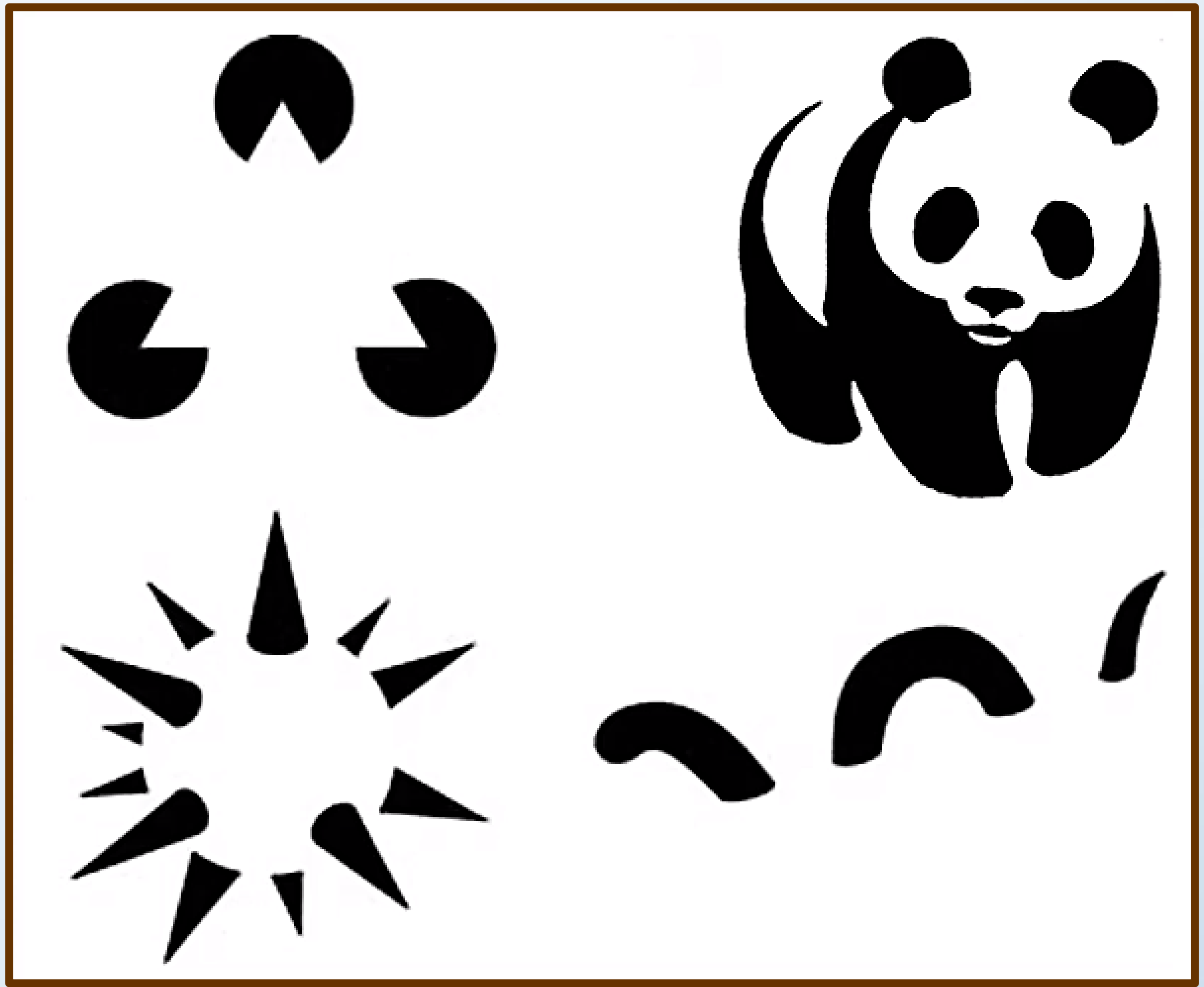
Symmetry



Parallel and symmetrical features are grouped together

Principles of Segmentation by Human Visual System

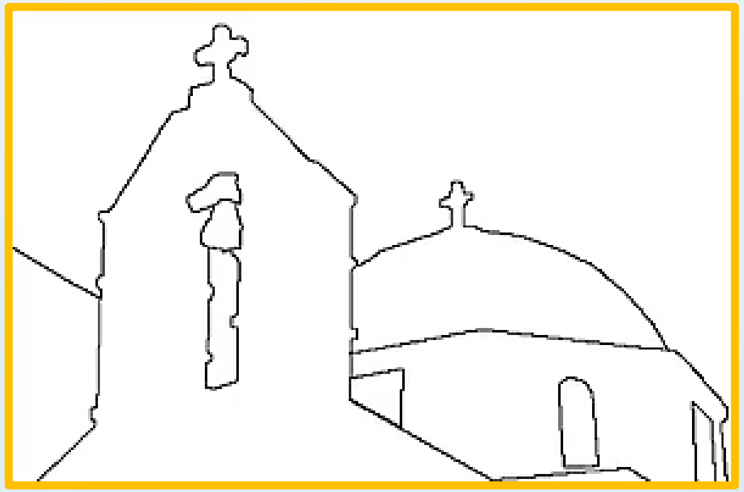
Illusion



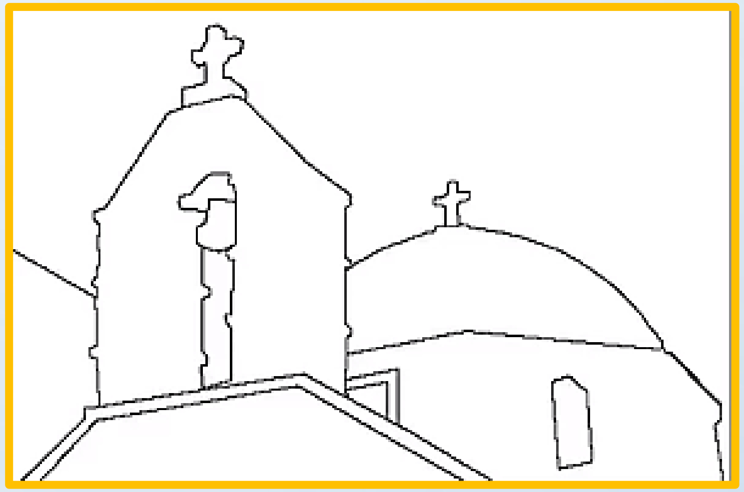
Illusory or subjective contours are perceived as objects

Segmentation by Human: Subjective

Input Image



User 1



User 2

User 3

Segmentation Strategies

Subjective and No Definite Guideline Segmentation is intuitive to us.

- Very hard to translate these intuitions to an algorithm.

Top-Down Segmentation: Pixel belong together because they come from same object.

Bottom-Up Segmentation: Pixel belong together because they look similar. [Better strategy to translate into algorithm]

Segmentation as Clustering: Pattern Analysis

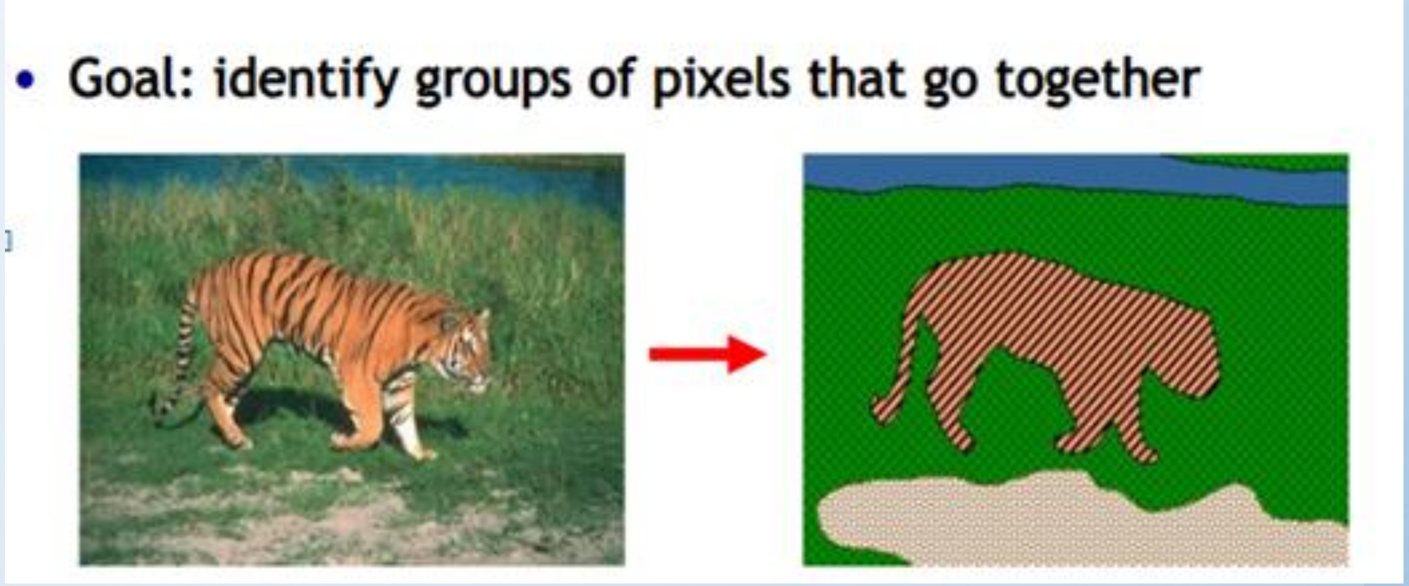
- ✓ One natural view of segmentation is that we are attempting to determine which components of a data set naturally "belong together".

- ✓ This is a problem known as clustering.

Image Features to Consider for Clustering:

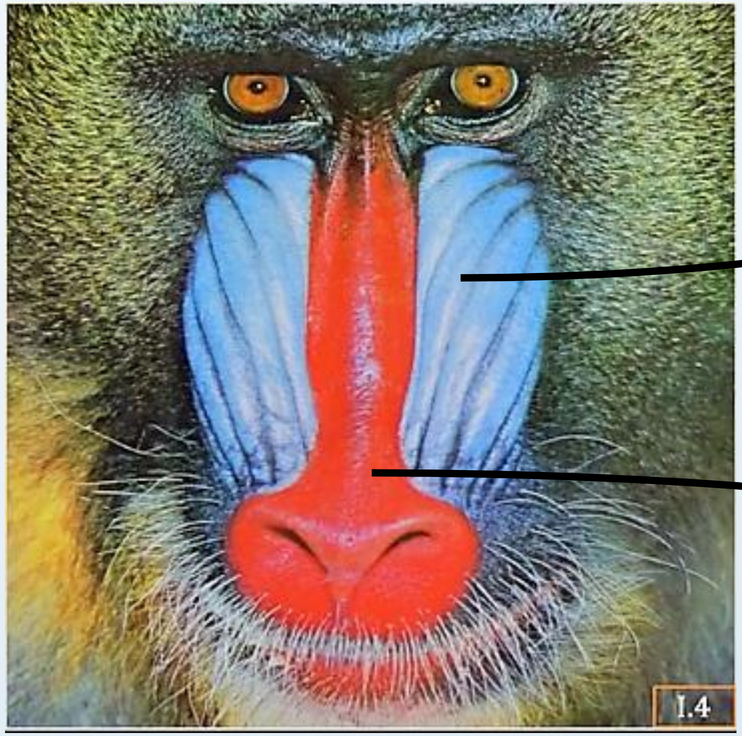
Visual similarity can be based on:

- ✓ Brightness.
- ✓ Color
- ✓ Position
- ✓ Depth
- ✓ Motion
- ✓ Texture
- ✓ Material ....

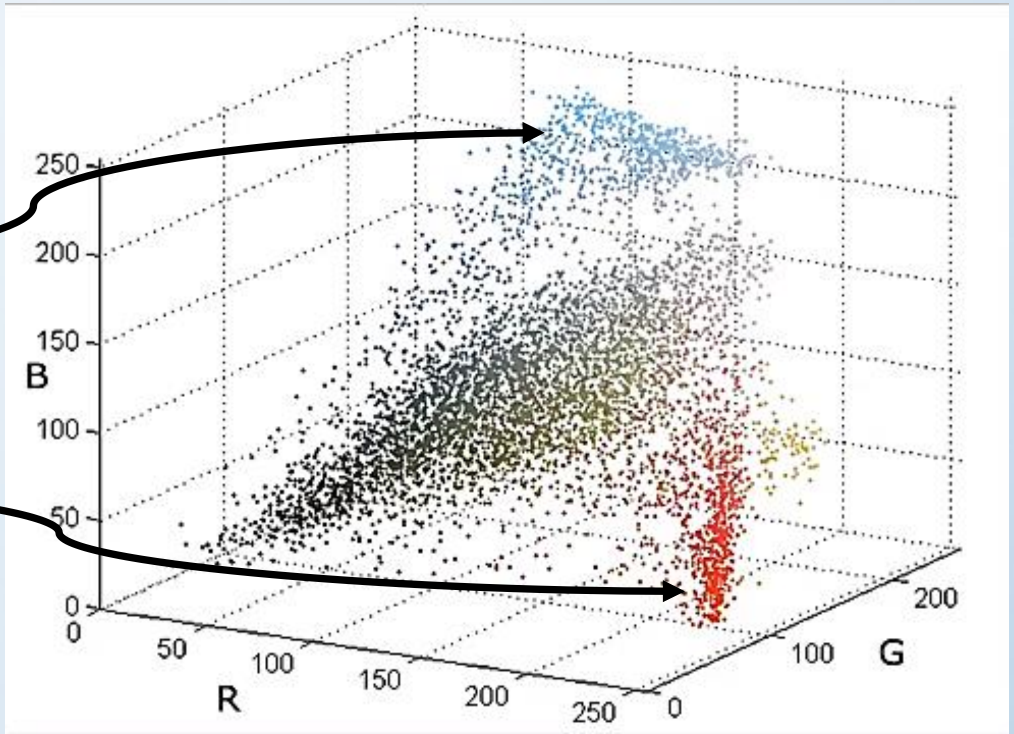


Segmentation as Clustering: Pixels in Euclidian Space

Euclidean Space: Generalization of 3D Cartesian space to higher dimension.



Input Image



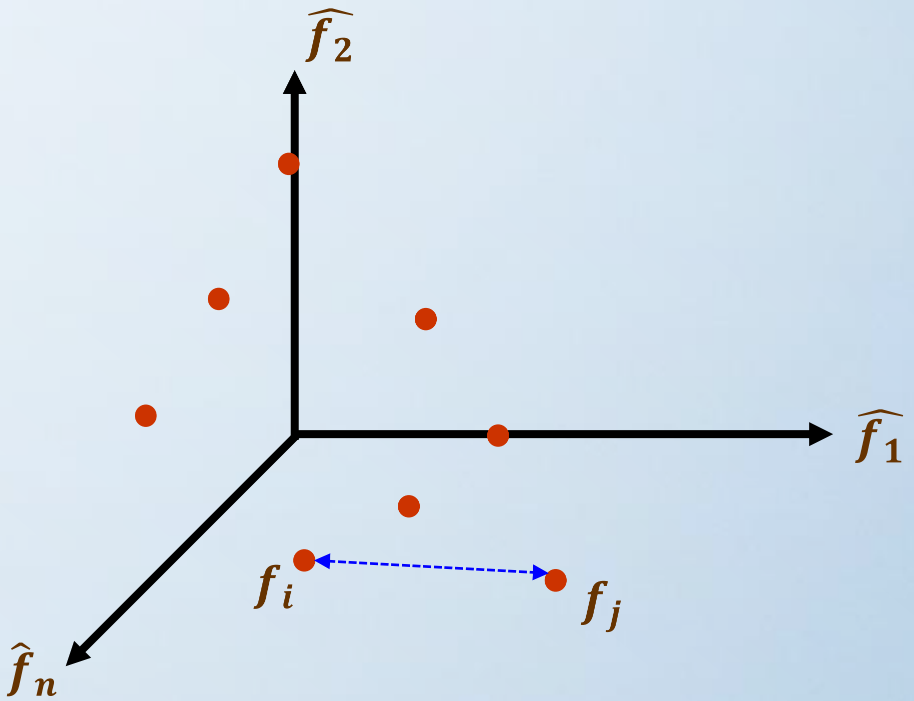
Color Distribution in (RGB) various channels.
Color of Feature point = Color of image pixel

Pixels as feature vectors: $[R, G, B, x, y, d, \ldots]$

Color, Position, Depth, etc….

Segmentation as Clustering: Pixels in Similarity

Euclidean Space: Let i and j be the two pixels whose features are f_i and f_j.



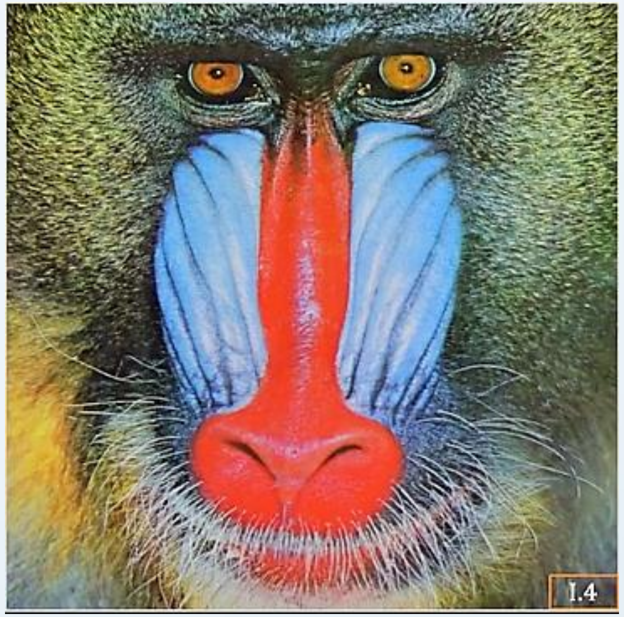
$[R, G, B, x, y, d, \ldots] - Space$

$\mathcal{L}^2$ **Distance** between f_i and f_j:

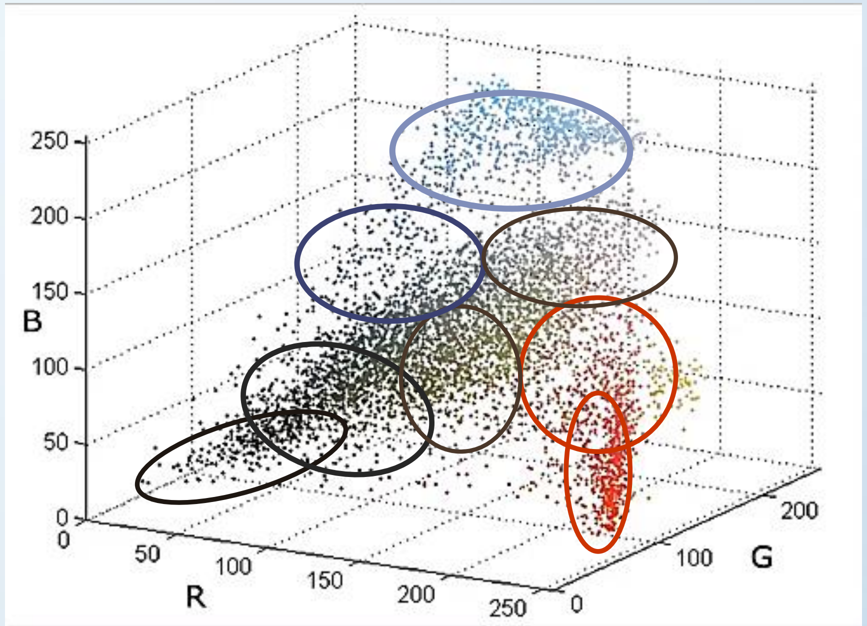
$$S(f_i, f_j) = \sqrt{\sum_k (f_{ik} - f_{jk})^2}$$

Smaller the distance greater the similarity.

Clustering Similar Pixels



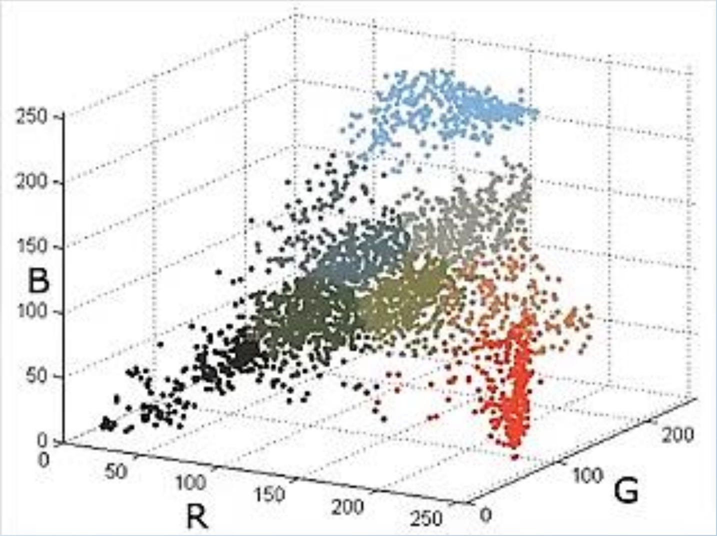
Input Image



Color Distribution in (RGB) various channels.
Color of Feature point = Color of image pixel

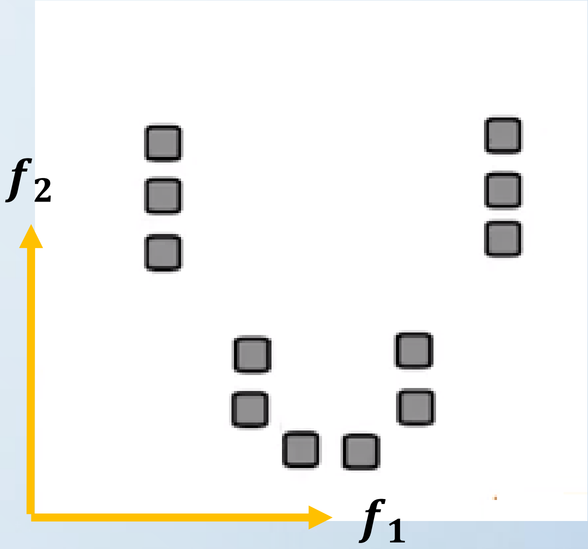
Segmented Image

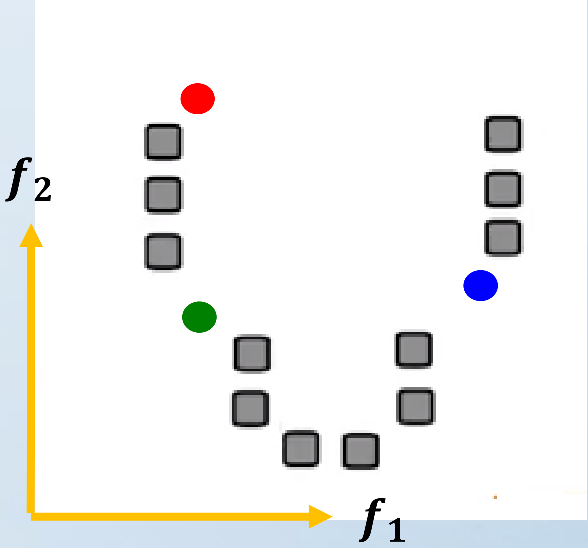
Color Coded Cluster.
Colors of Feature point = Color of image segment



K-Mean Clustering: 3 Means Clustering

Problem: Segment the given data point shown in 2D to exactly three clusters. (K=3)



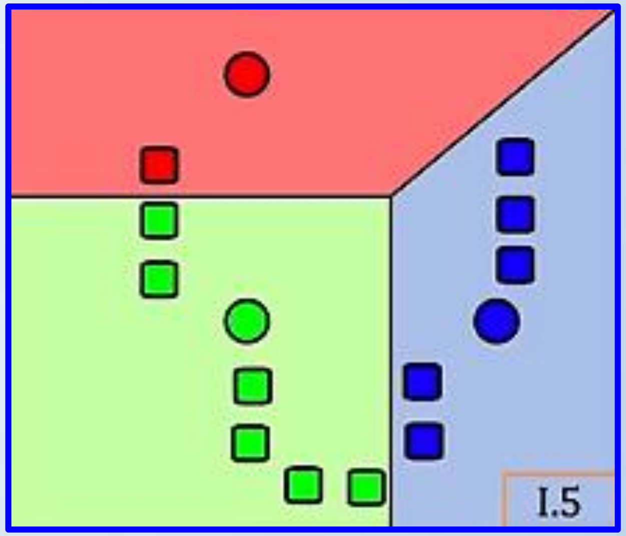


Solution:

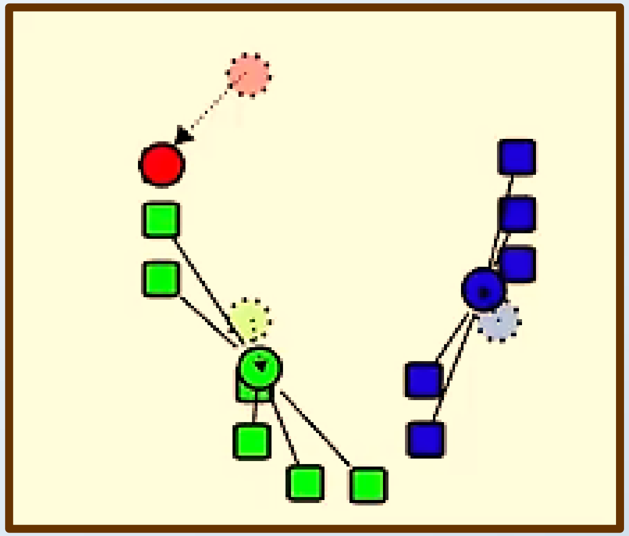
Step 1: Randomly generate the initial centroid (mean) of the three cluster. (K=3)

K-Mean Clustering: 3 Means Clustering

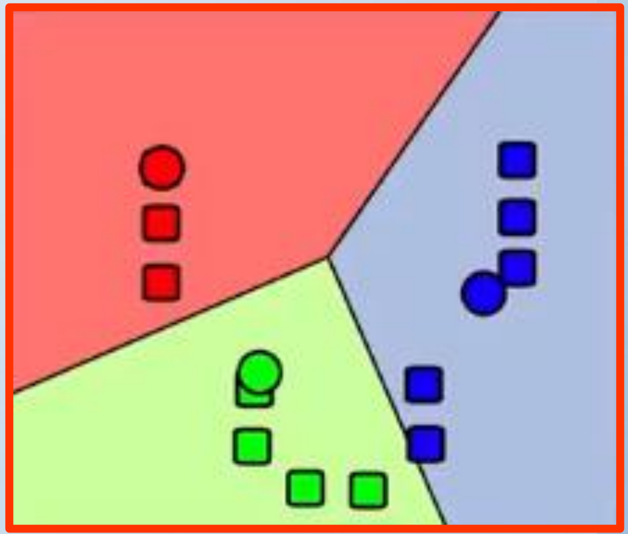
Solution:
Step 2: Create three clusters by assigning each feature point to the nearest mean. (K=3)



Solution:
Step 3: Re-compute the mean of each cluster. (K=3)



Solution:
Step 4: Repeat step 2 and 3 until reach convergence. (K=3)



K-Mean Clustering: Algorithm

Given: Image with N pixels and number of clusters k.

Task: Find the k clusters.

Clustering:

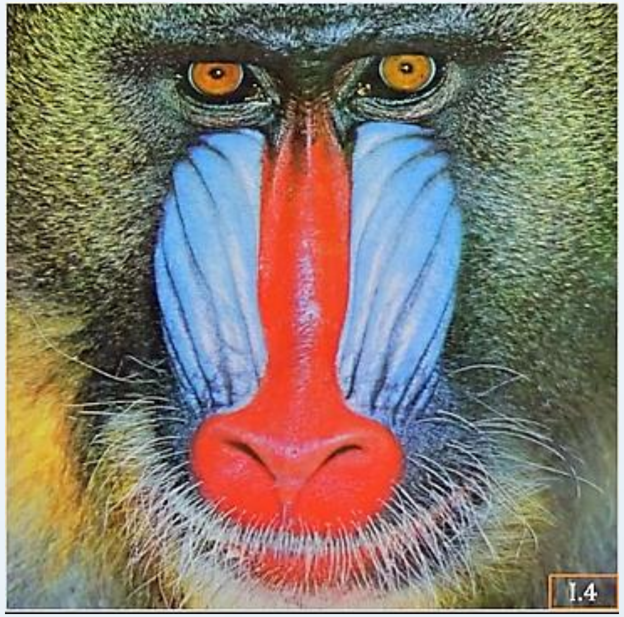
1: Pick k points randomly as the initial centroids (means) $\{\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_k\}$ of the k clusters in feature space.

2: For each pixel $\mathbf{x}_j$ find nearest cluster mean $\mathbf{m}_i$ to pixel's feature $\mathbf{f}_j$ and assign pixel to cluster i.

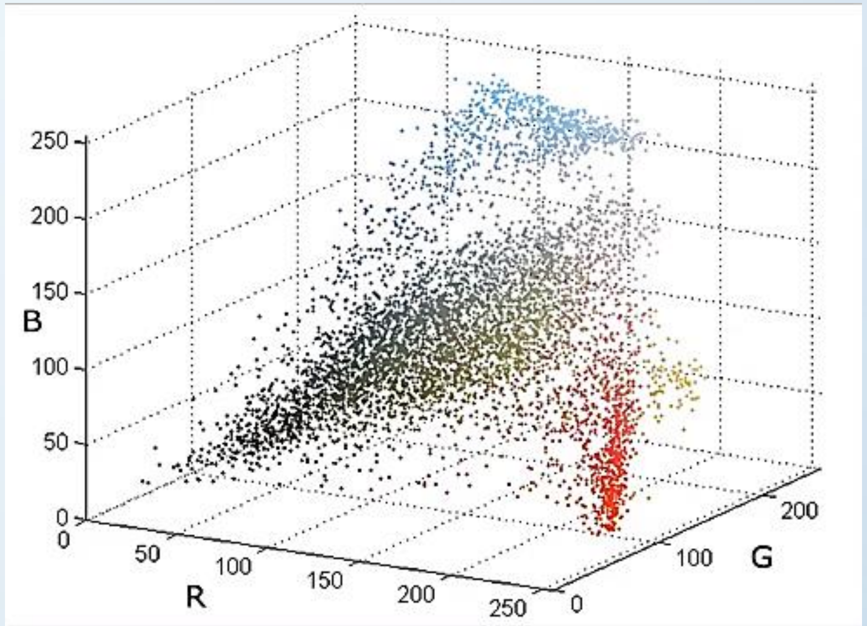
3: Recompute mean for each cluster using its assigned pixels.

4: If changes in all k means is less than a threshold ε, stop. Else go to step 2.

K-Mean Clustering: Results ($K = 2$)



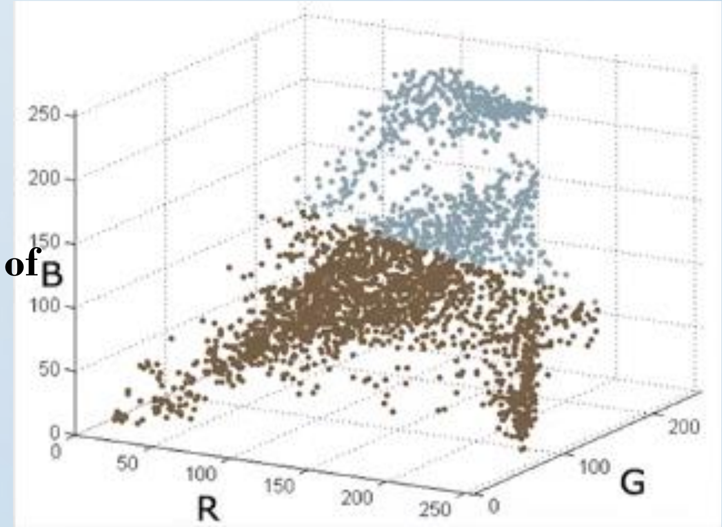
Input Image



Color Distribution in (RGB) various channels.
Color of Feature point = Color of image pixel

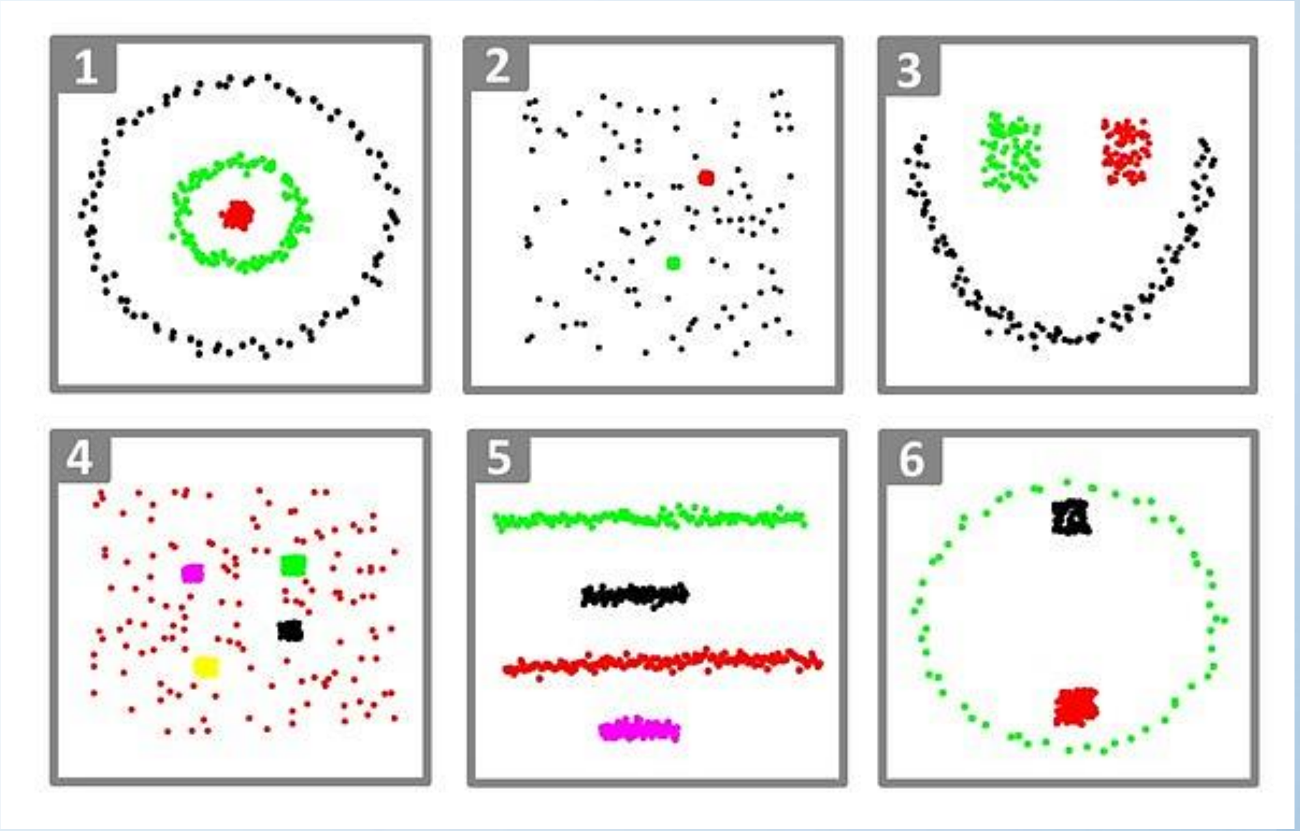
Segmented Image
$[R, G, B] - Space$
(K = 2)

Color Distribution in (RGB) various channels.
Color of Feature point = Color of image pixel

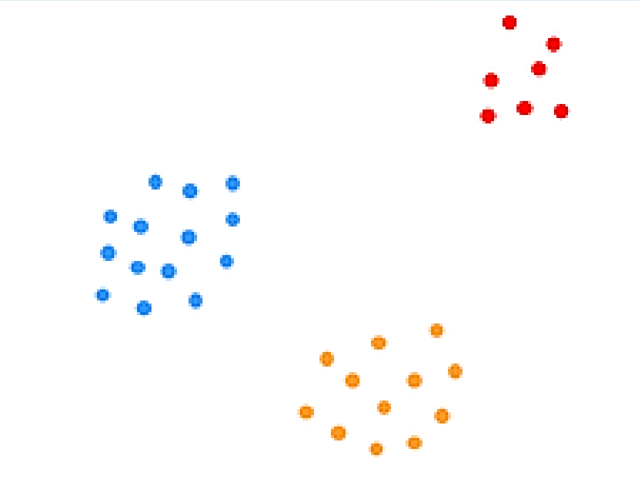


K-Mean Clustering: Limitations

- ✓ K-means will fail to effectively cluster these, even when the true number of clusters K is known to the algorithm.



This is because K-means, as a data-clustering algorithm, is ideal for discovering globular clusters like the ones shown below, where all members of each cluster are in close proximity to each other (in the Euclidean sense).



Data-clustering vs Graph-clustering

K-Mean Clustering: Limitations

✓ The results are at least somewhat dependent on its starting point.

✓ There is no statistical indication of the "right" number of clusters.

- *How can we choose a "good" K for K-means clustering?*

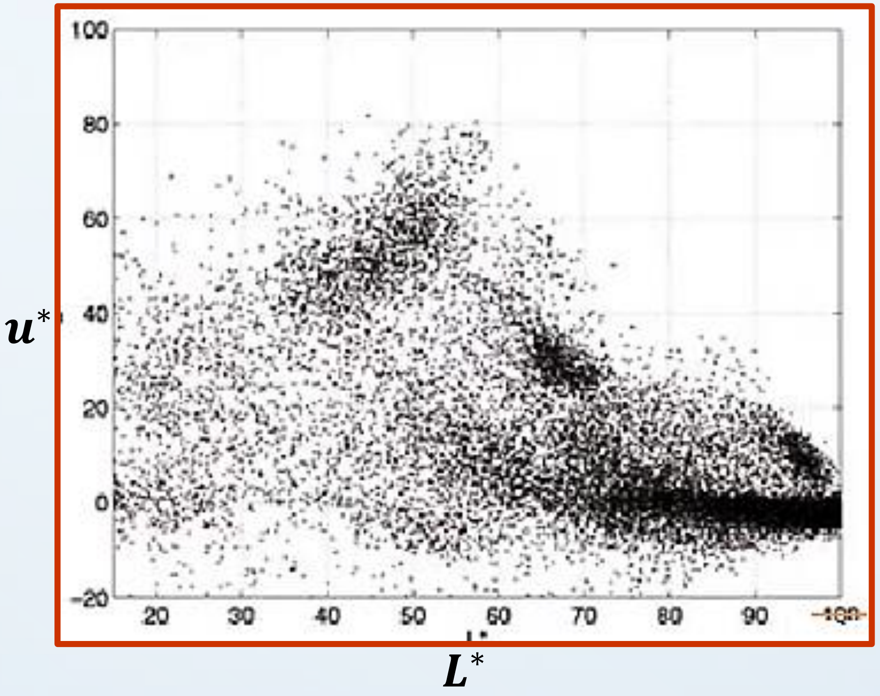
  You can choose the number of clusters by visually inspecting your data points, but you will soon realize that there is a lot of ambiguity in this process for all except the simplest data sets.

✓ The multiple sources of variation make it almost impossible to replicate; the only way to reliably reproduce a k-mean cluster is by ascription (using DFA or logistic regression).

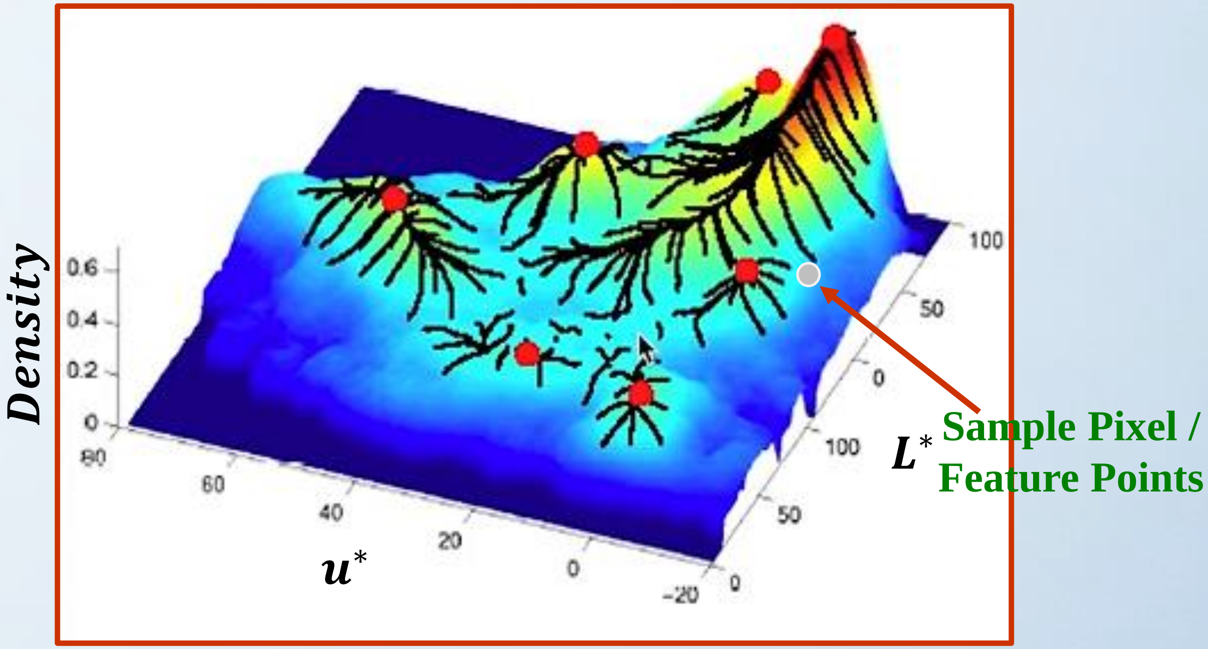
✓ Clustering outliers.

Centroids can be dragged by outliers, or outliers might get their own cluster instead of being ignored. Consider removing or clipping outliers before clustering.

Mean Shift Segmentation



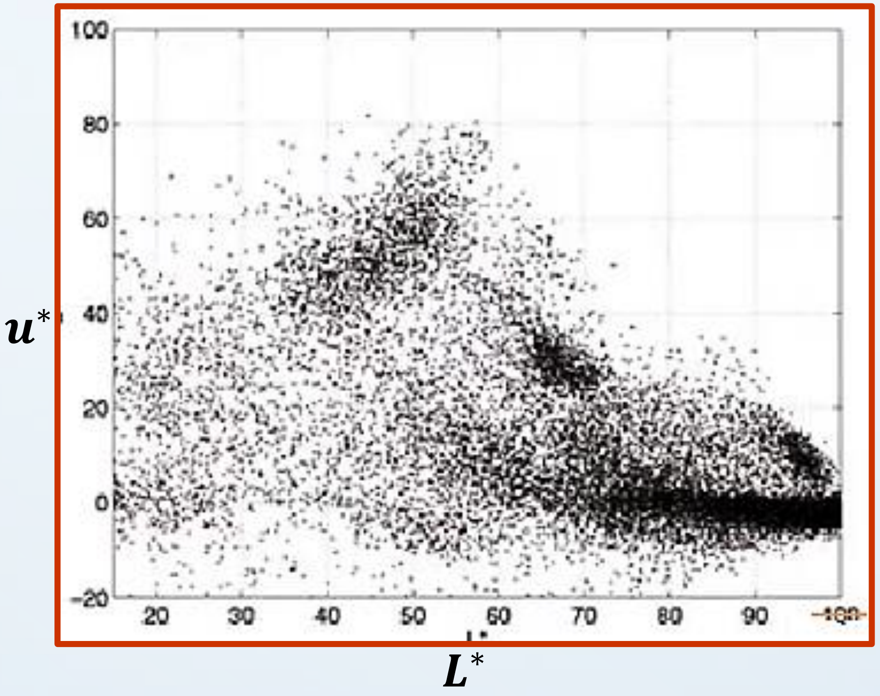
Pixel Feature Distribution



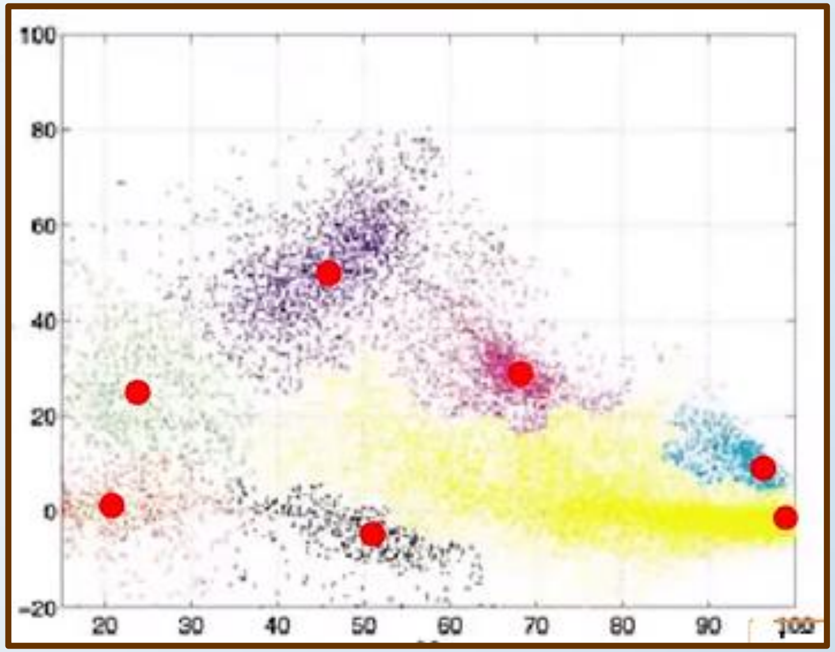
Normalized Density

- ✓ Each hill represents a cluster.
- ✓ Peak (mode) of hill represents “center” of the cluster.
- ✓ Each pixel climbs the steepest hill with in its neighborhood.
- ✓ Pixel assigned to the hill (cluster) it climb.

The Concept of Mean Shift Segmentation



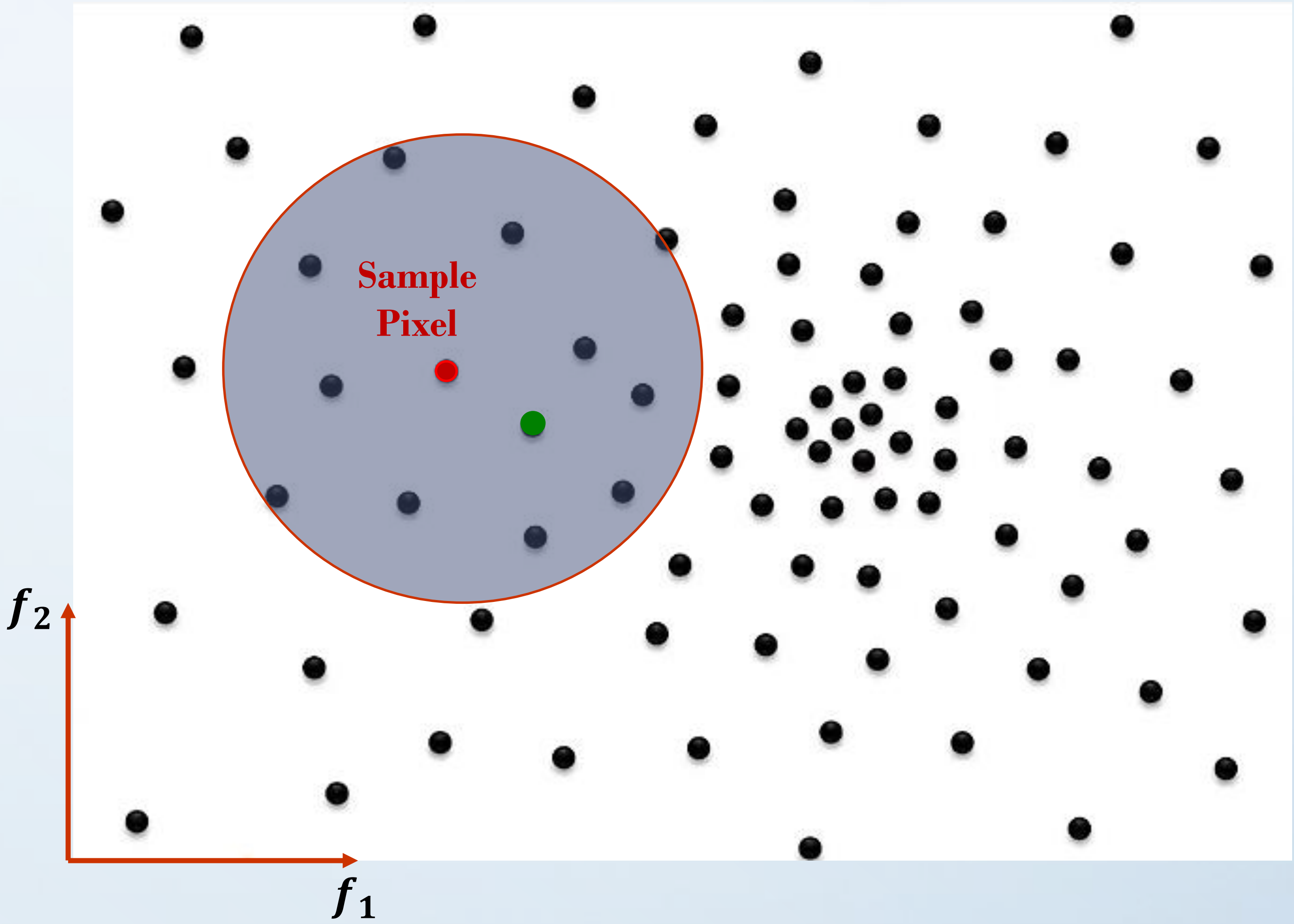
Pixel Feature Distribution



Leveled Clusters and Their Center

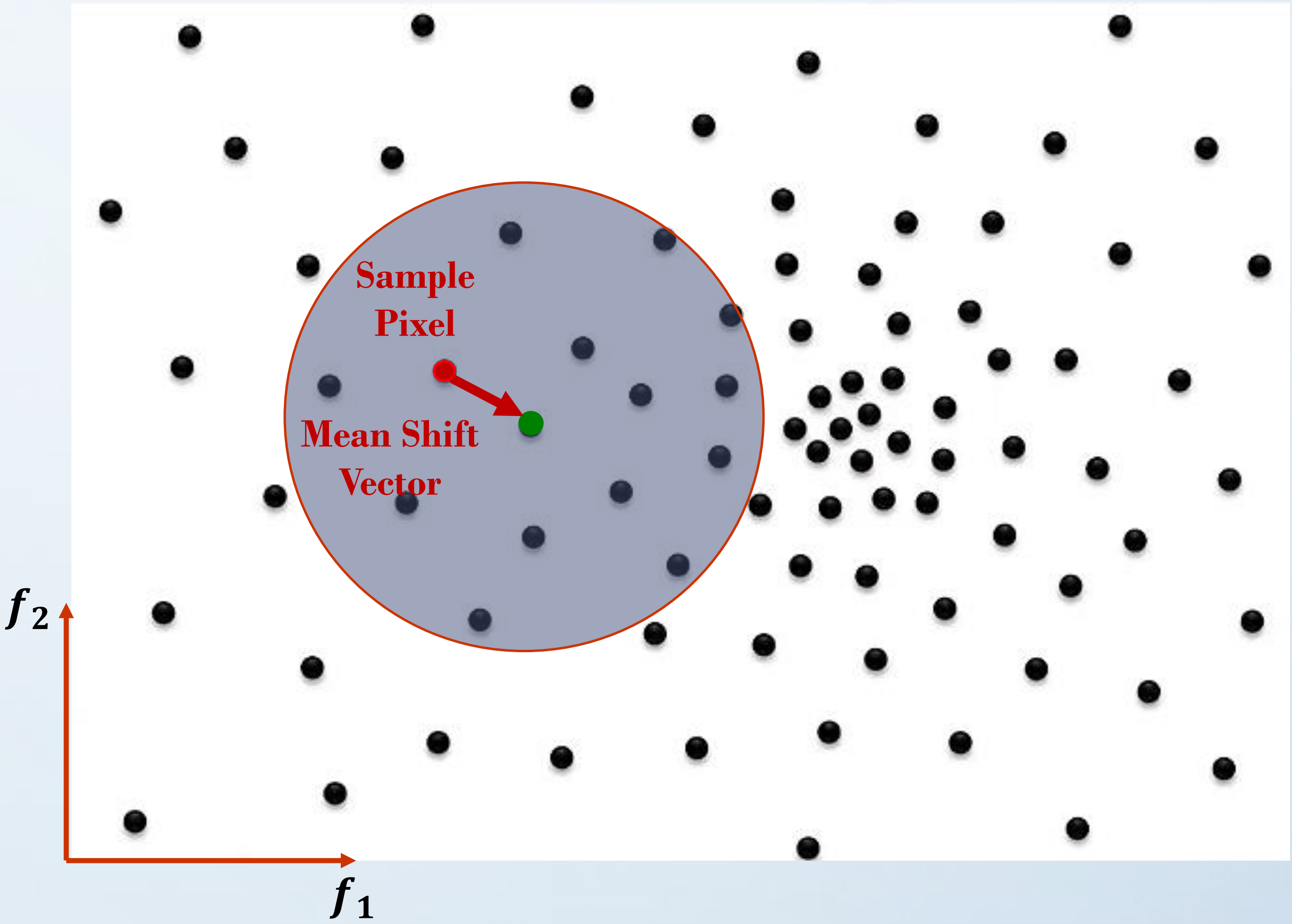
- ✓ Each hill represents a cluster.
- ✓ Peak (mode) of hill represents “center” of the cluster.
- ✓ Each pixel climbs the steepest hill with in its neighborhood.
- ✓ Pixel assigned to the hill (cluster) it climb.

Hill Climbing using Mean Shift



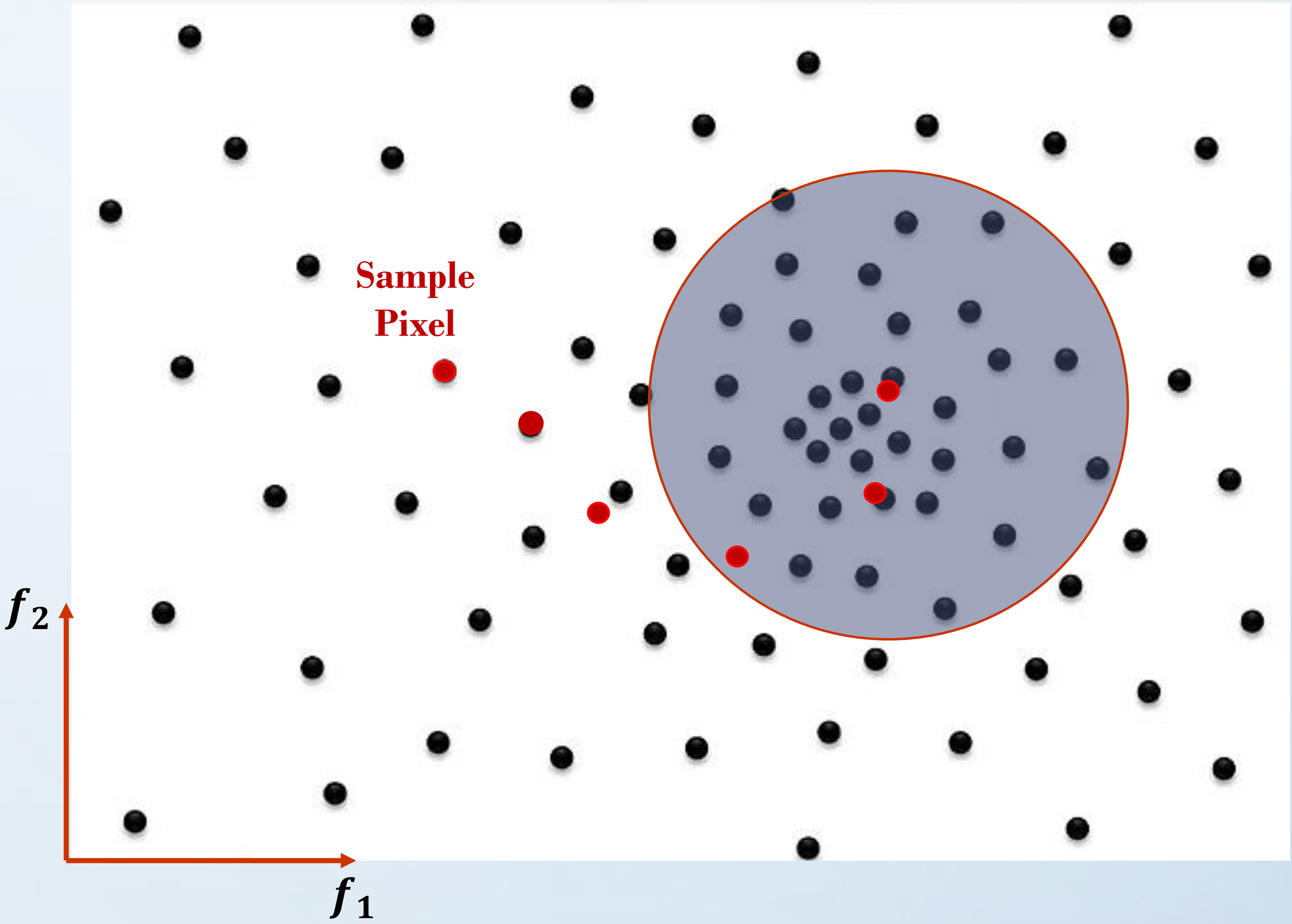
- ✓ Compute centroid with in the window W.

Hill Climbing using Mean Shift



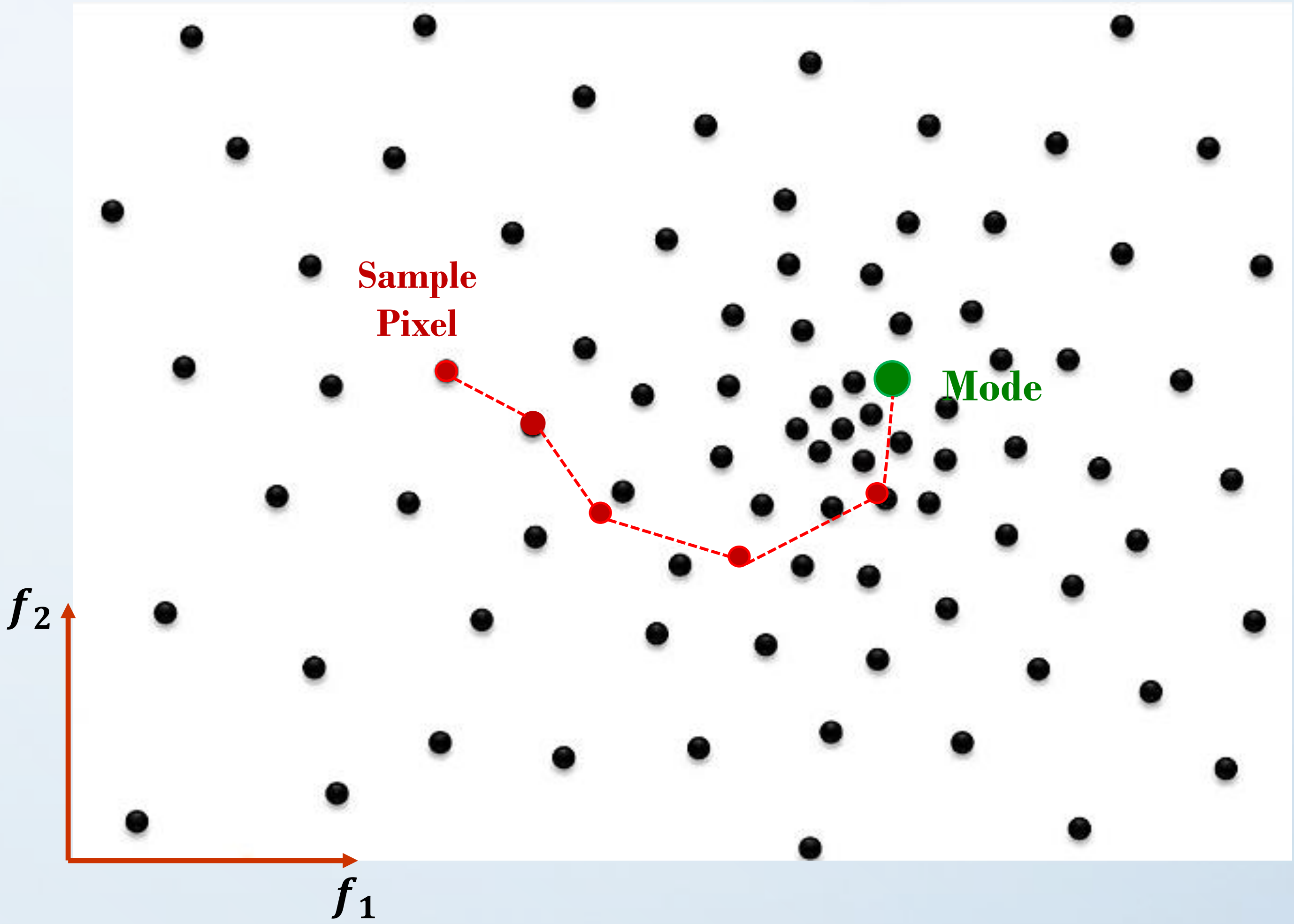
✓ Shift window to the centroid and the vector is called Mean Shift vector (Indicates the direction of mean shift).

Hill Climbing using Mean Shift



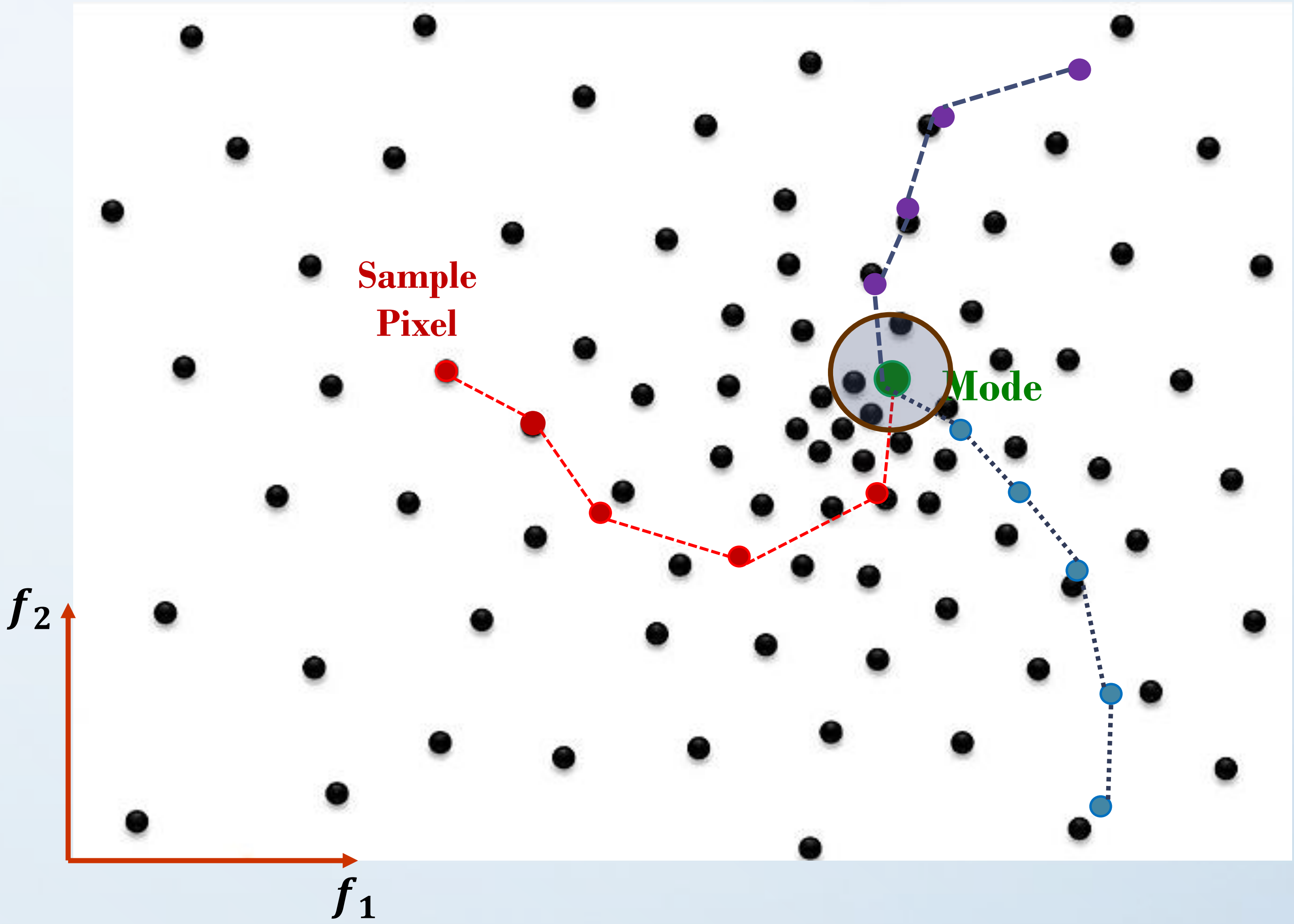
- ✓ Repeat the process until convergence.

Hill Climbing using Mean Shift



✓ Declare mode and assign it a cluster level. The convergence reaches when we re-compute the mean it doesn't changes any more.

Hill Climbing using Mean Shift



✓ Features that convergences to same mode.

Mean Shift Algorithm

Given: Distribution of N pixels in feature space.

Task: Find modes (clusters) of distribution.

Clustering Algorithm:

Step 1: Set $m_i = f_i$ as initial mean for each pixel i.

Step 2: Repeat the following for each mean m_i.

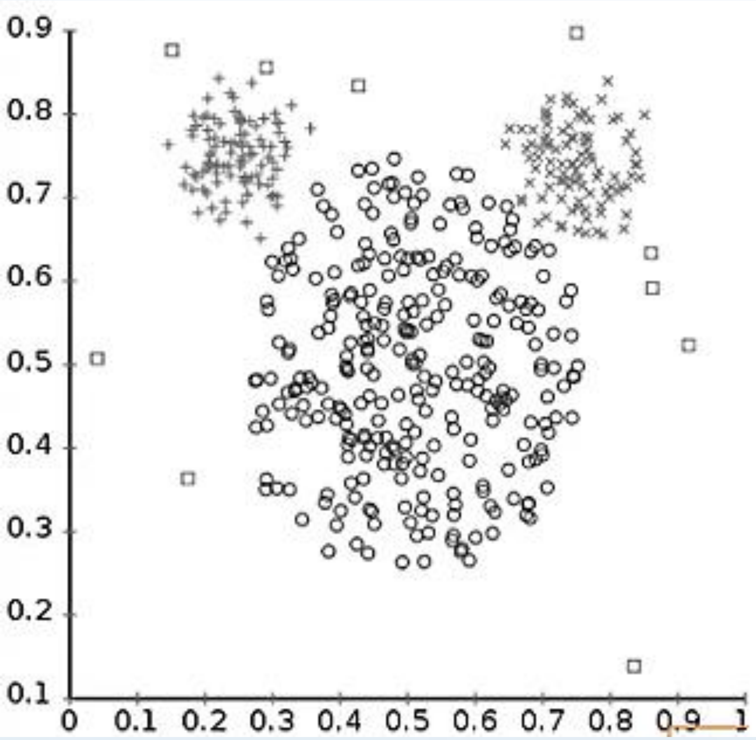
a. Place a window w_i around m_i.

b. Compute centroid m with in the window. Set $m_i = m$.

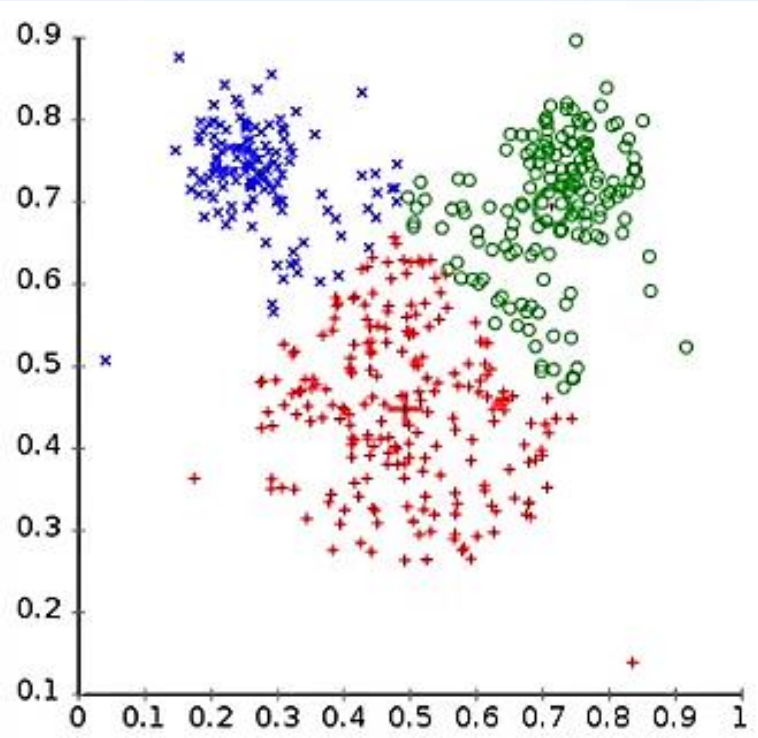
c. Stop if shift in mean m_i is less than a threshold ϵ. m_i is the mode.

Step 3: Level all pixels that have same mode as belonging to same cluster

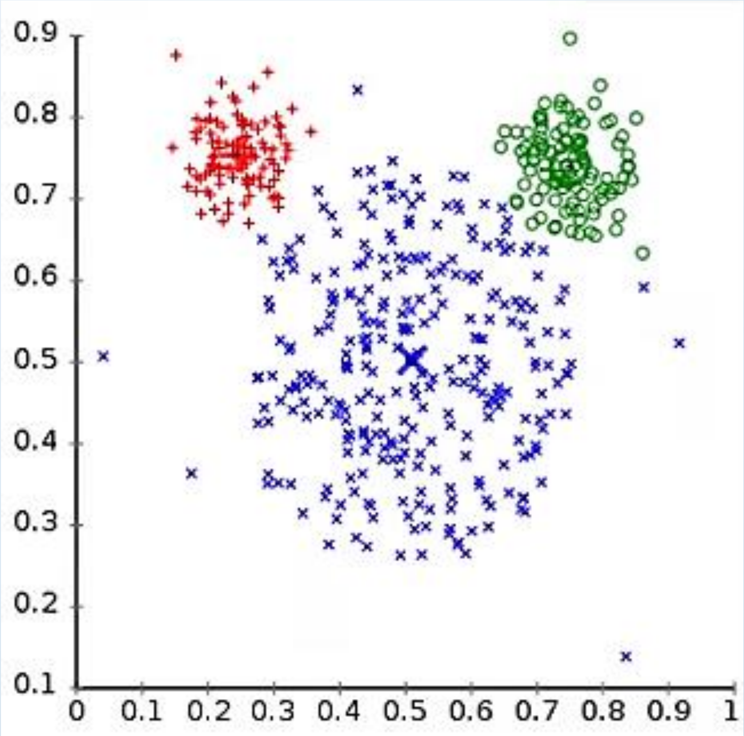
K-Mean vs Mean Shift



Original Data
[Samples are grouped together as a part of some experiment]
Reference Cluster



K-Mean (K=3) Clustering
[Does a fairly good job in grouping similar data together]
But mis-classifies many around the outliers



Mean Shift Clustering
[Classifies More accurately]

K-Mean vs Mean Shift: Natural Images

Input Image

K-Mean Clustering
[K = 16]
Feature consisting of five dimensional values R, G and B (Color Information) and Pixel Position

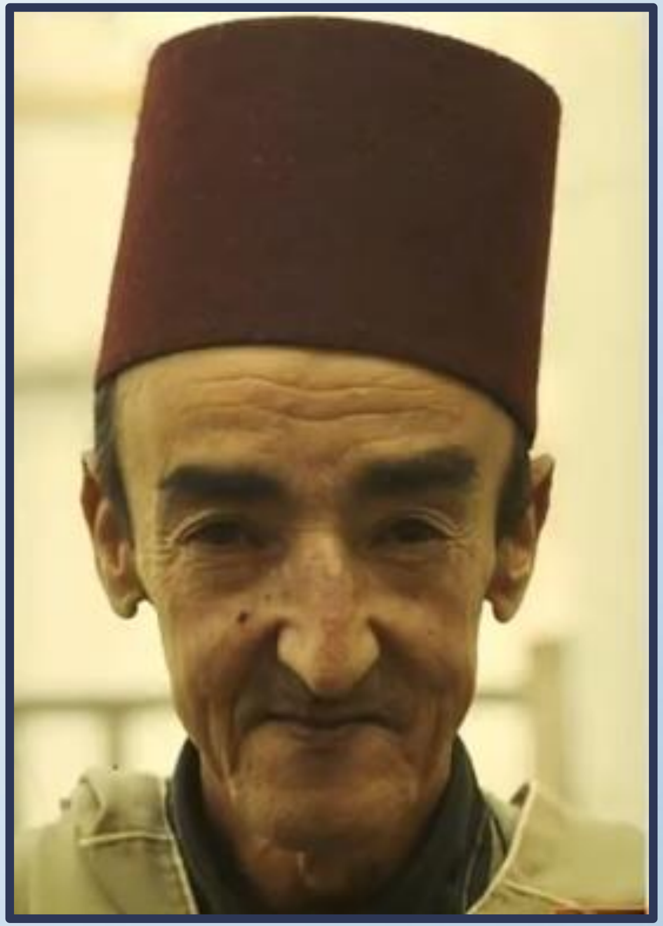
Mean Shift
[W=21]
No need to provide the number of clusters or K-Value

Graph based Image Segmentation

- ✓ Images are represented as graph and segmentation essentially is finding cuts in the graph.
- ✓ In recent years, graphs have emerged as a unified representation for image analysis and processing.
- ✓ Graph-based methods for:
  - ✓ Segmentation
  - ✓ Filtering
  - ✓ Classification and clustering

Images as Graph

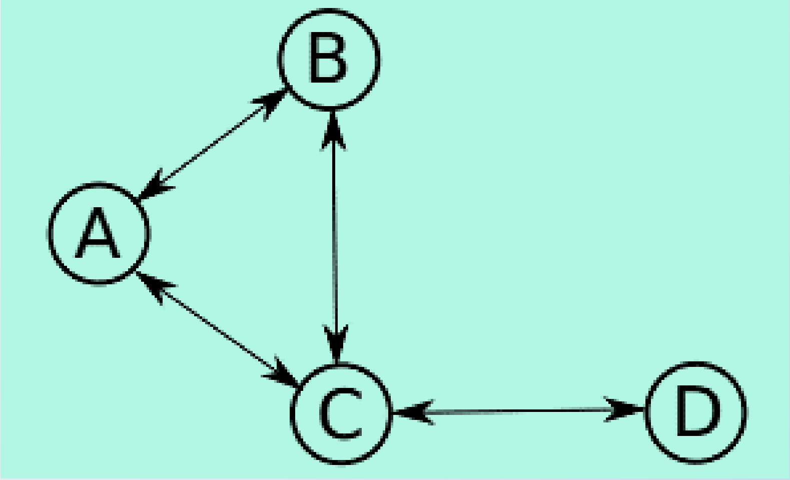
Each pixel a Vertex



Graph and Images: Basic Definition

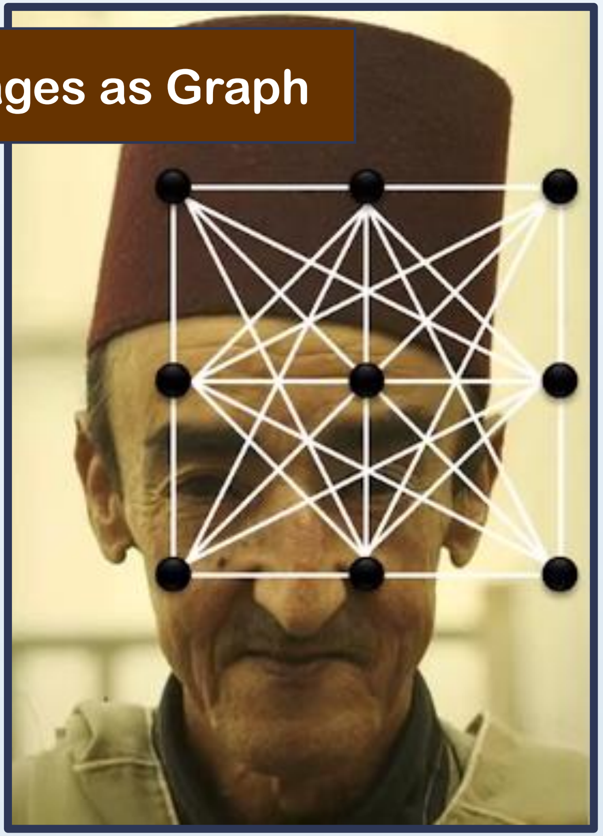
- ✓ A graph is a pair $G = (V, E)$, where.
  - ✓ V is a set.
  - ✓ E consists of pairs of elements in V .

- ✓ The elements of V are called the vertices of G.
- ✓ The elements of E are called the *edges* of G.



A drawing of an undirected graph with four vertices $\{A, B, C, D\}$ and four edges $\{e_{A,B}, e_{A,C}, e_{B,C}, e_{C,D}\}$.

Images as Graph



- ✓ Each edge is weighted by the affinity or **similarity** between its two vertices.

- ✓ Note vertices are pixels in an image.

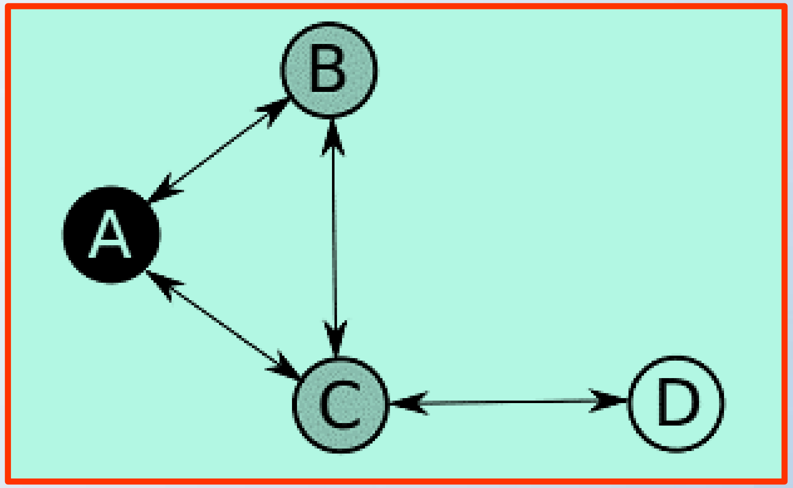
Graph and Images: Basic Definition

- ✓ Graph based image processing methods typically operate on pixel adjacency graphs, i.e., graphs whose vertex set is the set of image elements, and whose edge set is given by an adjacency relation on the image elements.

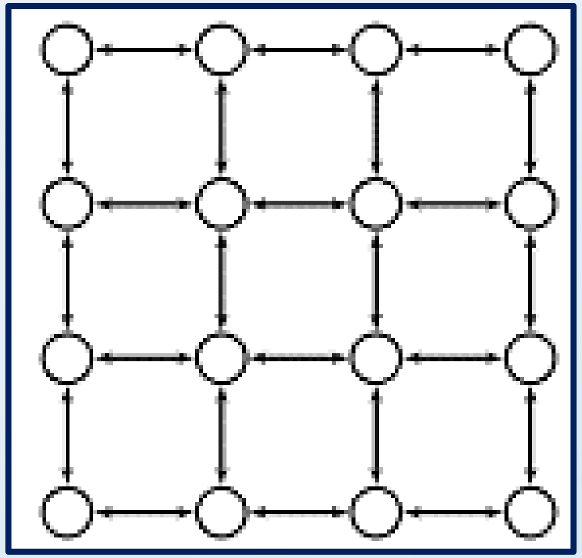
- ✓ Commonly, the edge set is defined as all vertices $\boldsymbol{v}, \boldsymbol{w}$ such that:

$$\boldsymbol{d}(\boldsymbol{v}, \boldsymbol{w}) < \boldsymbol{\rho}$$

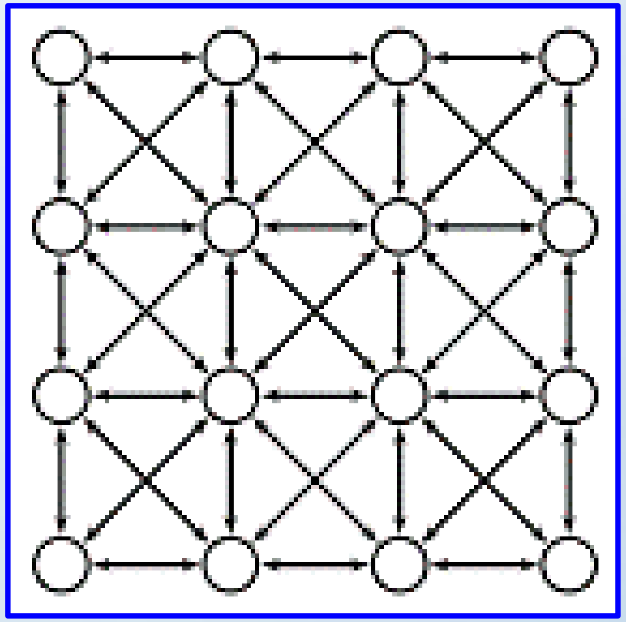
- ✓ This is called the *Euclidean adjacency relation*.



The set $\mathcal{N}(A) = \{B, C\}$ of vertices adjacent to A.



A 4-connected pixel adjacency graph.



A 8-connected pixel adjacency graph.

Graph and Images: Measuring Similarity

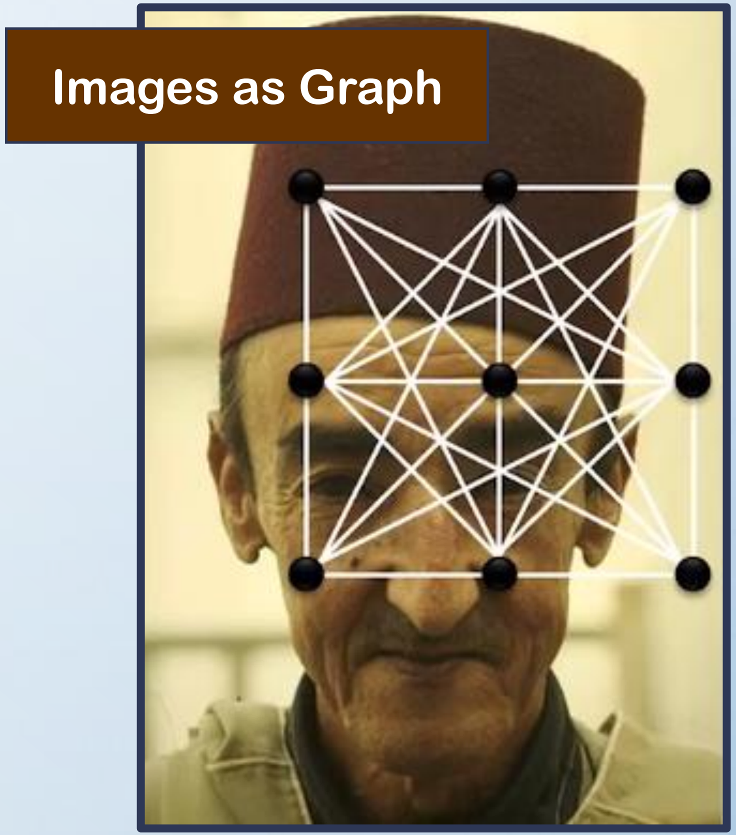
- ✓ Let i and j be two pixels whose features are f_i and f_j.

Pixel Dissimilarity:

$$S(\boldsymbol{f}_i, \boldsymbol{f}_j) = \sqrt{\left(\sum_k (\boldsymbol{f}_{ik} - \boldsymbol{f}_{jk})^2\right)}$$

- ✓ **Pixel Affinity**

$$w(i,j) = A(f_i, f_j) = \exp\left\{\frac{1}{2\sigma^2} S(\boldsymbol{f}_i, \boldsymbol{f}_j)\right\}$$

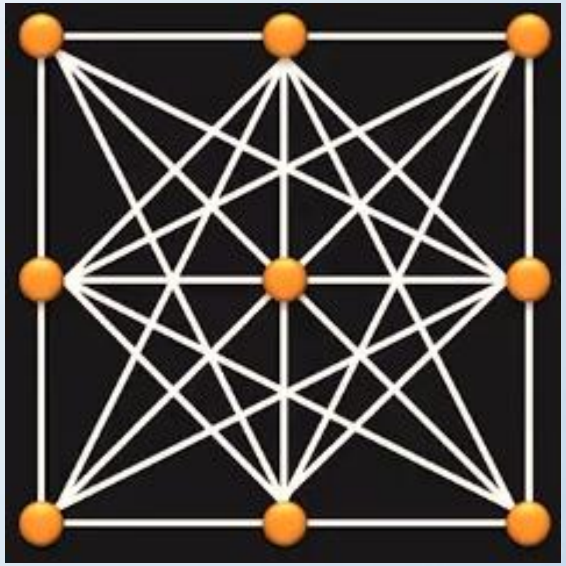


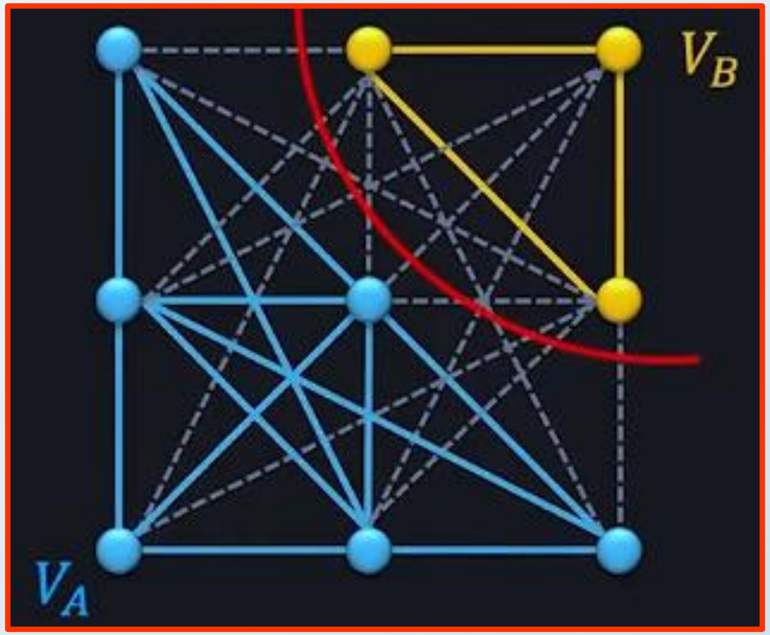
Smaller the dissimilarity. Larger the affinity.

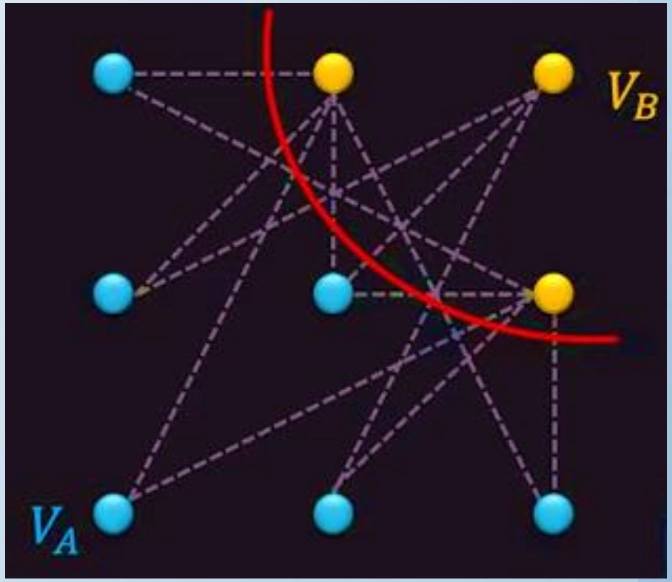
Graph Cut

- ✓ Cut $C = (V_A, V_B)$ is a partition of vertices $\boldsymbol{V}$ of a graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ in to two disjoint subsets V_A and V_B.

- ✓ **Cut Set**: Set of edges whose vertices are in different subsets of portions.





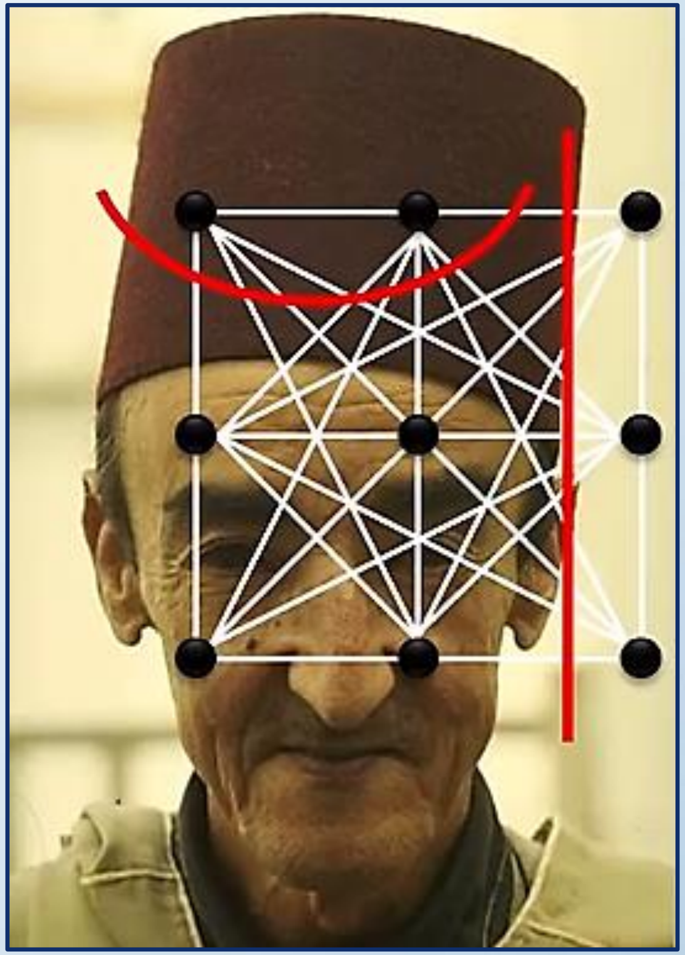


Cost of Cut: Sum of weights of cut-set edges:

$$cut(V_A, V_B) = \sum_{u \in V_A, v \in V_B} w(u, v)$$

Graph Cut Segmentation

✓ Cut $C = (V_A, V_B)$ is a partition of vertices $\boldsymbol{V}$ of a graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ in to two disjoint subsets V_A and V_B.



Criteria for Graph-Cut:

✓ A pair of vertices (pixels) with in a subgraph have high affinity.

✓ A pair of vertices (pixels) from two different subgraphs have low affinity.

That is minimizing the cost of cut also called **MIN-CUT**.

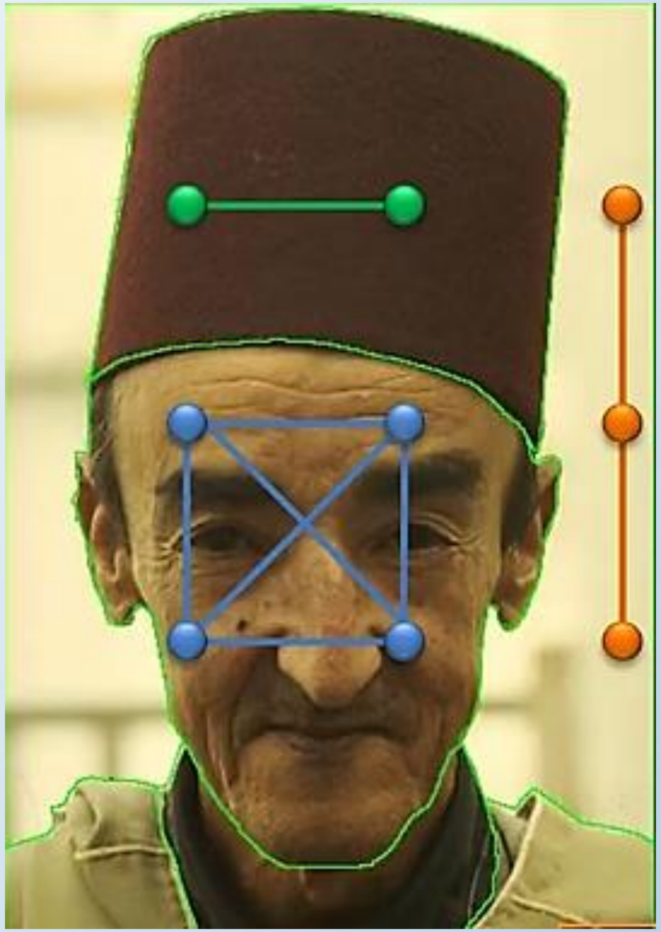
Graph Cut Segmentation

Criteria for Graph-Cut:

✓ A pair of vertices (pixels) with in a subgraph have high affinity.

✓ A pair of vertices (pixels) from two different subgraphs have low affinity.

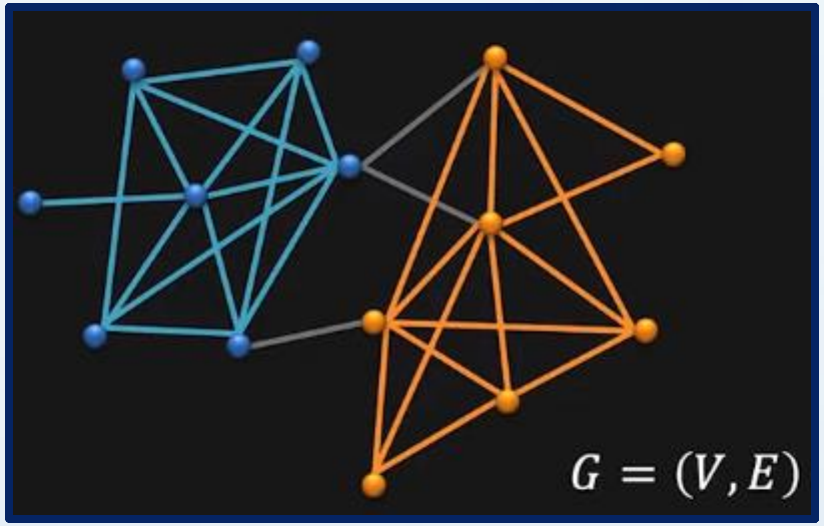
That is minimizing the cost of cut also called **MIN-CUT**.

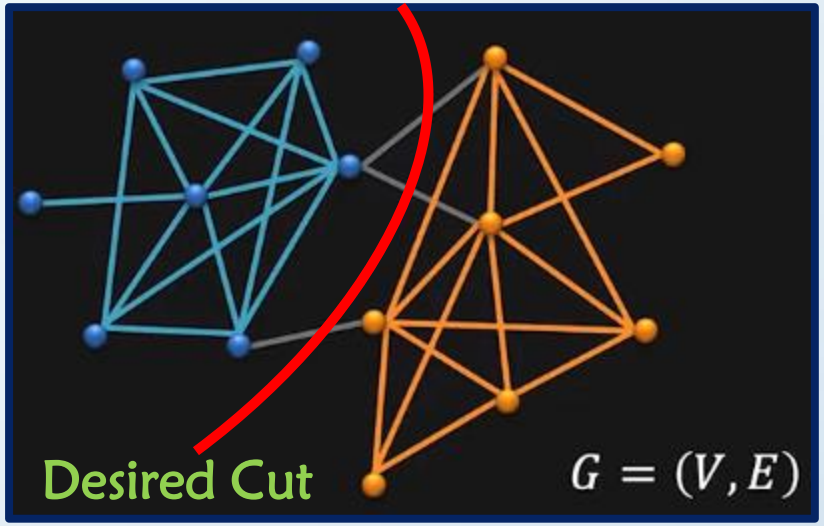


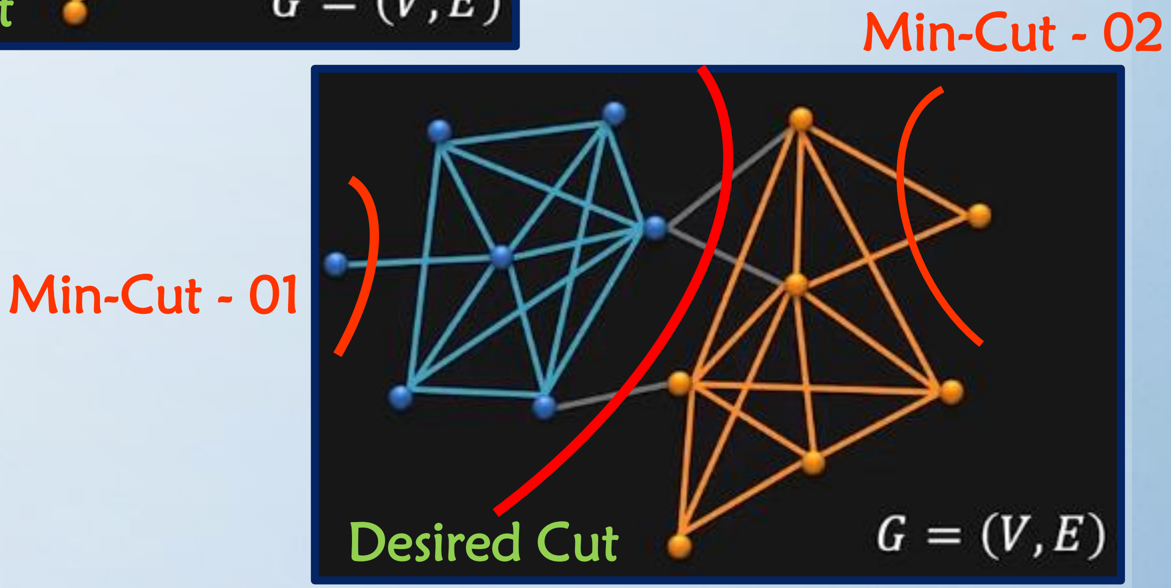
Each subgraph is an image segment.

Problem with Min-Cut

There is a bias to cut small, isolated segments.





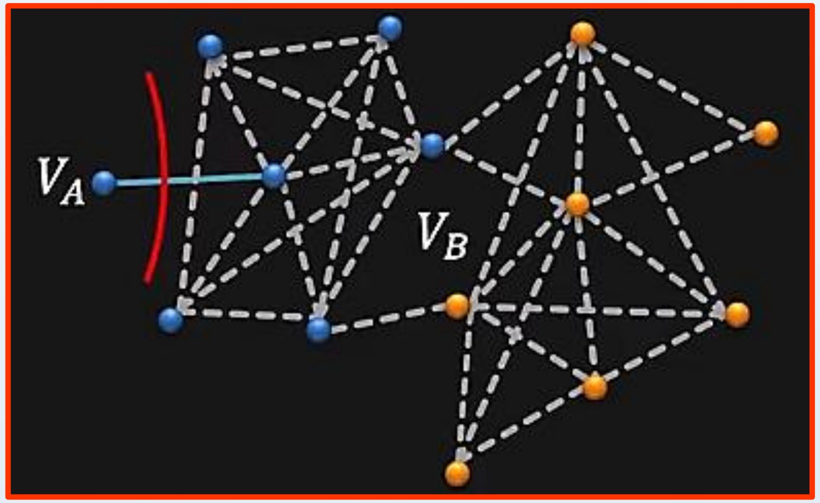


Solution: Normalize cut to favor larger subgraphs.

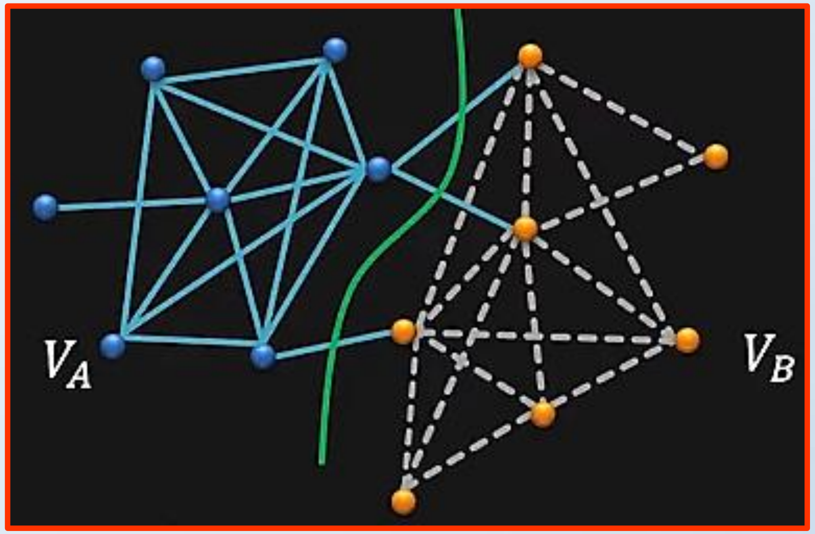
Measure of Subgraph Size

Compute how strongly vertices V_A are associated with vertices V.

$$assoc(V_A, V) = \sum_{u \in V_A, v \in V} w(u, v)$$



Weak Association (V_A, V)



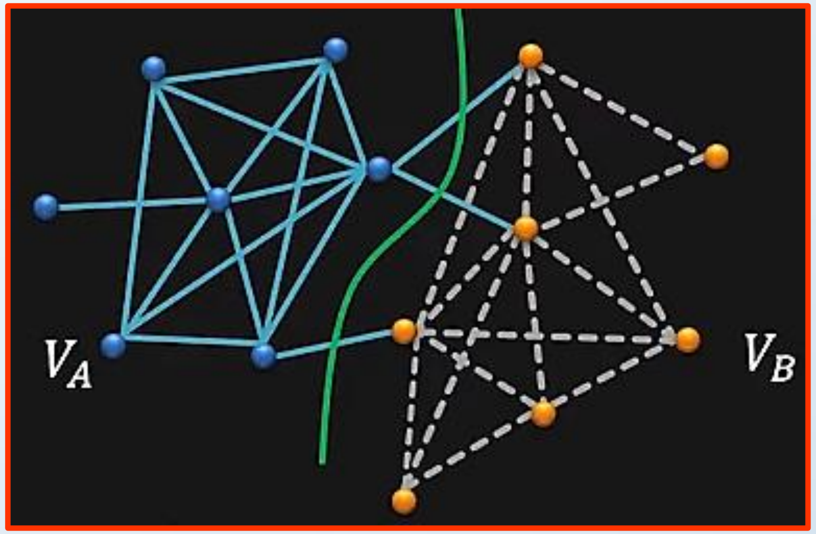
Strong Association (V_A, V)

$assoc()$ is the sum of weights of the solid edges.

Normalized Cut: NCut

Minimize the cost of normalized cost during partition.

$$NCut(V_A, V_B) = \frac{cut(V_A, V_B)}{assoc(V_A, V)} + \frac{cut(V_A, V_B)}{assoc(V_B, V)}$$



Minimizing Ncut has no known polynomial time solution.
It is NP-Complete.

First Eigenvector-based approximation exist [Shi, 2000].

Normalized Cut: NCut

Segmented Image
Feature (Brightness, Location)