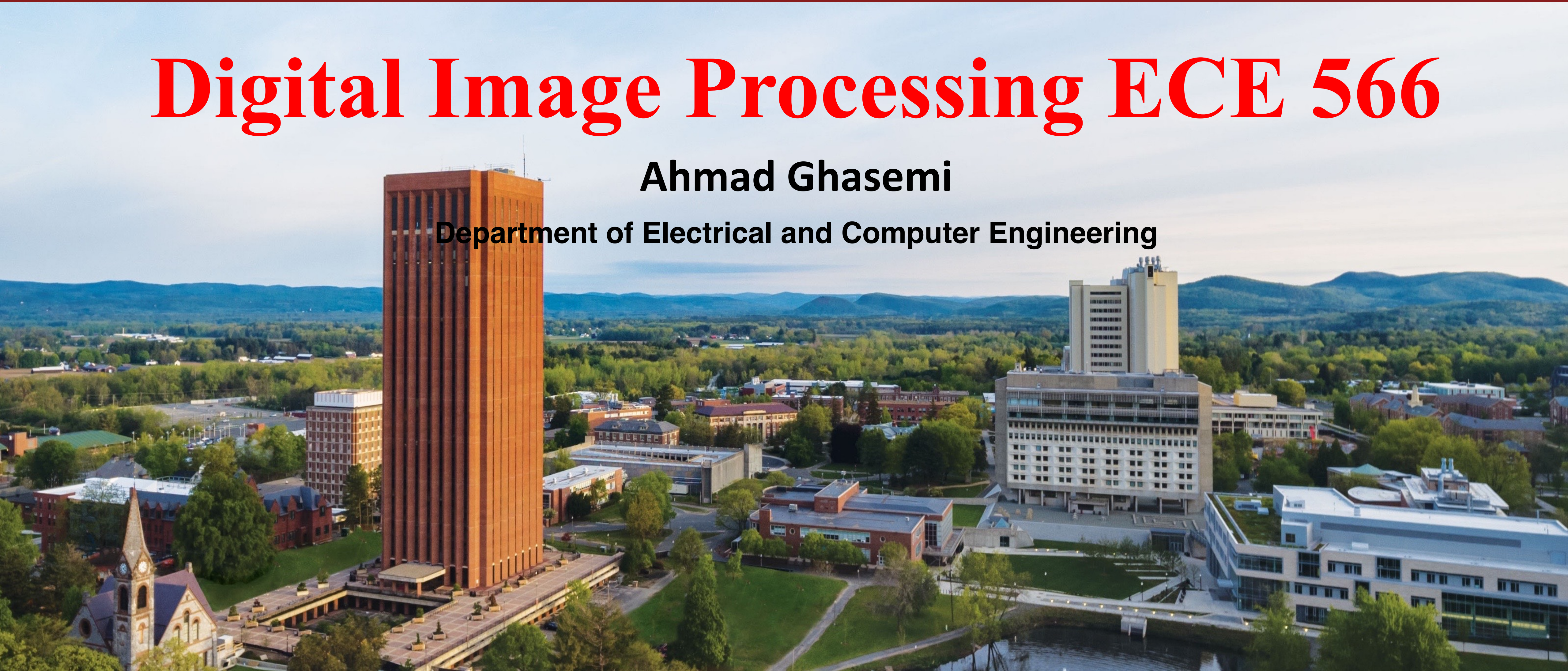
University of Massachusetts Amherst

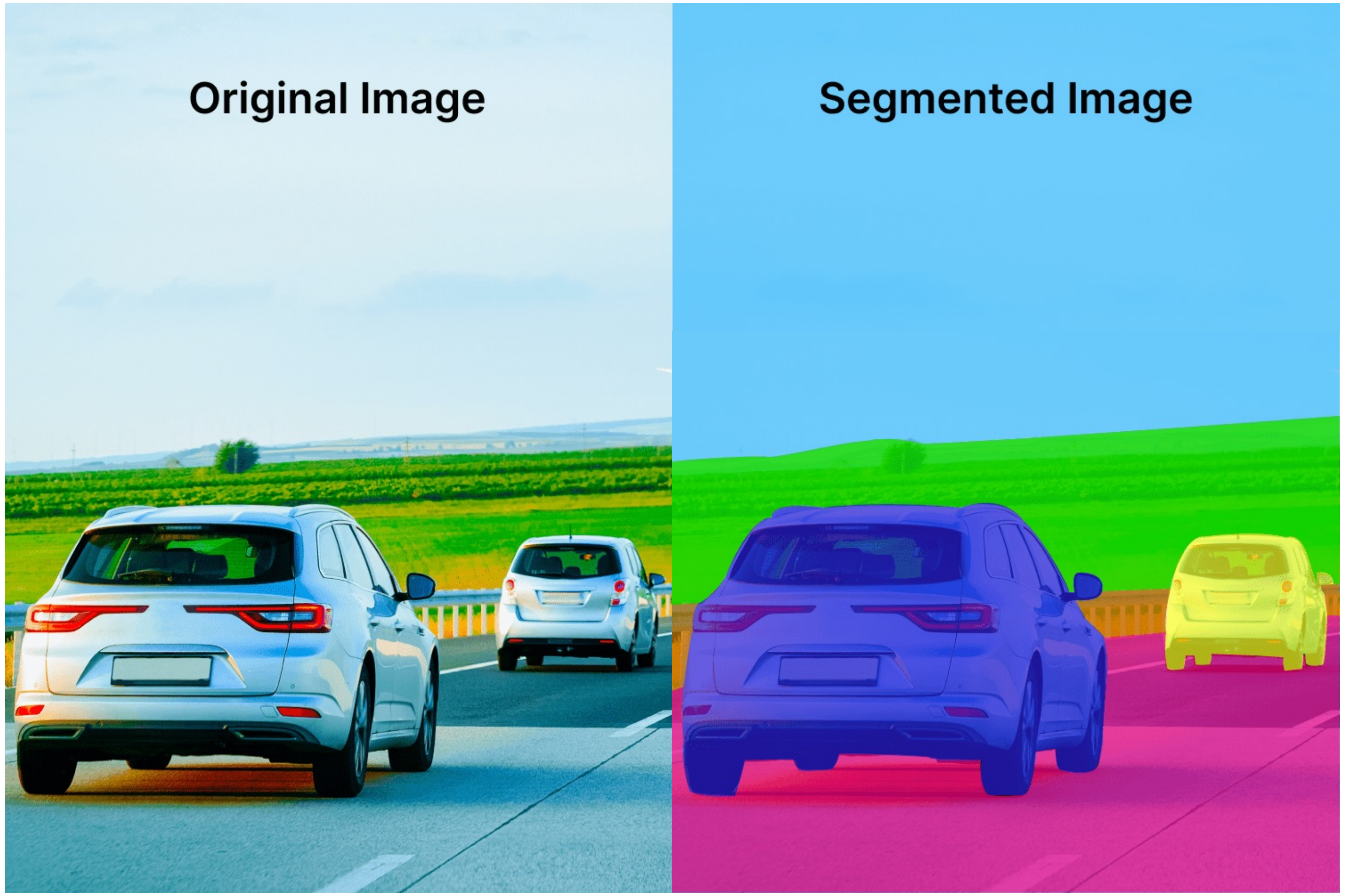


Digital Image Processing ECE 566

Ahmad Ghasemi

Department of Electrical and Computer Engineering

Image Segmentation



Definition of segmentation

Partition the region R occupied by an image into $R_1, \ldots, R_N$ such that

(a) $\bigcup_{i=1}^{n} R_i = R.$ → complete

Definition of segmentation

Partition the region R occupied by an image into $R_1, \ldots, R_N$ such that

(a) $\bigcup_{i=1}^{n} R_i = R.$ → complete

(b) R_i is a connected set, for $i = 0, 1, 2, \ldots, n.$ → Points in a region connected

Definition of segmentation

Partition the region R occupied by an image into $R_1, \ldots, R_N$ such that

(a) $\bigcup_{i=1}^{n} R_i = R.$ → complete

(b) R_i is a connected set, for $i = 0, 1, 2, \ldots, n.$ → Points in a region connected

(c) $R_i \cap R_j = \varnothing$ for all i and j, $i \neq j.$ → Regions are disjoint

Definition of segmentation

Partition the region R occupied by an image into $R_1, \ldots, R_N$ such that

(a) $\bigcup_{i=1}^{n} R_i = R.$ → complete

(b) R_i is a connected set, for $i = 0, 1, 2, \ldots, n.$ → Points in a region connected

(c) $R_i \cap R_j = \varnothing$ for all i and j, $i \neq j.$ → Regions are disjoint

(d) $Q(R_i) = \text{TRUE}$ for $i = 0, 1, 2, \ldots, n.$ → Points in a region share a similarity (e. g., Same color)

Definition of segmentation

Partition the region R occupied by an image into $R_1, \ldots, R_N$ such that

(a) $\bigcup_{i=1}^{n} R_i = R.$ → complete

(b) R_i is a connected set, for $i = 0, 1, 2, \ldots, n.$ → Points in a region connected

(c) $R_i \cap R_j = \varnothing$ for all i and j, $i \neq j.$ → Regions are disjoint

(d) $Q(R_i) = \text{TRUE}$ for $i = 0, 1, 2, \ldots, n.$ → Points in a region share a similarity (e. g., Same color)

(e) $Q(R_i \cup R_j) = \text{FALSE}$ for any adjacent regions R_i and R_j. → Points from different regions are dissimilar (e. g., Different color)

Segmentation Applications

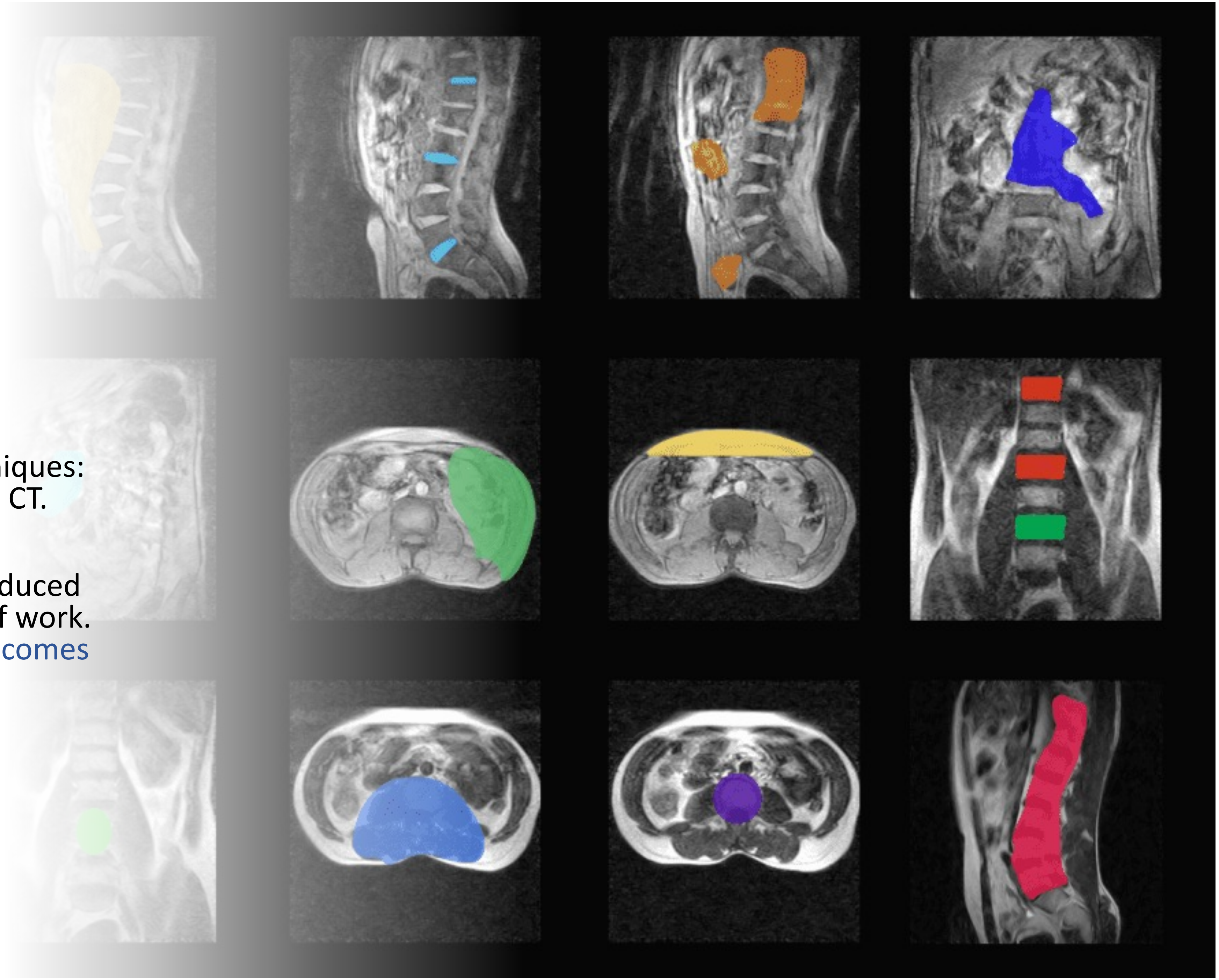
Medical imaging

- Diagnose diseases and injuries.

- Common medical imaging techniques: Radiography, MRI, ultrasound, and CT.

- Making sense of the images produced by these machines requires a lot of work. This is where image segmentation comes in.

  - ➢ Tumor detection
  - ➢ Brain segmentation
  - ➢ Diagnosis of diseases



Segmentation Applications

Autonomous vehicles

- Crucial for the proper functioning of autonomous vehicles.

- Three important types of image segmentation that enable autonomous vehicles to interpret and understand their environment

  - Object detection
  - lane segmentation
  - semantic segmentation

- Accurate image segmentation, autonomous vehicles can navigate safely and efficiently on the road.

Segmentation Applications

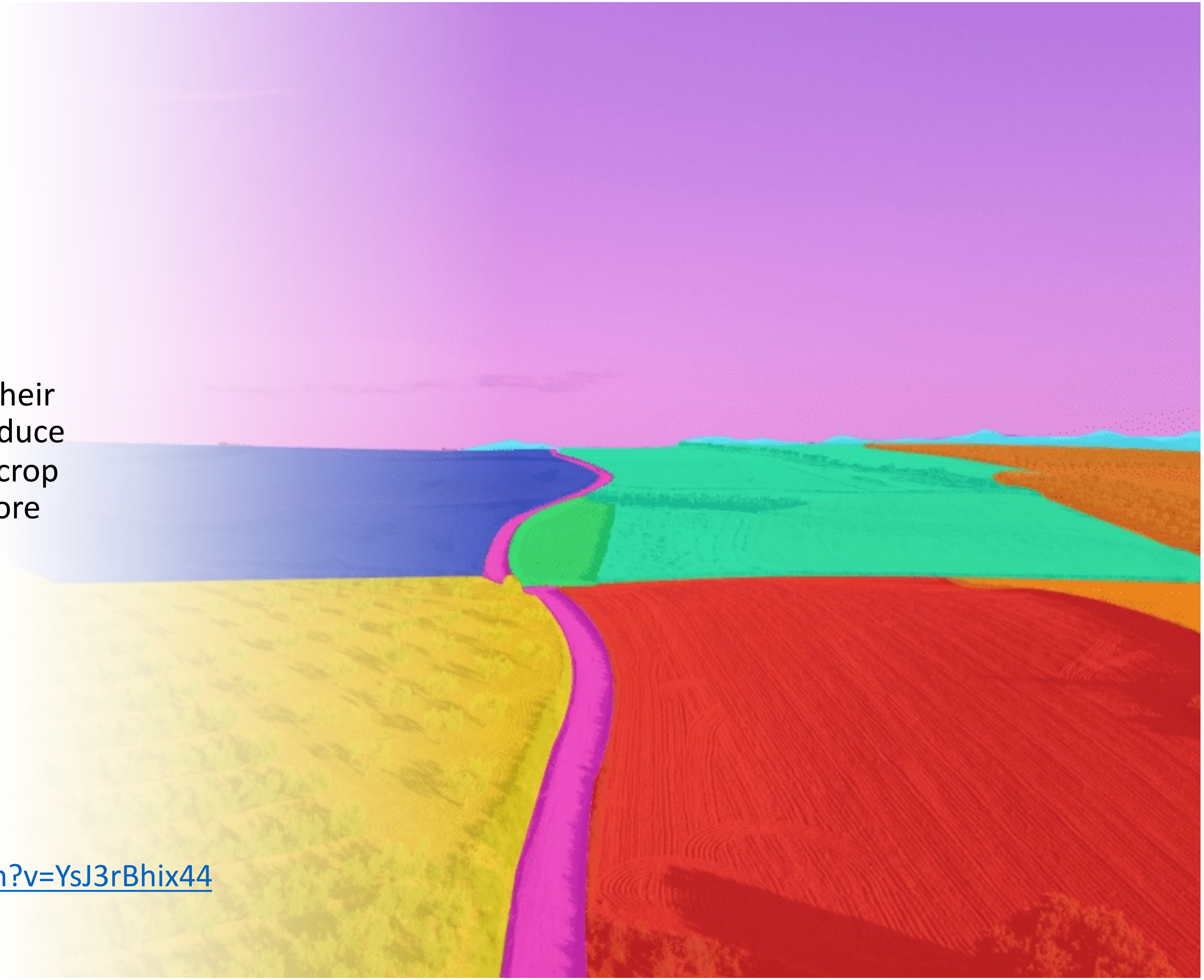
Agriculture

- It can help farmers improve their crop management practices, reduce production costs, and increase crop yield, ultimately leading to a more sustainable food supply.

  - ➢ Estimate crop yield
  - ➢ Detect weeds

https://youtu.be/wfObVKKJkE

https://www.youtube.com/watch?v=YsJ3rBhix44

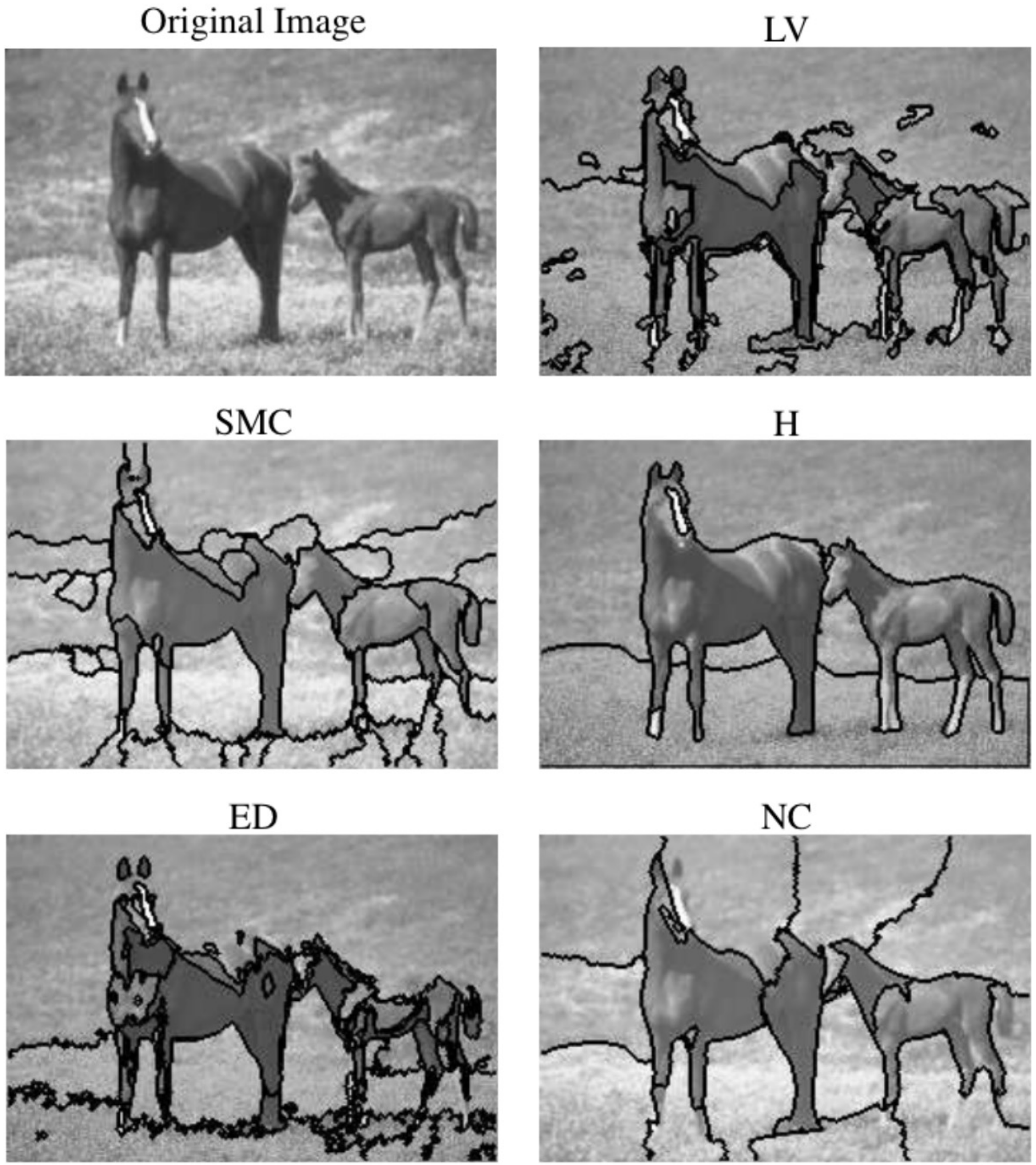


Segmentation Approaches

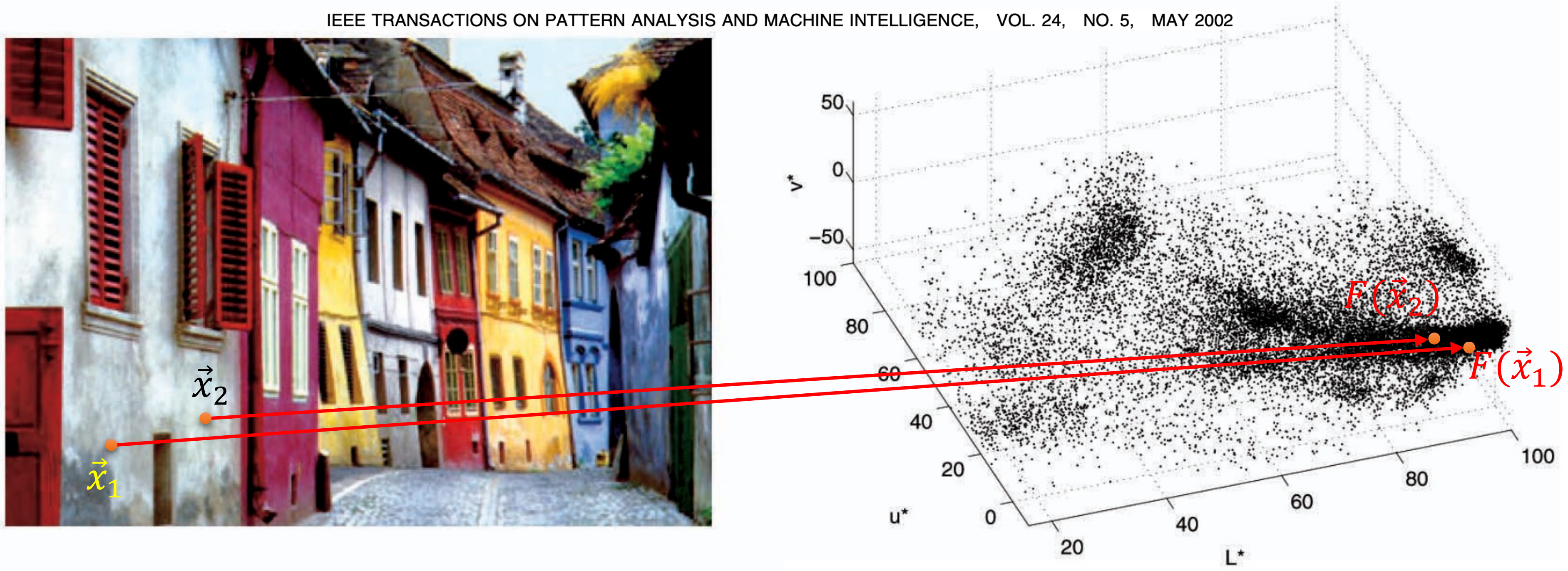
Edge based: boundaries of regions are different enough to allow boundary detection

Region based: partition into regions that are similar according to a criterion.

Segmentation (example): Which one is the best?



Segmentation as feature-based clustering



Image

Scatter plot of pixel feature vectors

General Approach:

1. Assign each pixel $\vec{x}_i$ to a *"feature vector"* $F(\vec{x}_i)$
2. Group pixels based on feature vector similarity

Assigning feature vectors to pixels

Given an image $I(\vec{x})$, consider feature vectors $\vec{F}(\vec{x})$ of the form

$$\vec{F}(\vec{x}) = \begin{pmatrix} \vec{x} \\ I(\vec{x}) \\ \vec{L}(\vec{x}) \end{pmatrix} .$$

Image position

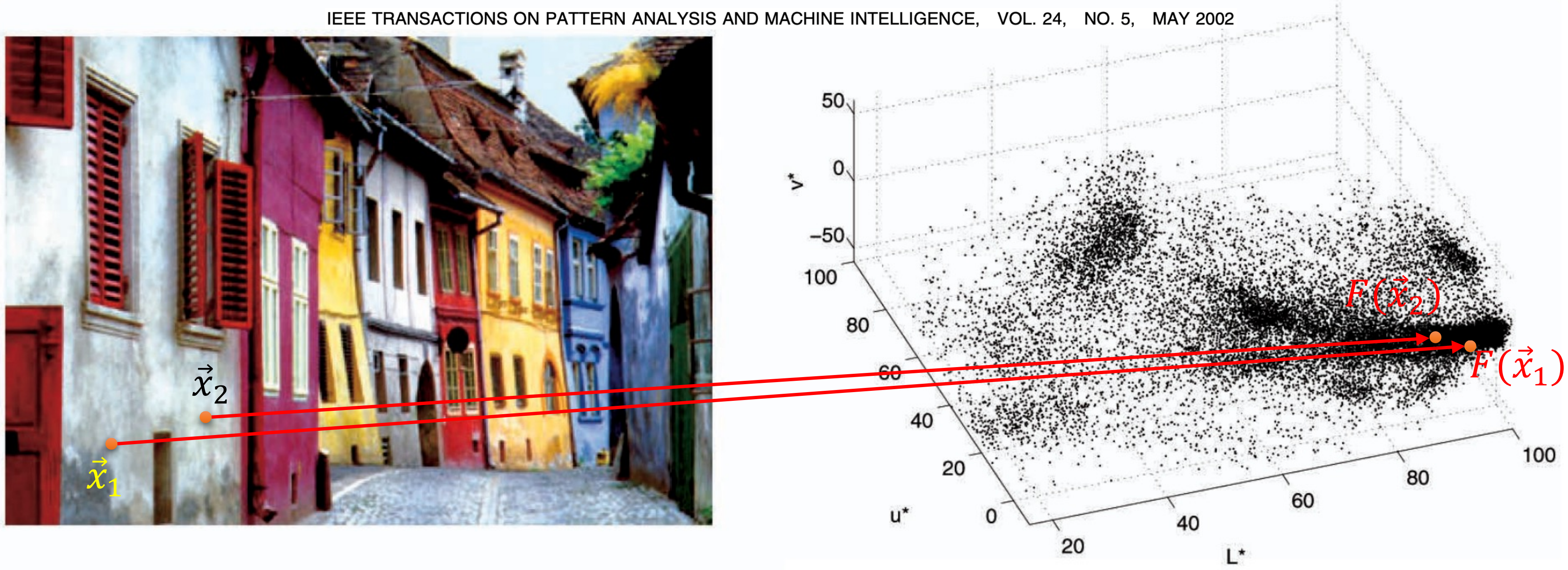
Intensity filter responses

Color filter responses

$\vec{L}(\vec{x})$ is a vector of local image features, perhaps bandpass filter responses. For color images, $\vec{F}(\vec{x})$ would also include information about the color at pixel $\vec{x}$.

To segment the image we might seek a clustering of the feature vectors $\vec{F}(\vec{x})$ observed in that image. A compact region of the image having a distinct gray-level or colour will correspond to a region in the feature space with a relatively high density of sampled feature vectors.

Probabilistic view of the segmentation problem



Image

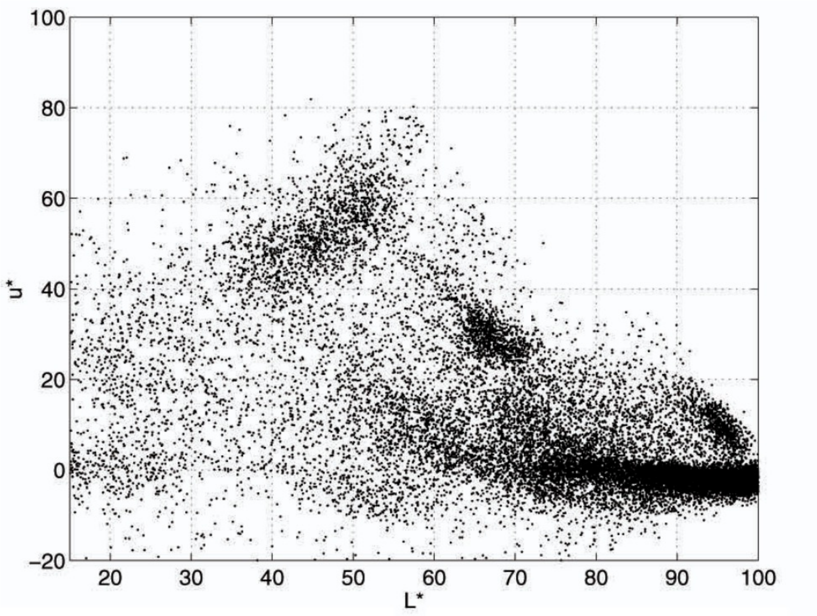
Scatter plot of pixel feature vectors

The feature vectors of pixels in a segment follow a “hidden” segment-specific distribution.

Task: To identify feature vectors arising from a same underlying distribution

Parametric vs. non-parametric methods

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 5, MAY 2002

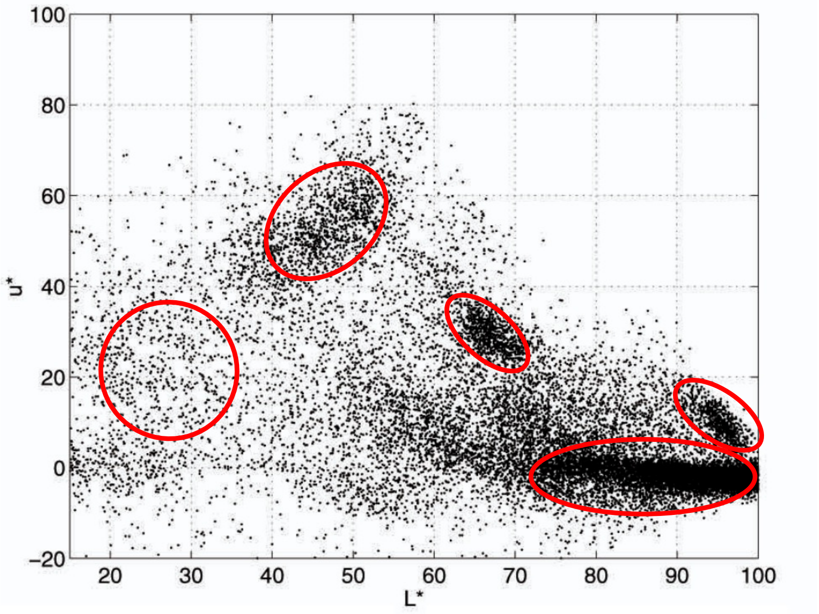


Mixture of Gaussians

Parametric: Assume a specific distribution for the feature vectors and estimate its parameters

Parametric vs. non-parametric methods

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 5, MAY 2002

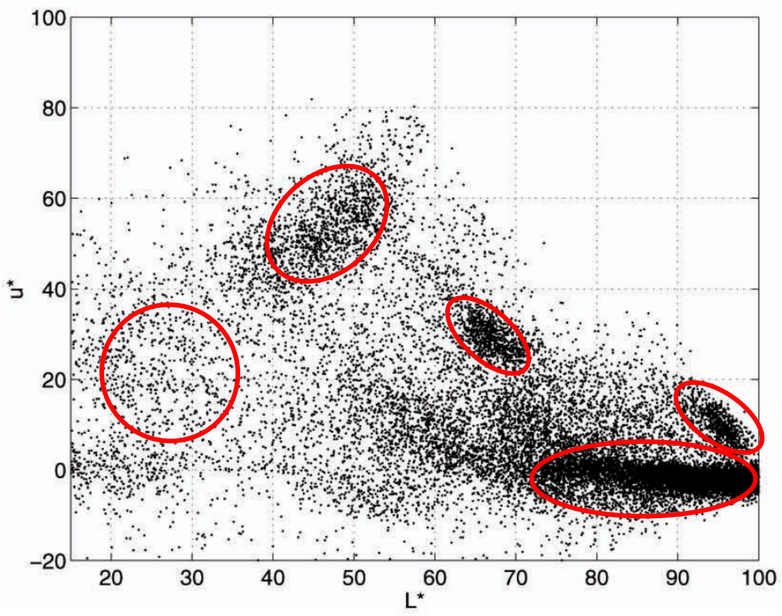


Mixture of Gaussians

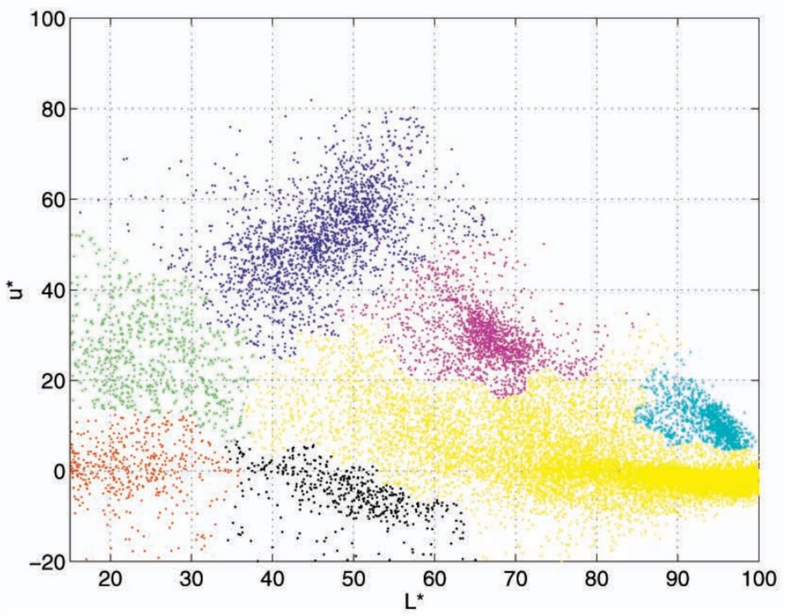
Parametric: Assume a specific distribution for the feature vectors and estimate its parameters

Parametric vs. non-parametric methods

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 5, MAY 2002



Mixture of Gaussians



Mean-shift segmentation

Parametric: Assume a specific distribution for the feature vectors and estimate its parameters

Non-parametric: Distribution defined implicitly using the feature vectors themselves

K-Means

- $x_1, \ldots, x_N$ are data points or vectors of observations
- Each observation (vector x_i) will be assigned to one and only one cluster
- $C(i)$ denotes cluster number for the i^{th} observation
- Dissimilarity measure: Euclidean distance metric
- K-means minimizes within-cluster point scatter:

$$W(C) = \frac{1}{2} \sum_{k=1}^{K} \sum_{C(i)=k} \sum_{C(j)=k} \left\| x_i - x_j \right\|^2 = \sum_{k=1}^{K} N_k \sum_{C(i)=k} \left\| x_i - m_k \right\|^2$$

where

m_k is the mean vector of the k^{th} cluster

N_k is the number of observations in k^{th} cluster

University of Massachusetts Amherst

Within and Between Cluster Criteria

Let's consider total point scatter for a set of N data points:

$$T = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} d(x_i, x_j)$$

Distance between two points

Within and Between Cluster Criteria

Let's consider total point scatter for a set of *N* data points:

$$T = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} d(x_i, x_j)$$

Distance between two points

T can be re-written as:

$$T = \frac{1}{2}\sum_{k=1}^{K}\sum_{C(i)=k}\left(\sum_{C(j)=k} d(x_i, x_j) + \sum_{C(j)\neq k} d(x_i, x_j)\right)$$
$$= W(C) + B(C)$$

Where,

$$W(C) = \frac{1}{2}\sum_{k=1}^{K}\sum_{C(i)=k}\sum_{C(j)=k} d(x_i, x_j)$$

Within cluster scatter

$$B(C) = \frac{1}{2}\sum_{k=1}^{K}\sum_{C(i)=k}\sum_{C(j)\neq k} d(x_i, x_j)$$

Between cluster scatter

Within and Between Cluster Criteria

Let' s consider total point scatter for a set of *N* data points:

$$T = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} d(x_i, x_j)$$

Distance between two points

T can be re-written as:

$$T = \frac{1}{2}\sum_{k=1}^{K}\sum_{C(i)=k}\Big(\sum_{C(j)=k} d(x_i, x_j) + \sum_{C(j)\neq k} d(x_i, x_j)\Big)$$
$$= W(C) + B(C)$$

Where,

$$W(C) = \frac{1}{2}\sum_{k=1}^{K}\sum_{C(i)=k}\sum_{C(j)=k} d(x_i, x_j)$$

Within cluster scatter

$$B(C) = \frac{1}{2}\sum_{k=1}^{K}\sum_{C(i)=k}\sum_{C(j)\neq k} d(x_i, x_j)$$

Between cluster scatter

If *d* is square Euclidean distance, then

$$W(C) = \sum_{k=1}^{K} N_k \sum_{C(i)=k} \left\| x_i - m_k \right\|^2$$

and $B(C) = \sum_{k=1}^{K} N_k \left\| m_k - m \right\|^2$

Grand mean

Within and Between Cluster Criteria

Let's consider total point scatter for a set of N data points:

$$T = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} d(x_i, x_j)$$

Distance between two points

T can be re-written as:

$$T = \frac{1}{2}\sum_{k=1}^{K}\sum_{C(i)=k}\left(\sum_{C(j)=k} d(x_i, x_j) + \sum_{C(j)\neq k} d(x_i, x_j)\right)$$
$$= W(C) + B(C)$$

Where,
$$W(C) = \frac{1}{2}\sum_{k=1}^{K}\sum_{C(i)=k}\sum_{C(j)=k} d(x_i, x_j)$$

Within cluster scatter

$$B(C) = \frac{1}{2}\sum_{k=1}^{K}\sum_{C(i)=k}\sum_{C(j)\neq k} d(x_i, x_j)$$

Between cluster scatter

If d is square Euclidean distance, then

$$W(C) = \sum_{k=1}^{K} N_k \sum_{C(i)=k} \|x_i - m_k\|^2$$

and $$B(C) = \sum_{k=1}^{K} N_k \|m_k - m\|^2$$

Grand mean

Minimizing $W(C)$ is equivalent to maximizing $B(C)$

K-Means Clustering: Expectation-Maximization

- Find values for $\{r_{nk}\}$ and $\{\mu_k\}$ to minimize:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \mu_k\|^2 \qquad (9.1)$$

K-Means Clustering: Expectation-Maximization

- Find values for $\{r_{nk}\}$ and $\{\mu_k\}$ to minimize:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \mu_k\|^2 \tag{9.1}$$

- Iterative procedure:
  1. Minimize J w.r.t. r_{nk}, keep μ_k fixed (Expectation)

$$r_{nk} = \begin{cases} 1 & \text{if } k = \arg\min_j \|\mathbf{x}_n - \mu_k\|^2 \\ 0 & \text{otherwise} \end{cases} \tag{9.2}$$

K-Means Clustering: Expectation-Maximization

- Find values for $\{r_{nk}\}$ and $\{\mu_k\}$ to minimize:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \mu_k\|^2 \qquad (9.1)$$

- Iterative procedure:
  1. Minimize J w.r.t. r_{nk}, keep μ_k fixed (Expectation)

$$r_{nk} = \begin{cases} 1 & \text{if } k = \arg\min_j \|\mathbf{x}_n - \mu_k\|^2 \\ 0 & \text{otherwise} \end{cases} \qquad (9.2)$$

  2. Minimize J w.r.t. μ_k, keep r_{nk} fixed (Maximization)

$$2 \sum_{n=1}^{N} r_{nk}(x_n - \mu_k) = 0 \qquad (9.3)$$

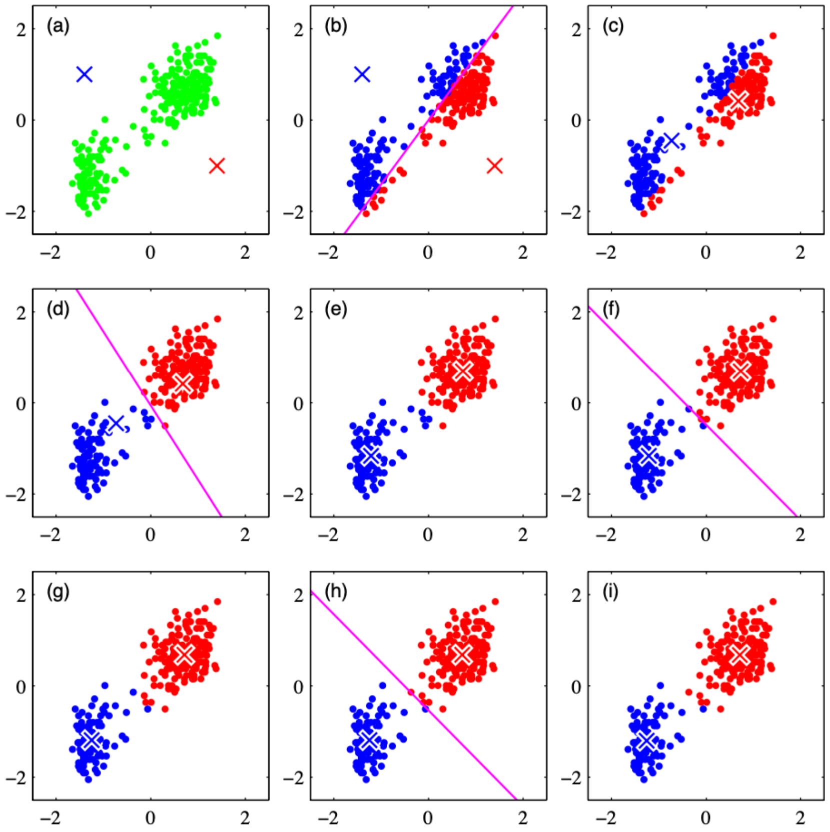
$$\mu_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}} \qquad (9.4)$$

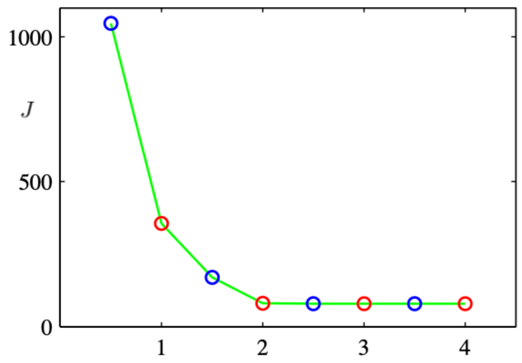
K-Means Clustering: Distortion Measure

- Dataset $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$
- Partition in K clusters
- Cluster prototype: μ_k
- Binary indicator variable, 1-of-K Coding scheme
  $r_{nk} \in \{0, 1\}$
  $r_{nk} = 1$, and $r_{nj} = 0$ for $j \neq k$.
  Hard assignment.
- Distortion measure

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \mu_k\|^2 \tag{9.1}$$

K-Means Clustering: Example





- Each E or M step reduces the value of the objective function J
- Convergence to a global or local maximum

K-Means Clustering: Choice of K

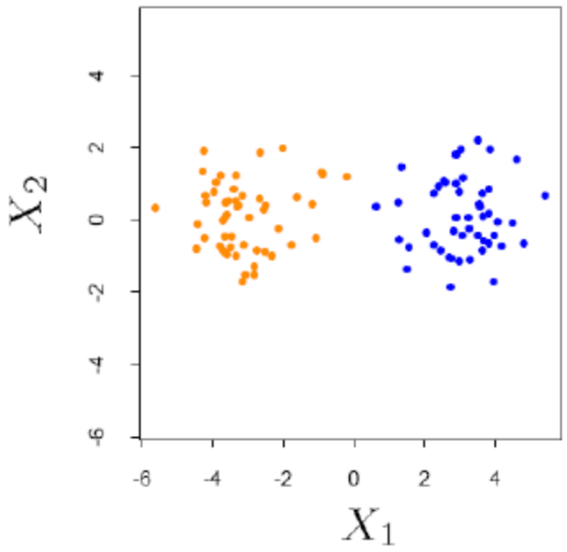
- Can $W_K(C)$, *i.e.*, the within cluster distance as a function of K serve as any indicator?
- Note that $W_K(C)$ decreases monotonically with increasing K. That is the within cluster scatter decreases with increasing centroids.

K-Means Clustering: Choice of K

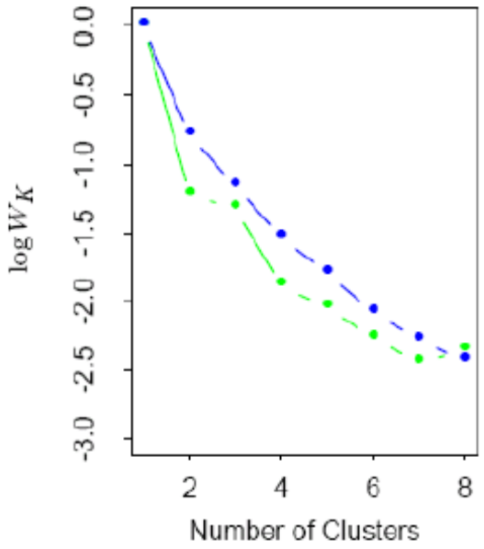
- Can $W_K(C)$, *i.e.*, the within cluster distance as a function of K serve as any indicator?
- Note that $W_K(C)$ decreases monotonically with increasing K. That is the within cluster scatter decreases with increasing centroids.
- Instead look for elbow by using successive difference between $W_K(C)$:

$$\{W_K - W_{K+1} : K < K^*\} >> \{W_K - W_{K+1} : K \geq K^*\}$$

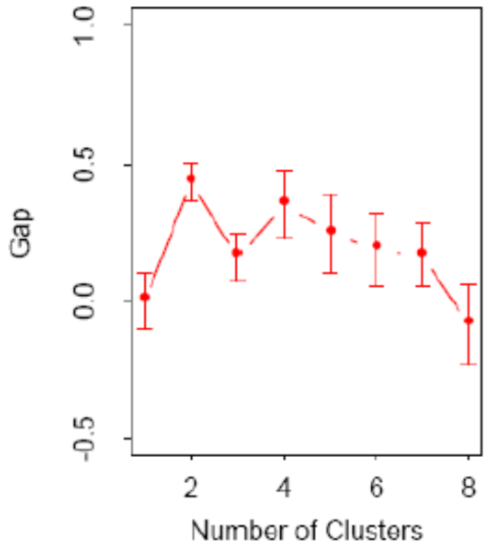
K-Means Clustering: Choice of K



Data points simulated from two pdfs



$\text{Log}(W_K)$ curve



Gap curve

$$Gap(k) = \frac{1}{B}\sum_{b=1}^{B} log(W^*_{kb}) - log(W_k)$$

This is essentially a visual heuristic

K-Means Clustering: Concluding Remarks

1. Direct implementation of K-Means can be slow
2. Online version:

$$\mu_k^{\text{new}} = \mu_k^{\text{old}} + \eta_n(\mathbf{x}_n - \mu_k^{\text{old}}) \tag{9.5}$$

K-Means Clustering: Concluding Remarks

1. Direct implementation of K-Means can be slow
2. Online version:

$$\mu_k^{\text{new}} = \mu_k^{\text{old}} + \eta_n(\mathbf{x}_n - \mu_k^{\text{old}}) \tag{9.5}$$

3. K-mediods, general distortion measure

$$\tilde{J} = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\mathcal{V}(\mathbf{x}_n, \mu_k) \tag{9.6}$$

where $\mathcal{V}(\cdot,\cdot)$ is any kind of dissimilarity measure

K-Means Clustering: Concluding Remarks

1. Direct implementation of K-Means can be slow
2. Online version:

$$\mu_k^{\text{new}} = \mu_k^{\text{old}} + \eta_n(\mathbf{x}_n - \mu_k^{\text{old}}) \tag{9.5}$$

3. K-mediods, general distortion measure

$$\tilde{J} = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \mathcal{V}(\mathbf{x}_n, \mu_k) \tag{9.6}$$

   where $\mathcal{V}(\cdot, \cdot)$ is any kind of dissimilarity measure

4. Image segmentation and compression example:

$K = 2$

4.2 %

$K = 3$

8.3 %

$K = 10$

16.7 %

Original image

100 %