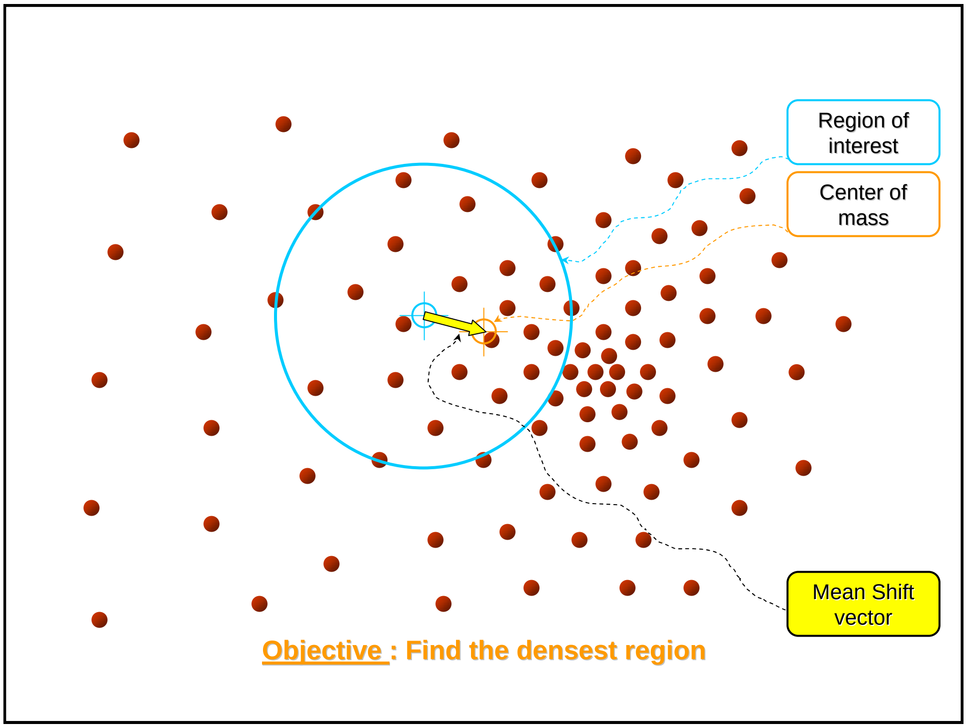
CAP6411
Computer Vision Systems
Lecture 13

Alper Yilmaz
Office: CSB 250
Email: yilmaz@cs.ucf.edu
Web: http://www.cs.ucf.edu/courses/**cap6411/cap6411**/spring2006

Slide credits go to Yaron Ukrainitz & Bernard Sarel

Mean-Shift Theory and Applications

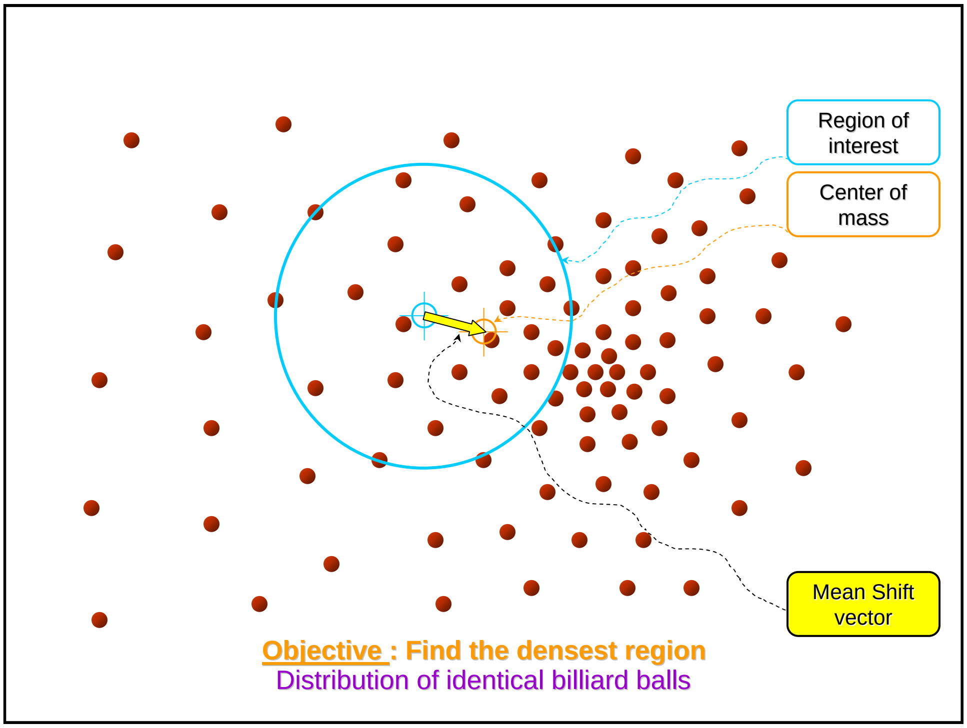


Region of interest

Center of mass

Mean Shift vector

Objective : Find the densest region



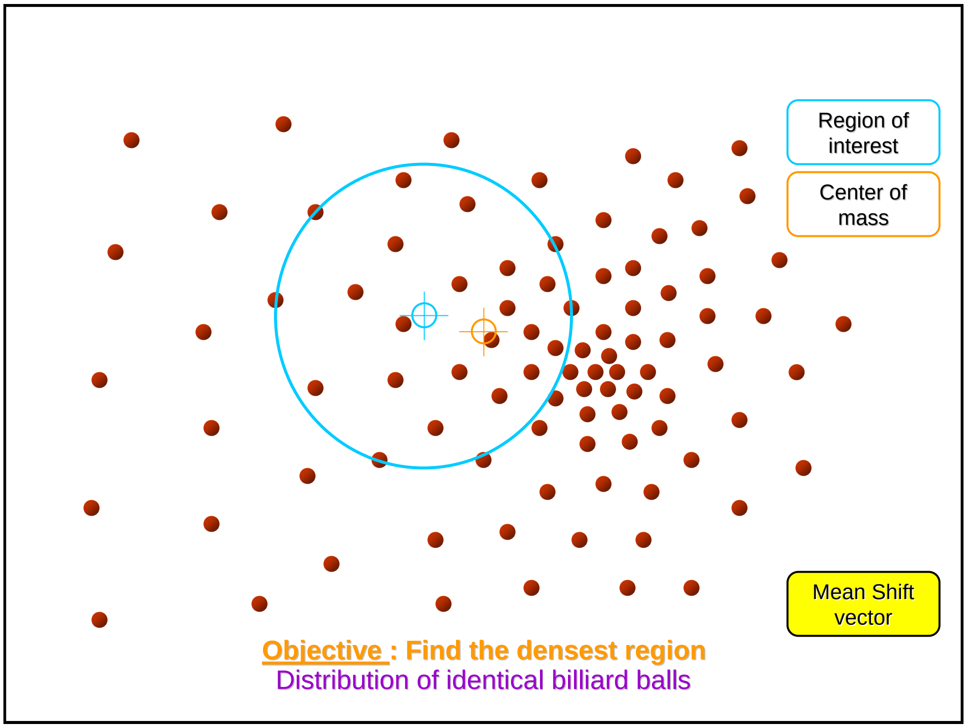
Region of interest

Center of mass

Mean Shift vector

Objective : Find the densest region

Distribution of identical billiard balls



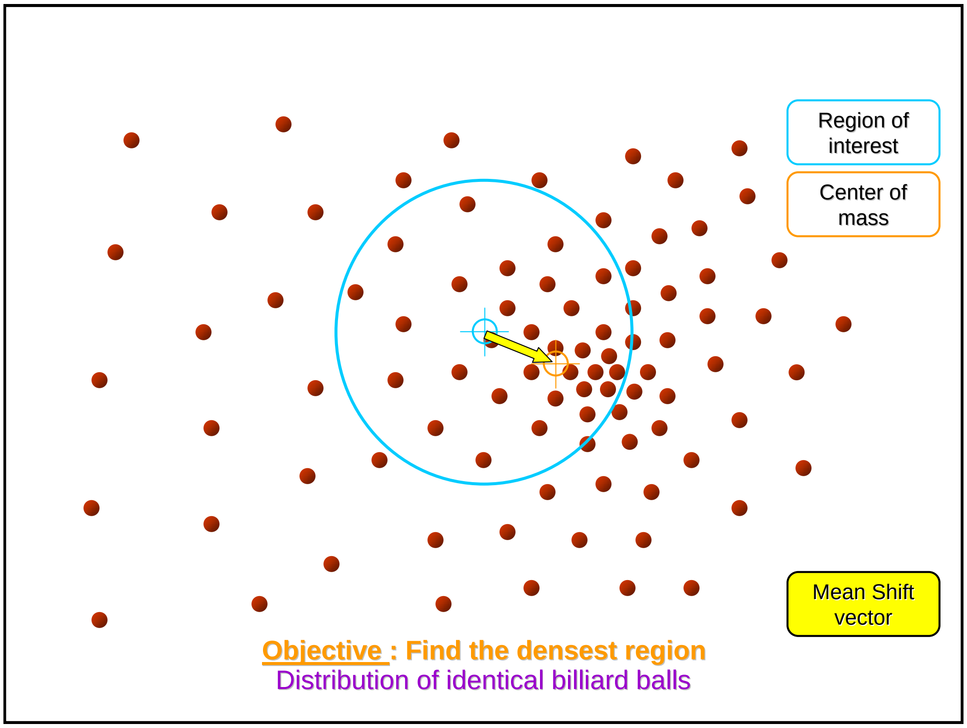
Region of interest

Center of mass

Mean Shift vector

Objective : **Find the densest region**

Distribution of identical billiard balls



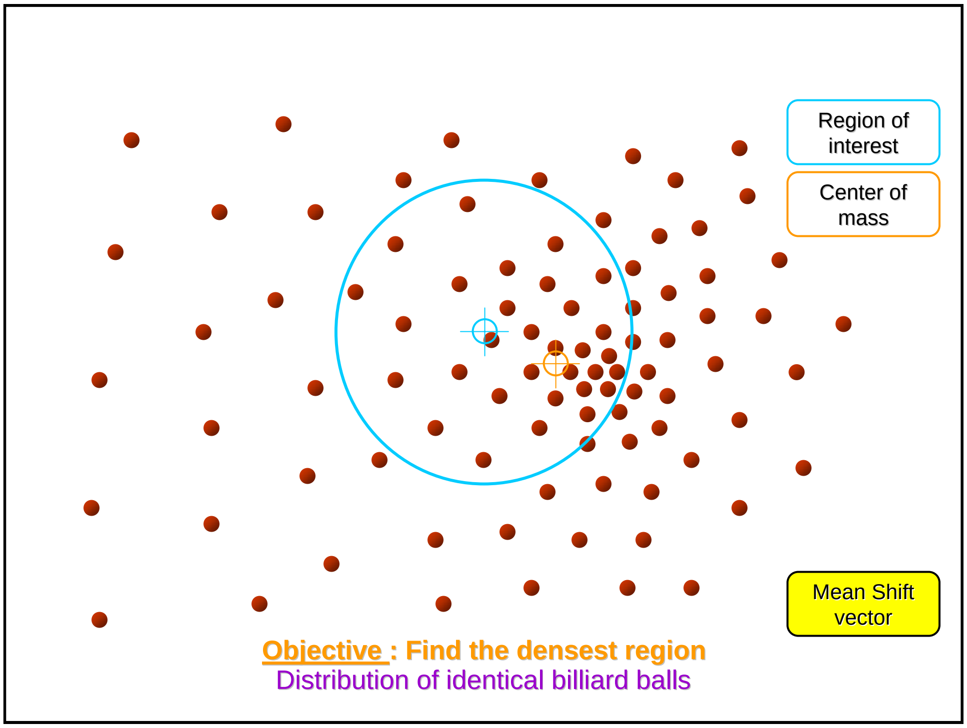
Region of interest

Center of mass

Mean Shift vector

Objective : **Find the densest region**

Distribution of identical billiard balls



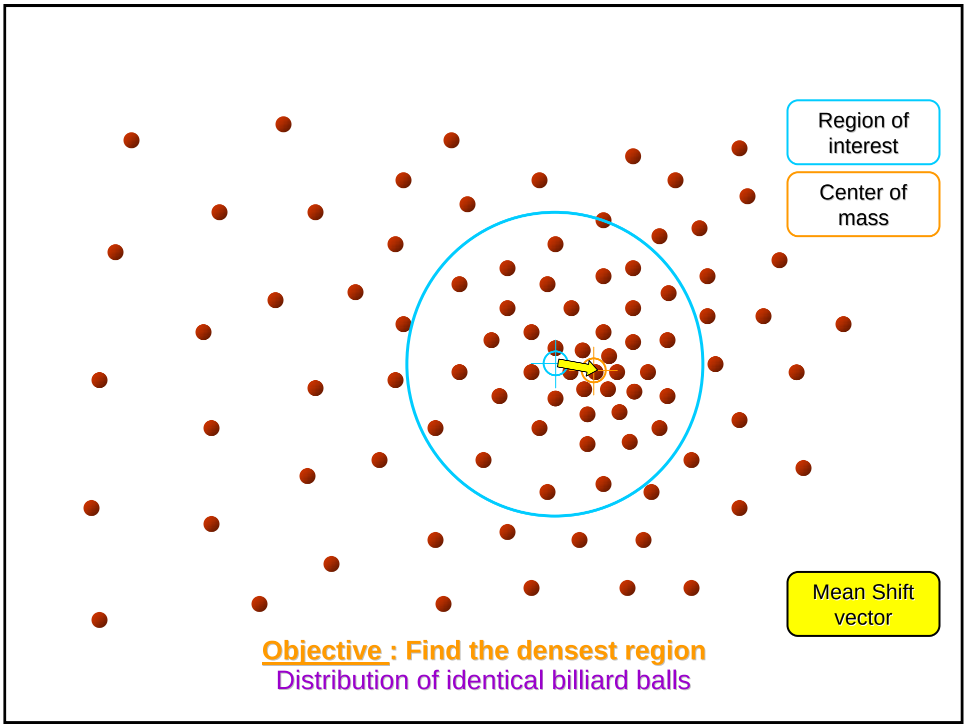
Region of interest

Center of mass

Mean Shift vector

Objective : Find the densest region

Distribution of identical billiard balls



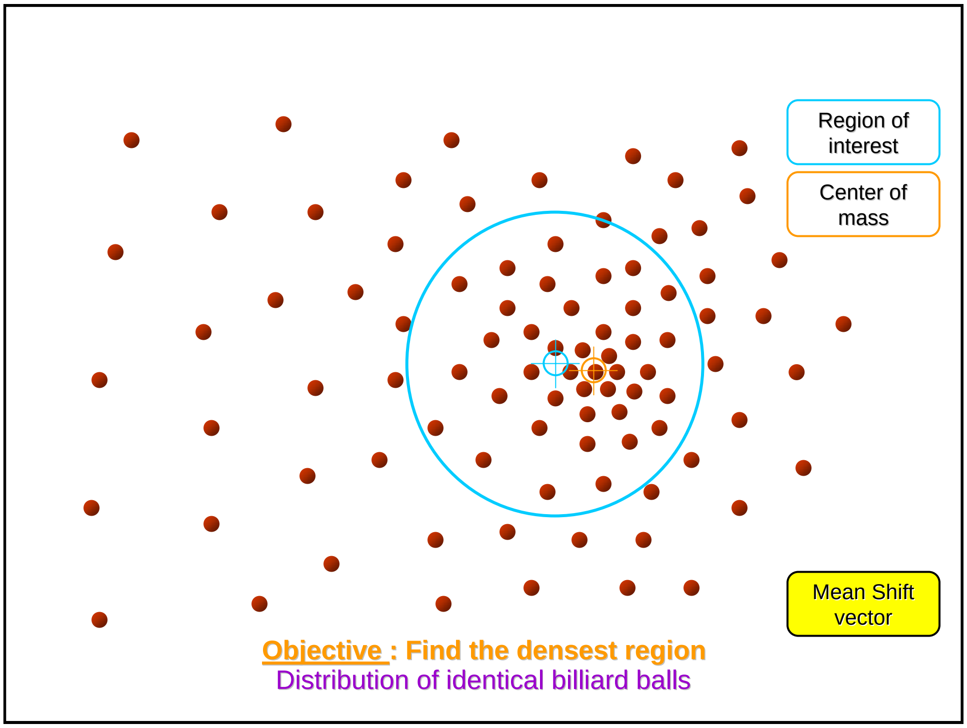
Region of interest

Center of mass

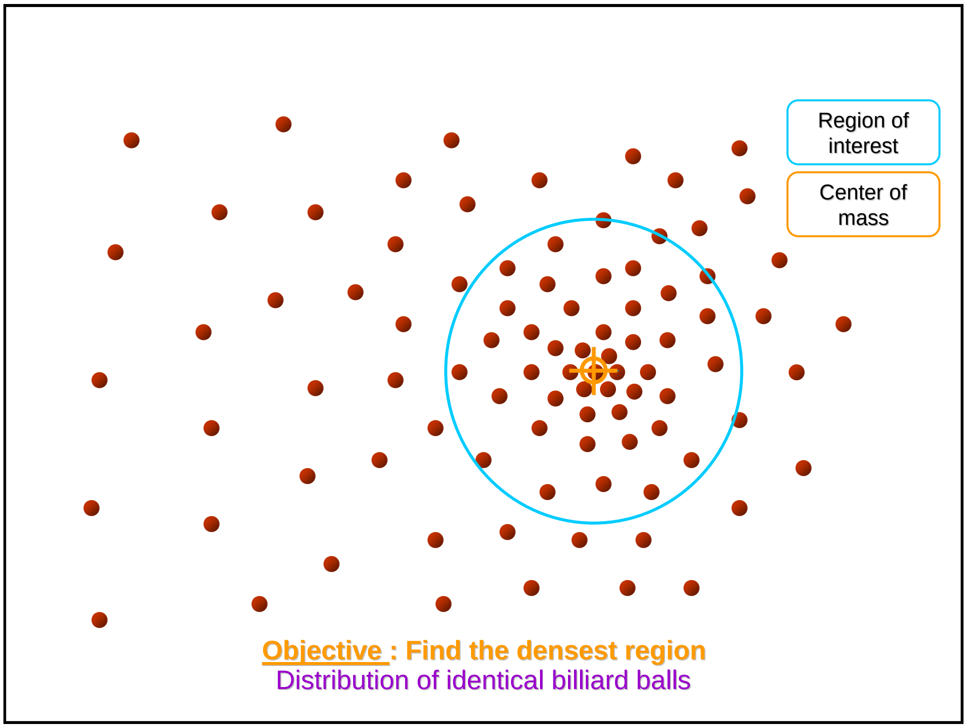
Mean Shift vector

Objective : Find the densest region

Distribution of identical billiard balls



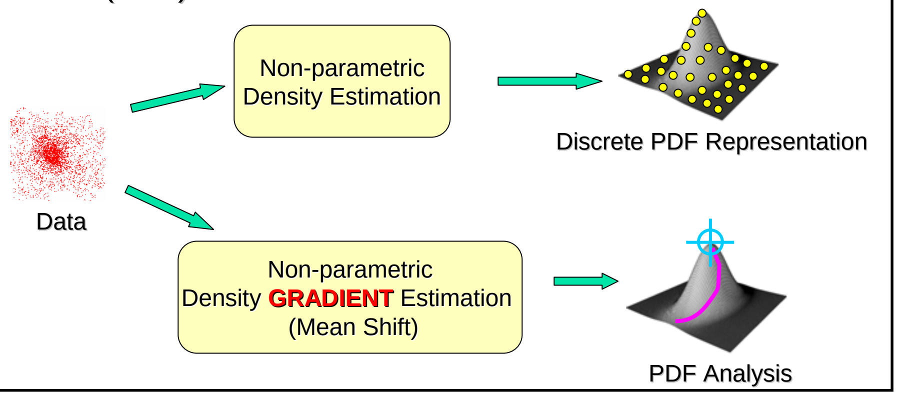
Objective : Find the densest region
Distribution of identical billiard balls



Objective : Find the densest region
Distribution of identical billiard balls

What is Mean-Shift?

- A tool for finding modes in a set of data samples, manifesting an underlying probability density function (PDF) in RN



Non-parametric Density Estimation

Discrete PDF Representation

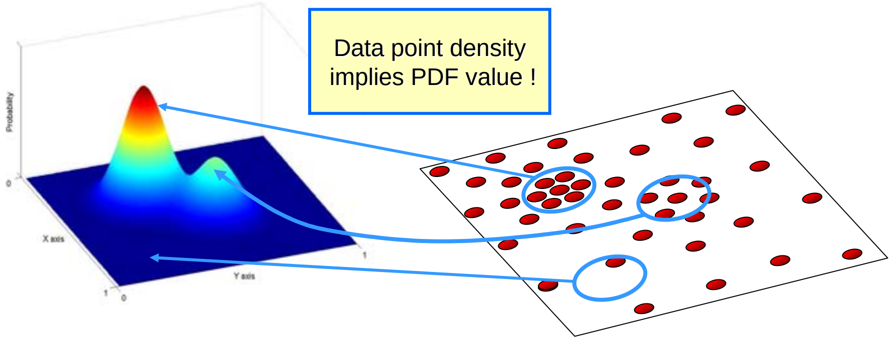
Data

Non-parametric Density **GRADIENT** Estimation (Mean Shift)

PDF Analysis

Non-Parametric Density Estimation

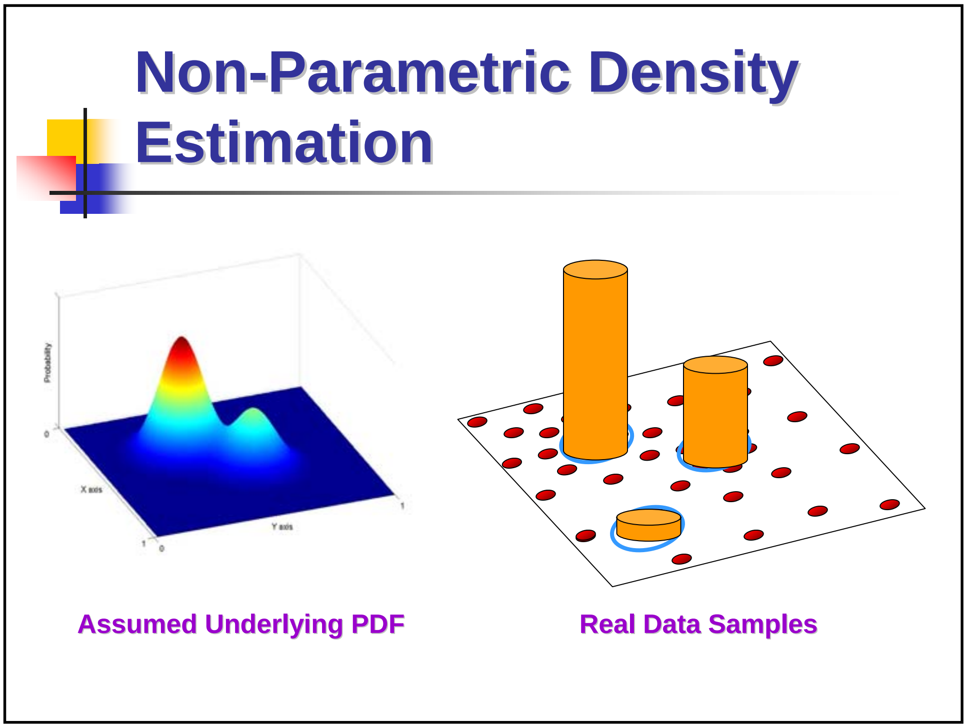
- Assumption : The data points are sampled from an underlying PDF



Data point density implies PDF value !

Assumed Underlying PDF

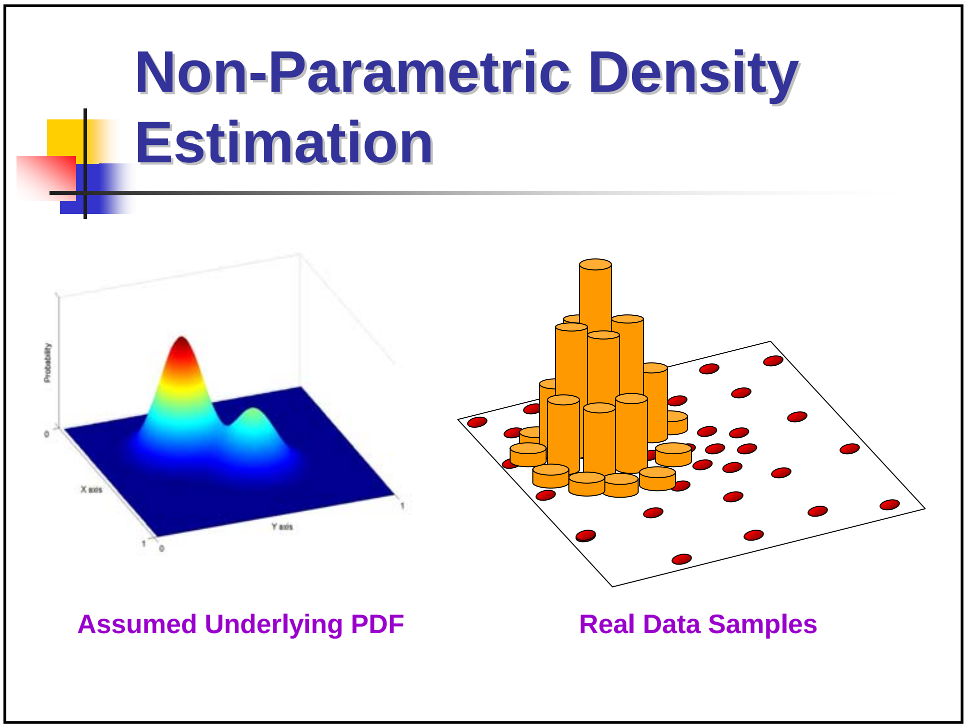
Real Data Samples



Non-Parametric Density Estimation

Assumed Underlying PDF

Real Data Samples



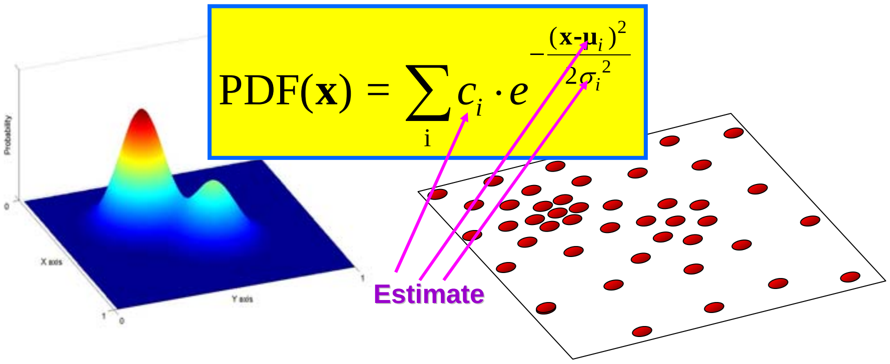
Non-Parametric Density Estimation

Assumed Underlying PDF

Real Data Samples

Parametric Density Estimation

- Assumption : The data points are sampled from an underlying PDF



$$\mathrm{PDF}(\mathbf{x}) = \sum_{\mathrm{i}} c_i \cdot e^{-\frac{(\mathbf{x}-\mathbf{u}_i)^2}{2\sigma_i^2}}$$

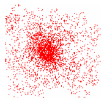
Assumed Underlying PDF **Real Data Samples**

Kernel Density Estimation

Parzen Windows - General Framework

$$P(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K(\mathbf{x}-\mathbf{x}_i)$$

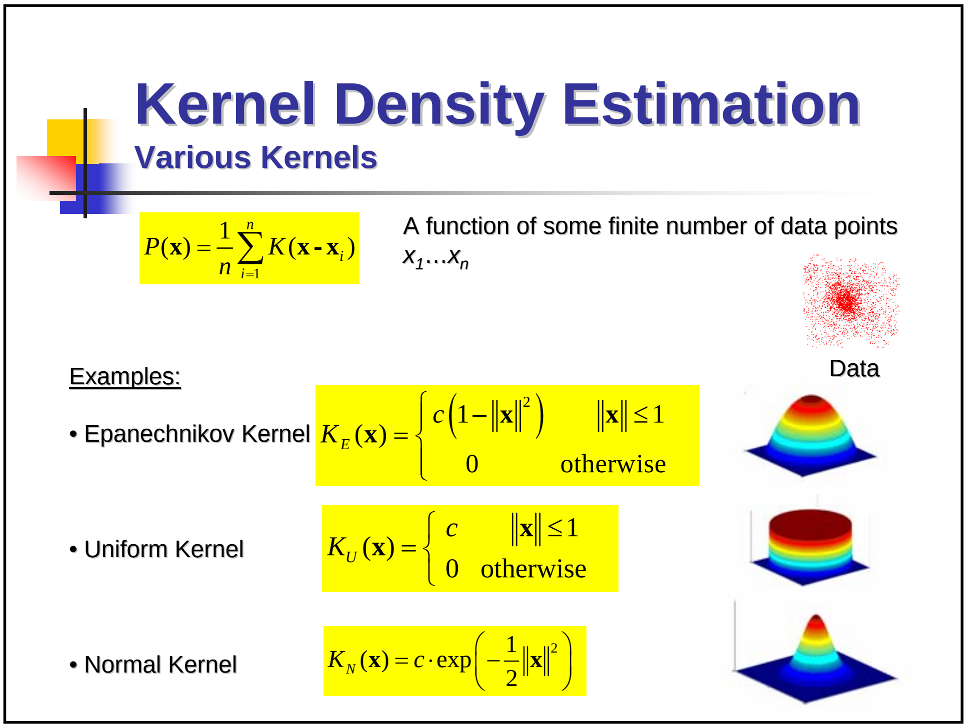
A function of some finite number of data points $x_1 \ldots x_n$



Data

Kernel Properties:

- Normalized $$\int_{R^d} K(\mathbf{x})d\mathbf{x} = 1$$
- Symmetric $$\int_{R^d} \mathbf{x}K(\mathbf{x})d\mathbf{x} = 0$$
- Exponential weight decay $$\lim_{\|\mathbf{x}\|\to\infty} \|\mathbf{x}\|^d K(\mathbf{x}) = 0$$
- ??? $$\int_{R^d} \mathbf{x}\mathbf{x}^T K(\mathbf{x})d\mathbf{x} = c\mathbf{I}$$



Kernel Density Estimation

Various Kernels

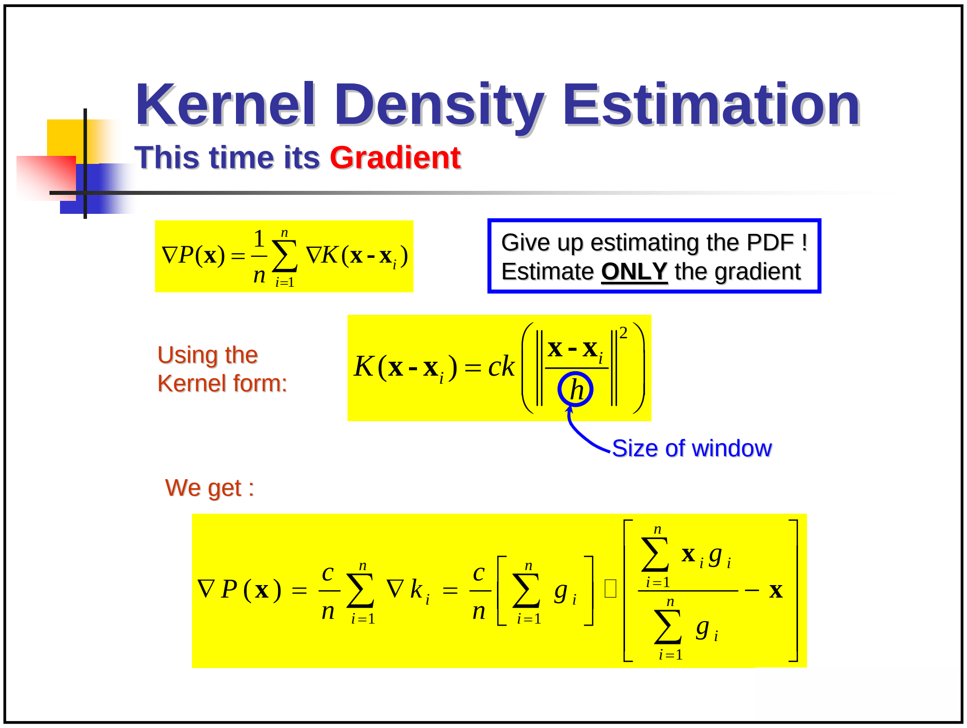
$P(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K(\mathbf{x} - \mathbf{x}_i)$

A function of some finite number of data points $x_1 \ldots x_n$

Data

<u>Examples:</u>

- Epanechnikov Kernel $K_E(\mathbf{x}) = \begin{cases} c\left(1 - \|\mathbf{x}\|^2\right) & \|\mathbf{x}\| \le 1 \\ 0 & \text{otherwise} \end{cases}$
- Uniform Kernel $K_U(\mathbf{x}) = \begin{cases} c & \|\mathbf{x}\| \le 1 \\ 0 & \text{otherwise} \end{cases}$
- Normal Kernel $K_N(\mathbf{x}) = c \cdot \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right)$



Kernel Density Estimation

This time its **Gradient**

$\nabla P(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} \nabla K(\mathbf{x} - \mathbf{x}_i)$

Give up estimating the PDF !
Estimate **<u>ONLY</u>** the gradient

Using the Kernel form: $K(\mathbf{x} - \mathbf{x}_i) = ck\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right)$

Size of window

We get :

$$\nabla P(\mathbf{x}) = \frac{c}{n}\sum_{i=1}^{n} \nabla k_i = \frac{c}{n}\left[\sum_{i=1}^{n} g_i\right] \square \left[\frac{\sum_{i=1}^{n} \mathbf{x}_i g_i}{\sum_{i=1}^{n} g_i} - \mathbf{x}\right]$$

Computing The Mean Shift

$$= \frac{c}{n}\left[\sum_{i=1}^{n} g_i\right] \square \left[\frac{\sum_{i=1}^{n} \mathbf{x}_i g_i}{\sum_{i=1}^{n} g_i} - \mathbf{x}\right]$$

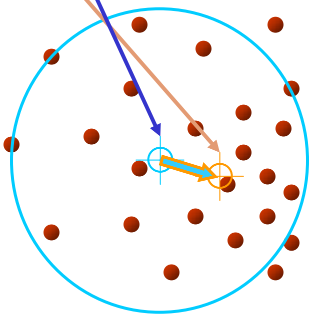
Yet another Kernel density estimation !

<u>Simple Mean Shift procedure</u>:

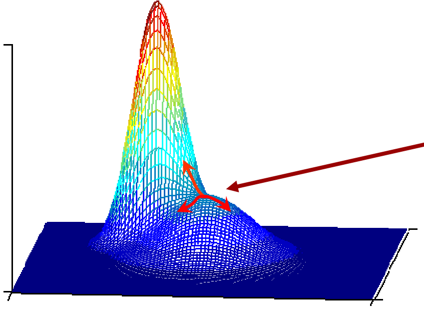
- Compute mean shift vector

$$\mathbf{m}(\mathbf{x}) = \left[\frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\frac{\|\mathbf{x}-\mathbf{x}_i\|^2}{h}\right)}{\sum_{i=1}^{n} g\left(\frac{\|\mathbf{x}-\mathbf{x}_i\|^2}{h}\right)} - \mathbf{x}\right]$$

•Translate the Kernel window by **m(x)**



Mean Shift Mode Detection



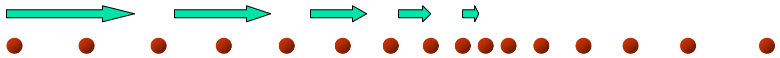
What happens if we reach a saddle point**?**

Perturb the mode position and check if we return back

<u>Updated Mean Shift Procedure:</u>

- Find all modes using the Simple Mean Shift Procedure
- Prune modes by perturbing them (find saddle points and plateaus)
- Prune nearby – take highest mode in the window

Mean Shift Properties

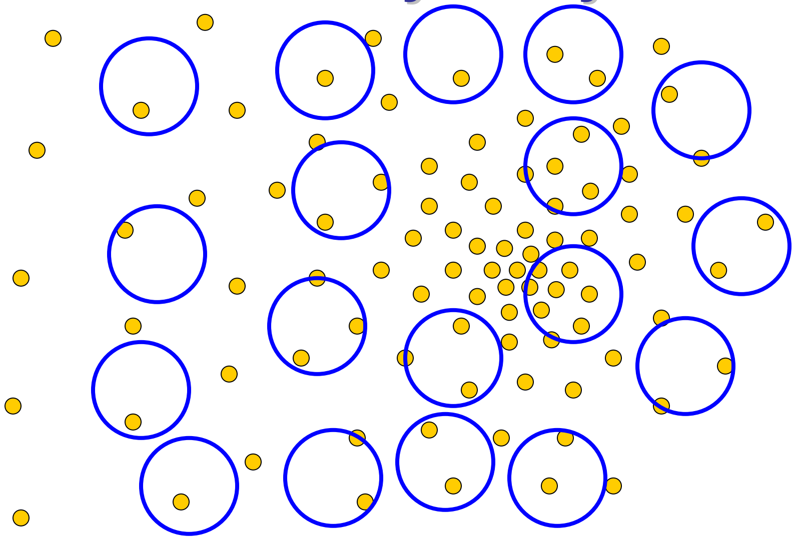


- Automatic convergence speed
  - mean shift vector size depends on gradient.
- Near maxima, the steps are small and refined

Adaptive Gradient Ascent

- Convergence is guaranteed for infinitesimal steps only infinitely convergent, (therefore set a lower bound)
- For Uniform Kernel ( ), convergence is achieved in a finite number of steps
- Normal Kernel (exhibits a smooth trajectory, but is slower than Uniform Kernel ( ).

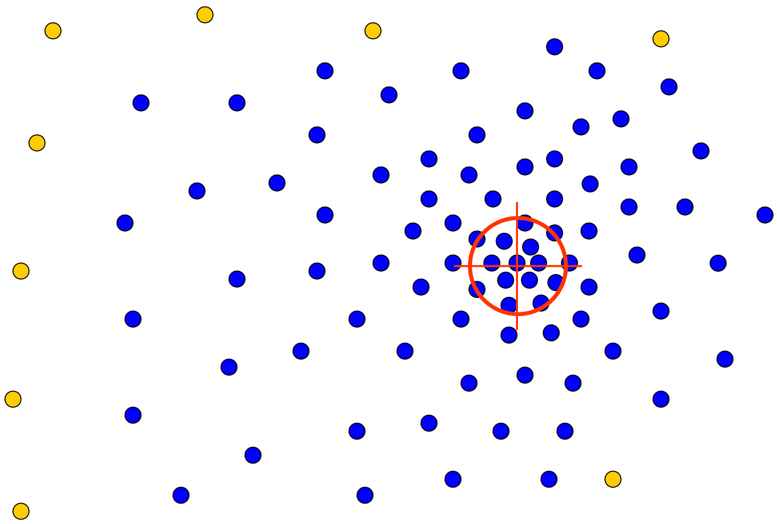
Real Modality Analysis



Tessellate the space with windows

Run the procedure in parallel

Real Modality Analysis

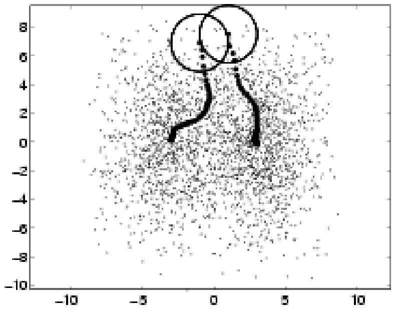


The **blue** data points were traversed by the windows towards the mode

---

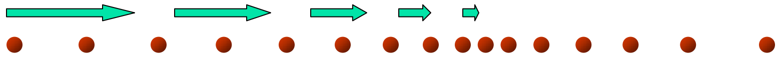
Real Modality Analysis

An example



Window tracks signify the steepest ascent directions

Mean Shift Strengths & Weaknesses



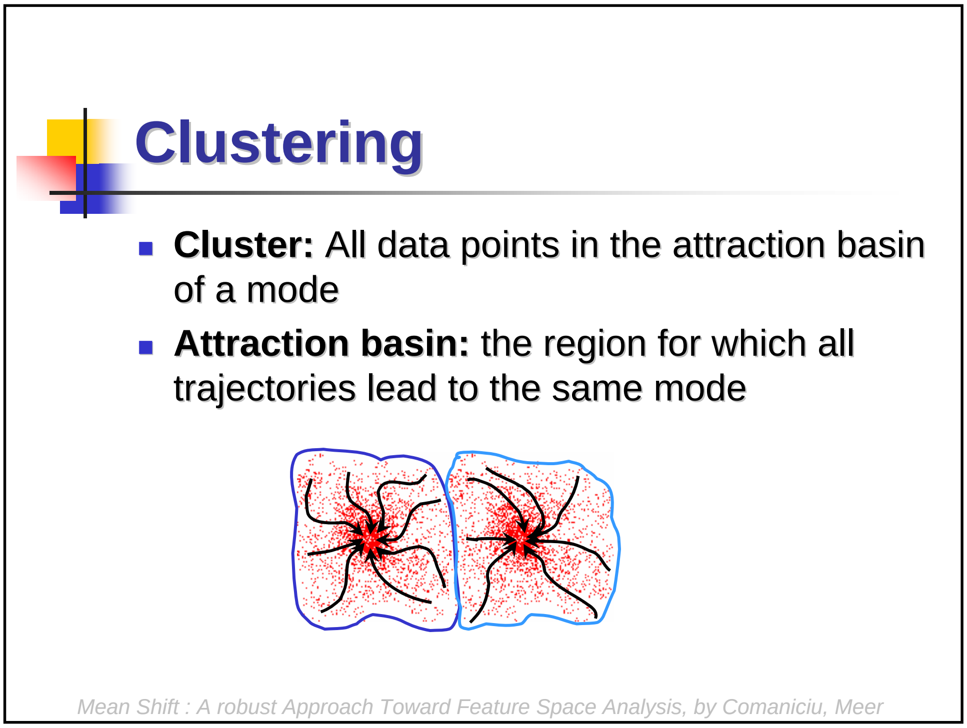
Strengths:

- Application independent tool
- Suitable for real data analysis
- Does not assume any prior shape (e.g. elliptical) on data clusters
- Can handle arbitrary feature spaces
- Only ONE parameter to choose
- h (window size) has a physical meaning, unlike K-Means

Weaknesses:

- The window size (bandwidth selection) is not trivial
- Inappropriate window size can cause modes to be merged, or generate additional “shallow” modes

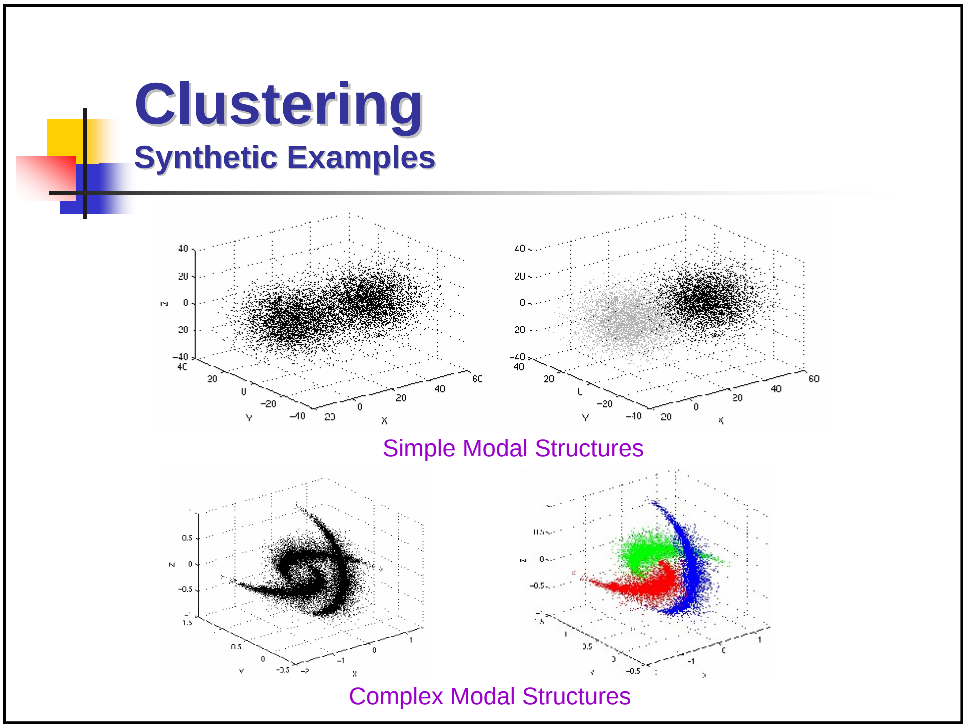
Mean Shift Applications



Clustering

- **Cluster:** All data points in the attraction basin of a mode
- **Attraction basin:** the region for which all trajectories lead to the same mode

Mean Shift : A robust Approach Toward Feature Space Analysis, by Comaniciu, Meer



Clustering

Synthetic Examples

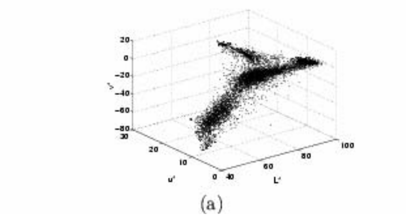
Simple Modal Structures

Complex Modal Structures

Feature space:
L*u*v representation

Clustering
Real Example

Initial window
enters

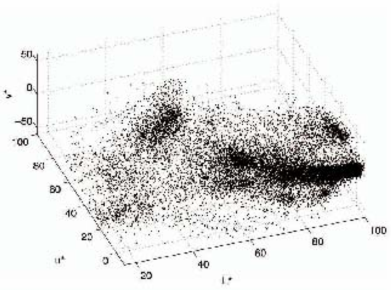


(a)

M

pruning

Clustering
Real Example

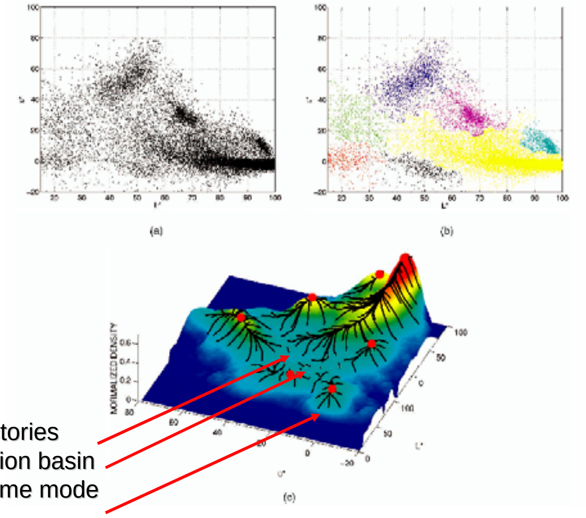


L*u*v space representation

Clustering

Real Example

2D (L*u)
space
representation



Final clusters

Not all trajectories
in the attraction basin
reach the same mode

Discontinuity Preserving Smoothing

Feature space : Joint domain = spatial coordinates + color space

$$K(\mathbf{x}) = C \cdot k_s\left(\left\|\frac{\mathbf{x}^s}{h_s}\right\|\right) \cdot k_r\left(\left\|\frac{\mathbf{x}^r}{h_r}\right\|\right)$$

Meaning : treat the image as data points in the spatial and gray level domain

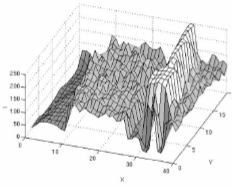
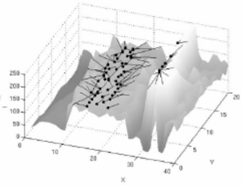
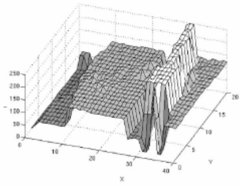


Image Data (slice)



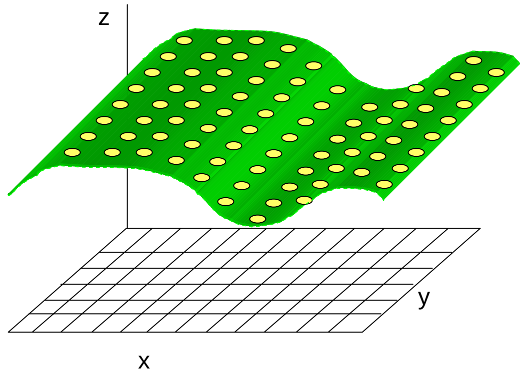
Mean Shift vectors



Smoothing result

Mean Shift : A robust Approach Toward Feature Space Analysis, by Comaniciu, Meer

Discontinuity Preserving Smoothing

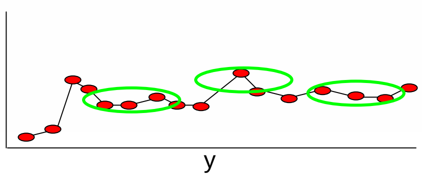


The image gray levels…

… can be viewed as data points in the *x, y, z* space (joined spatial And color space)

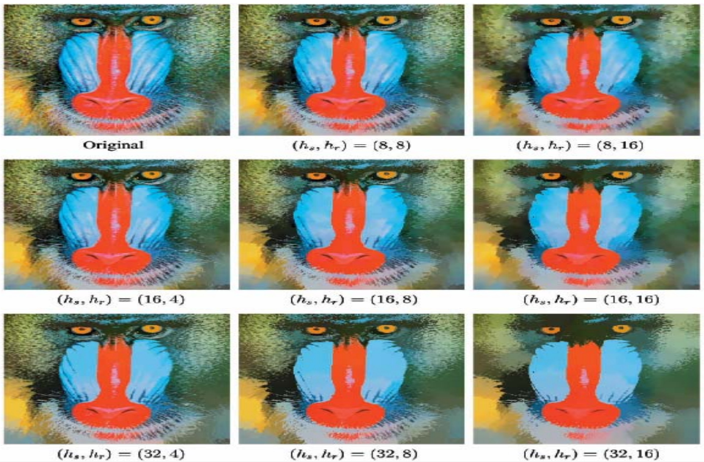
Discontinuity Preserving Smoothing

Flat regions induce the modes !



Discontinuity Preserving Smoothing

- The effect of window size in spatial and range spaces



Original

$(h_s, h_r) = (8, 8)$

$(h_s, h_r) = (8, 16)$

$(h_s, h_r) = (16, 4)$

$(h_s, h_r) = (16, 8)$

$(h_s, h_r) = (16, 16)$

$(h_s, h_r) = (32, 4)$

$(h_s, h_r) = (32, 8)$

$(h_s, h_r) = (32, 16)$

Discontinuity Preserving Smoothing

Example

Discontinuity Preserving Smoothing

Example

Mean-Shift Segmentation

Algorithm

- Run Filtering (discontinuity preserving smoothing)
- Cluster the clusters which are closer than window size

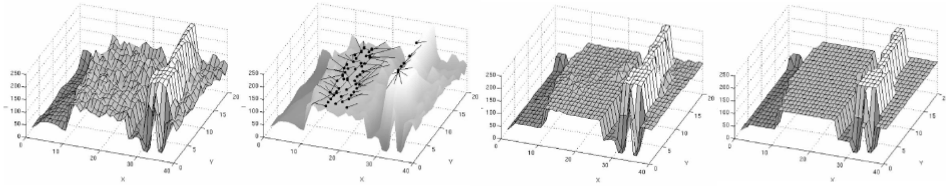


Image Data (slice) Mean-shift vectors Smoothing result Segmentation result

Mean Shift : A robust Approach Toward Feature Space Analysis, by Comaniciu, Meer
http://www.caip.rutgers.edu/~comanici

Segmentation

Example

…when feature space is only gray levels…

Segmentation

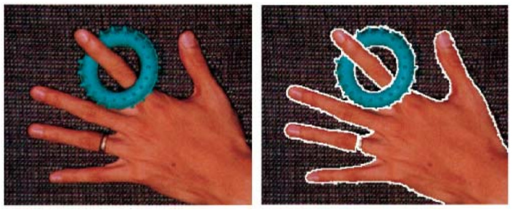
Example

Segmentation

Example

Segmentation

Example

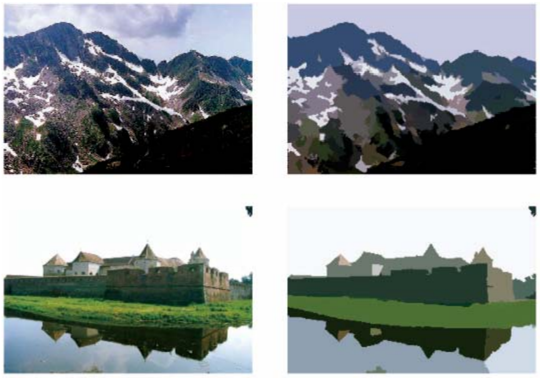


Segmentation

Example

Segmentation

Example



Segmentation

Example