Mean Shift

Theory and Applications

Reference

D. Comaniciu and P. Meer, “Mean shift: A robust approach toward feature space analysis,” IEEE T. PAMI, vol. 24, no. 5, pp. 603-619, May 2002.

Lecturer: Sang Hwa Lee

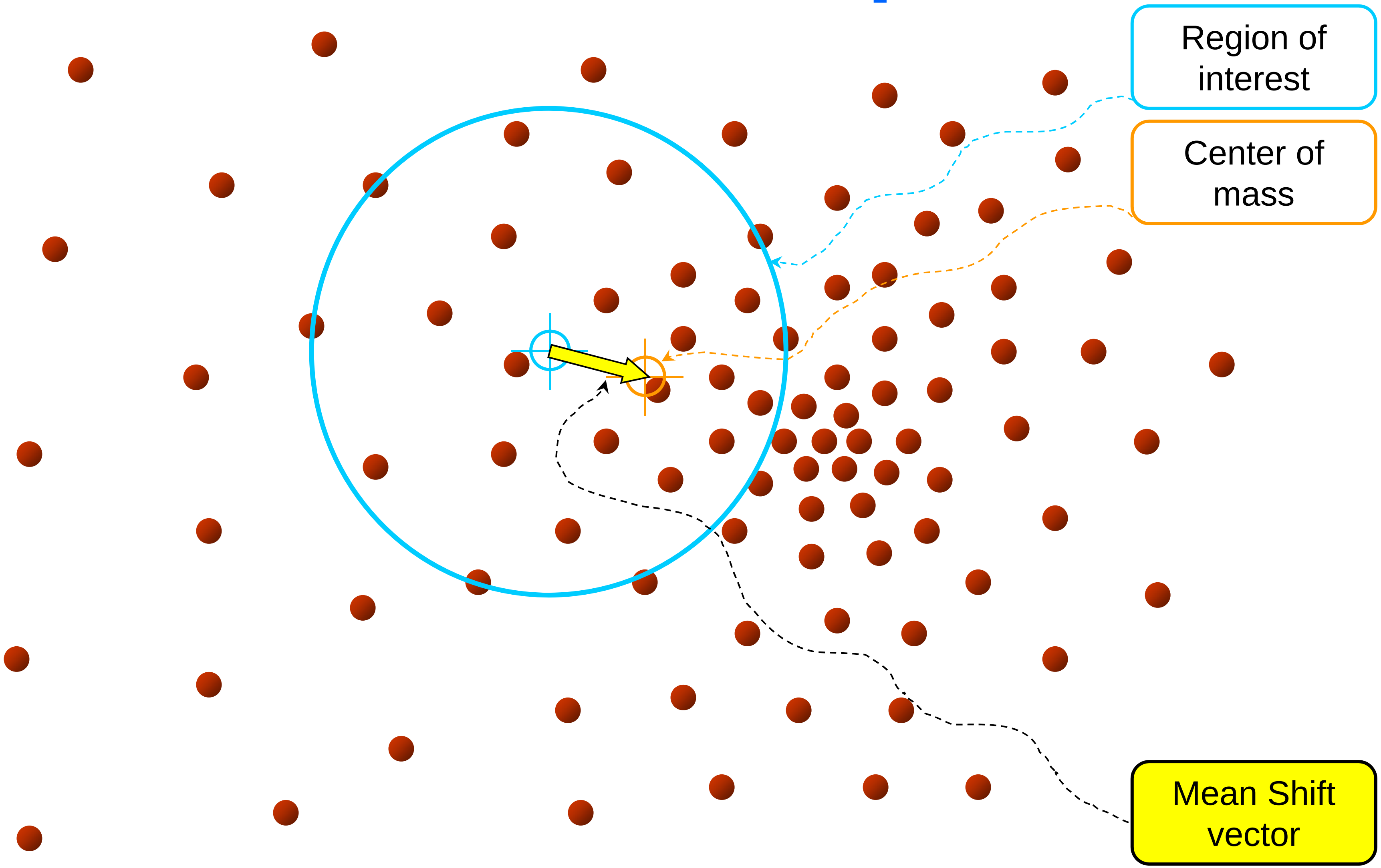
Agenda

- **Mean Shift Theory**
  - What is Mean Shift ?
  - Density Estimation Methods
  - Deriving the Mean Shift
  - Mean shift properties

- **Applications**
  - Clustering
  - Discontinuity Preserving Smoothing
  - Object Contour Detection
  - Segmentation
  - Object Tracking

Mean Shift Theory

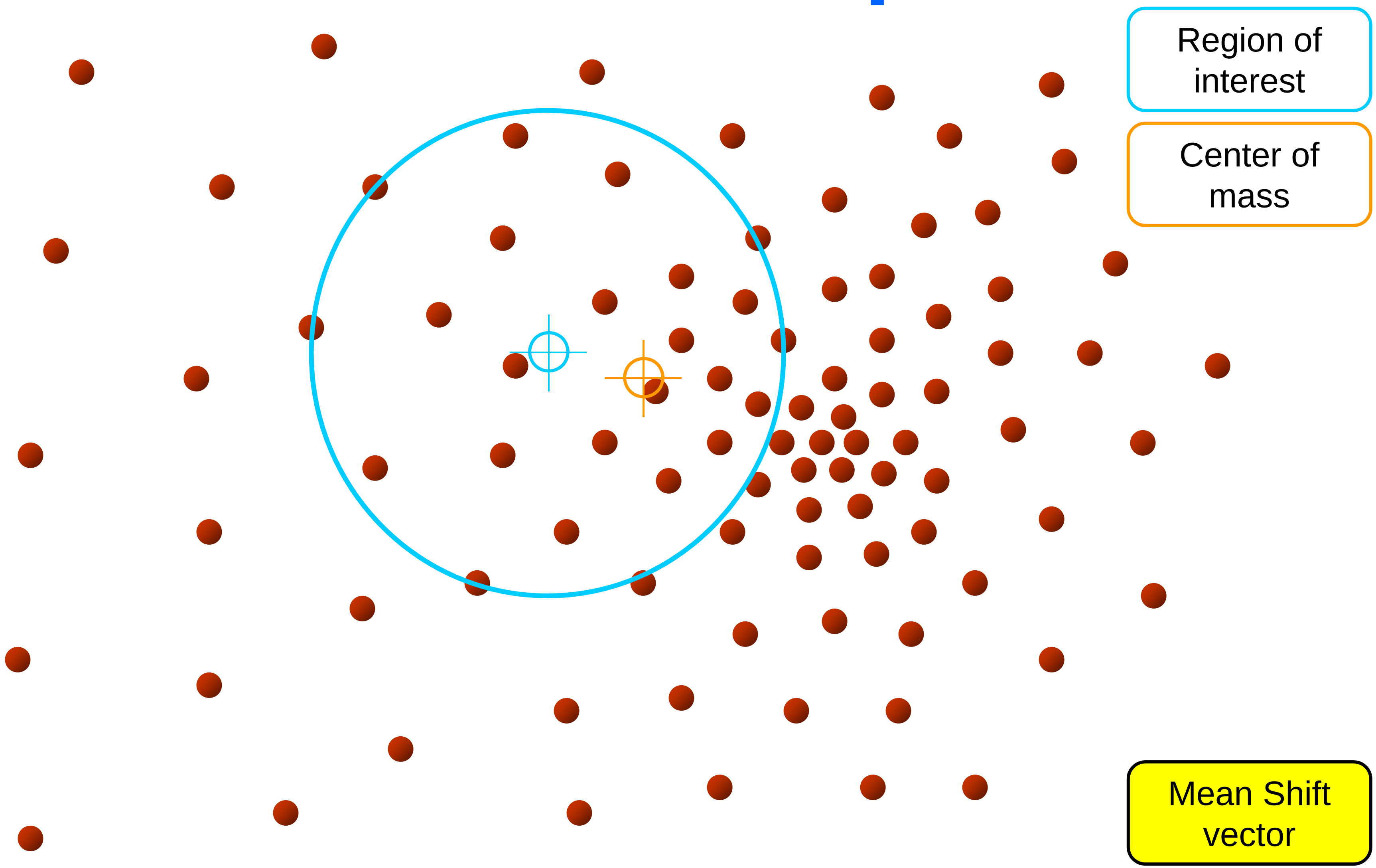
Intuitive Description



Objective : Find the densest region

Distribution of identical billiard balls

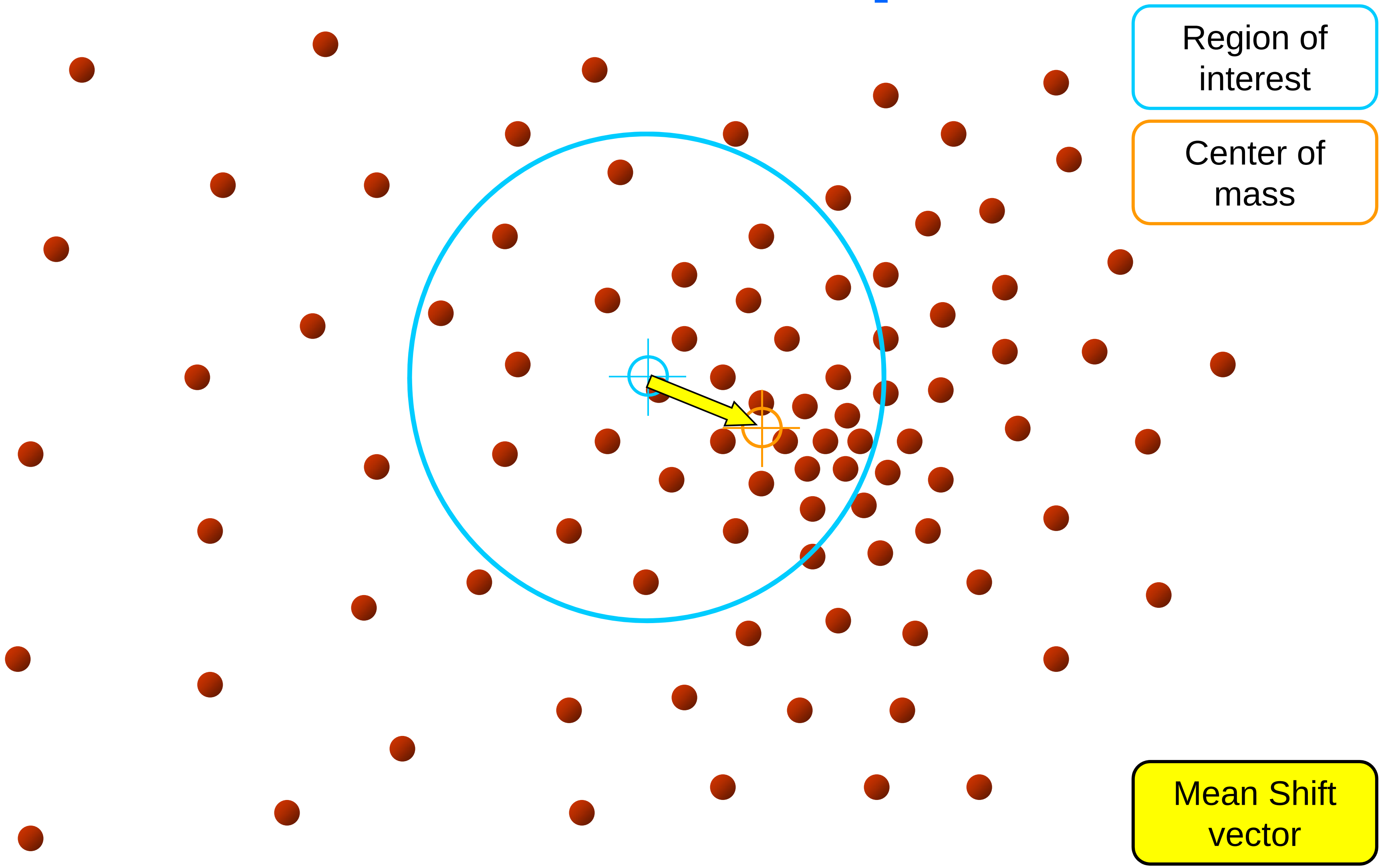
Intuitive Description



Objective : Find the densest region

Distribution of identical billiard balls

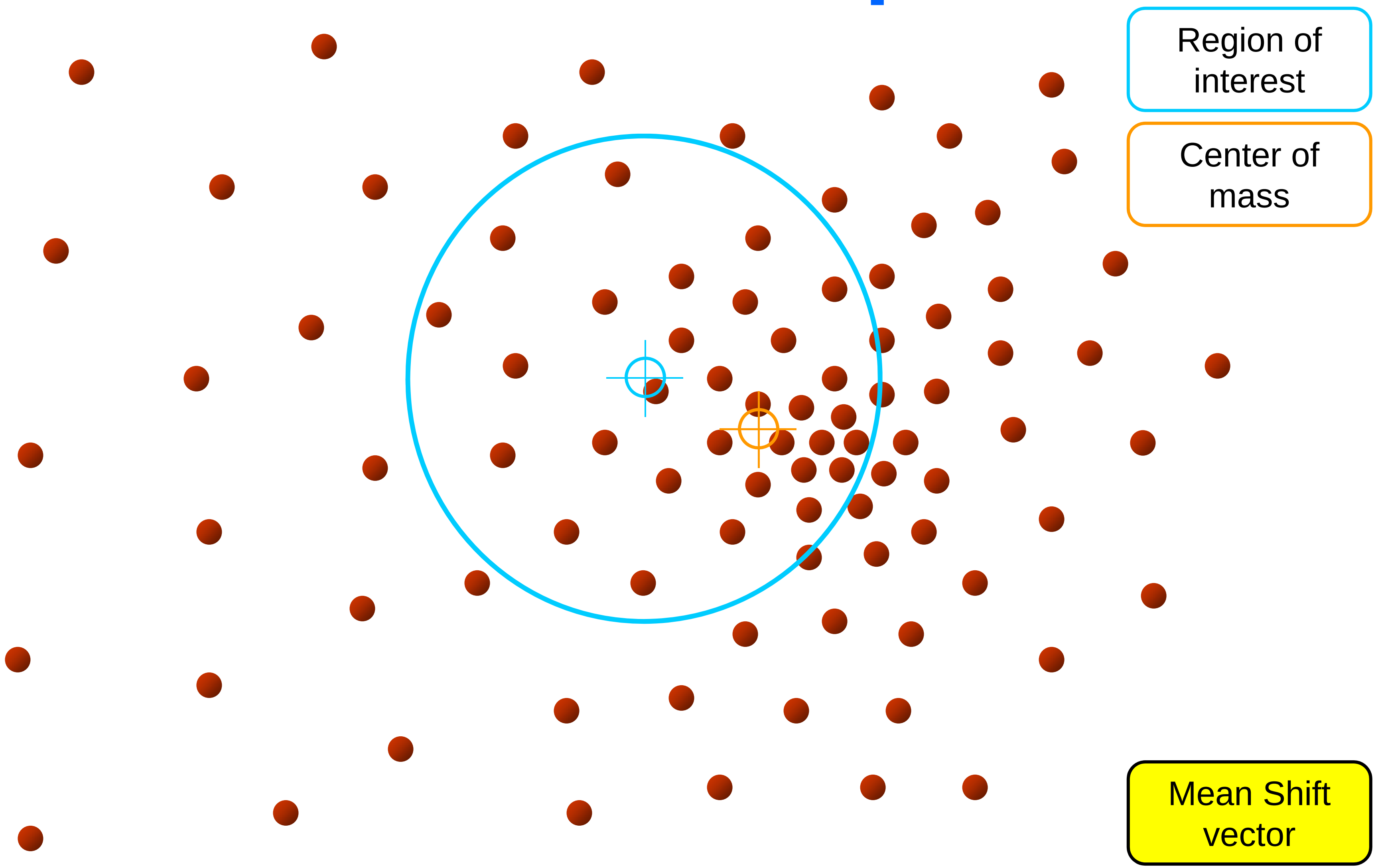
Intuitive Description



Objective : Find the densest region

Distribution of identical billiard balls

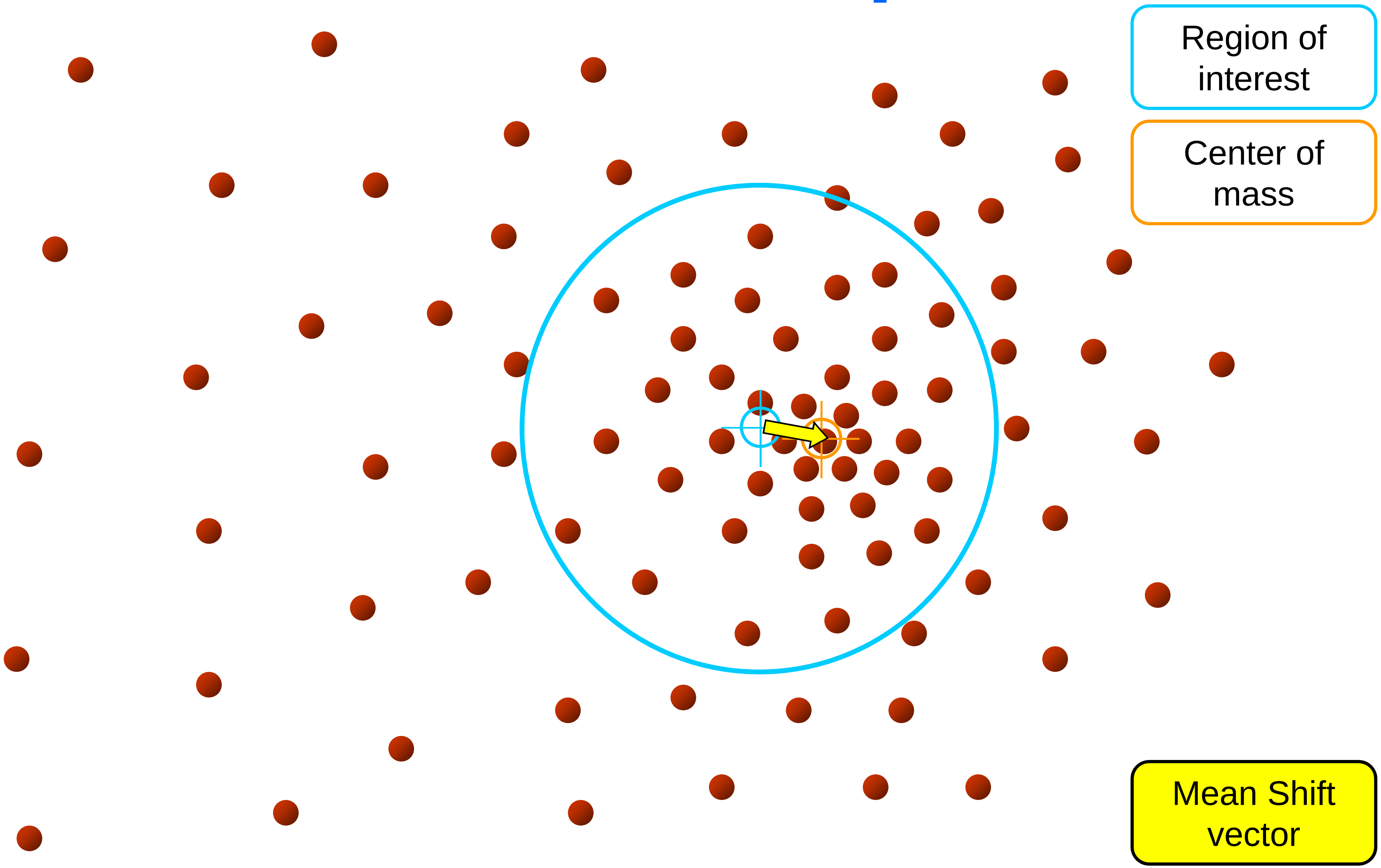
Intuitive Description



Objective : Find the densest region

Distribution of identical billiard balls

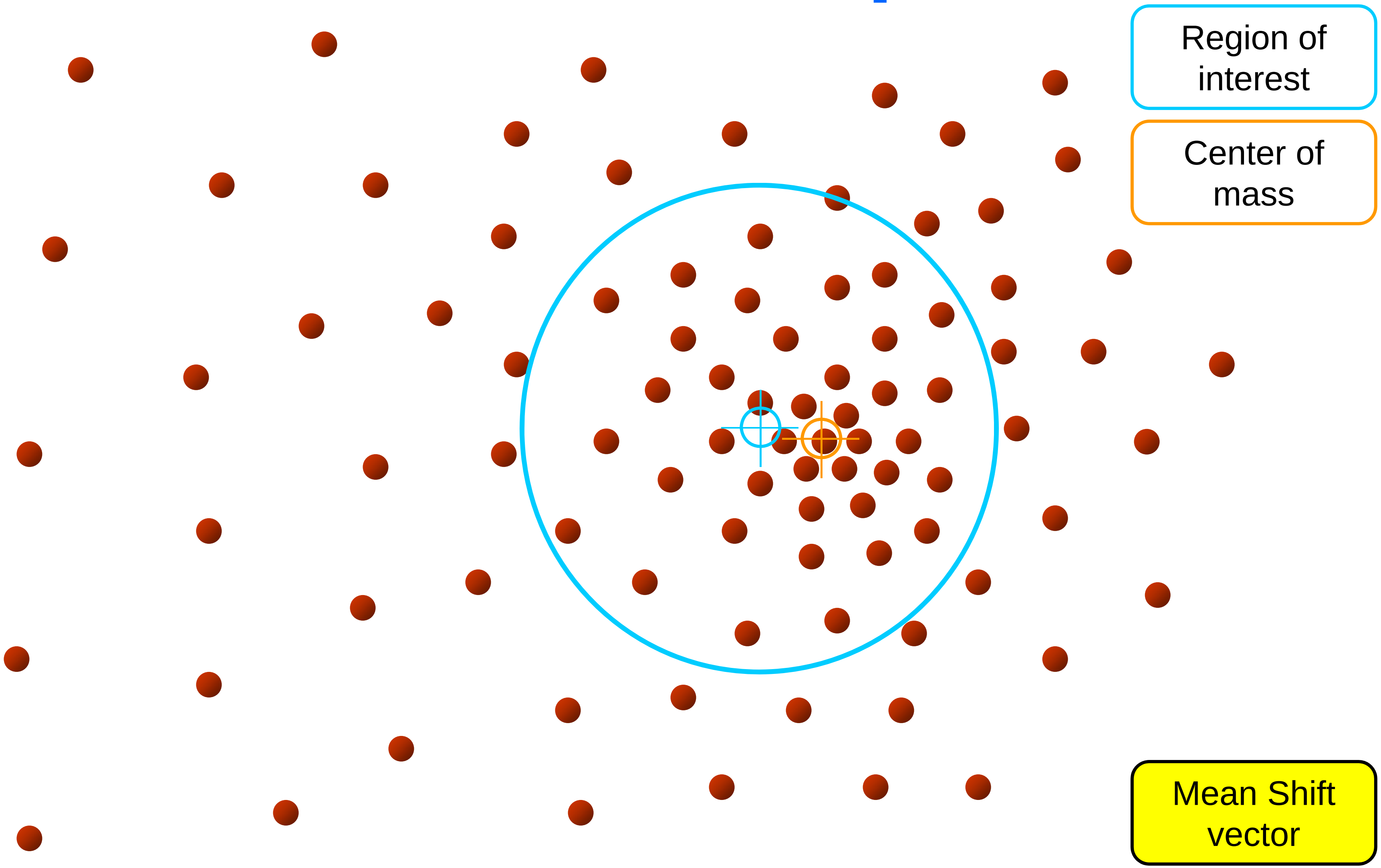
Intuitive Description



Objective : Find the densest region

Distribution of identical billiard balls

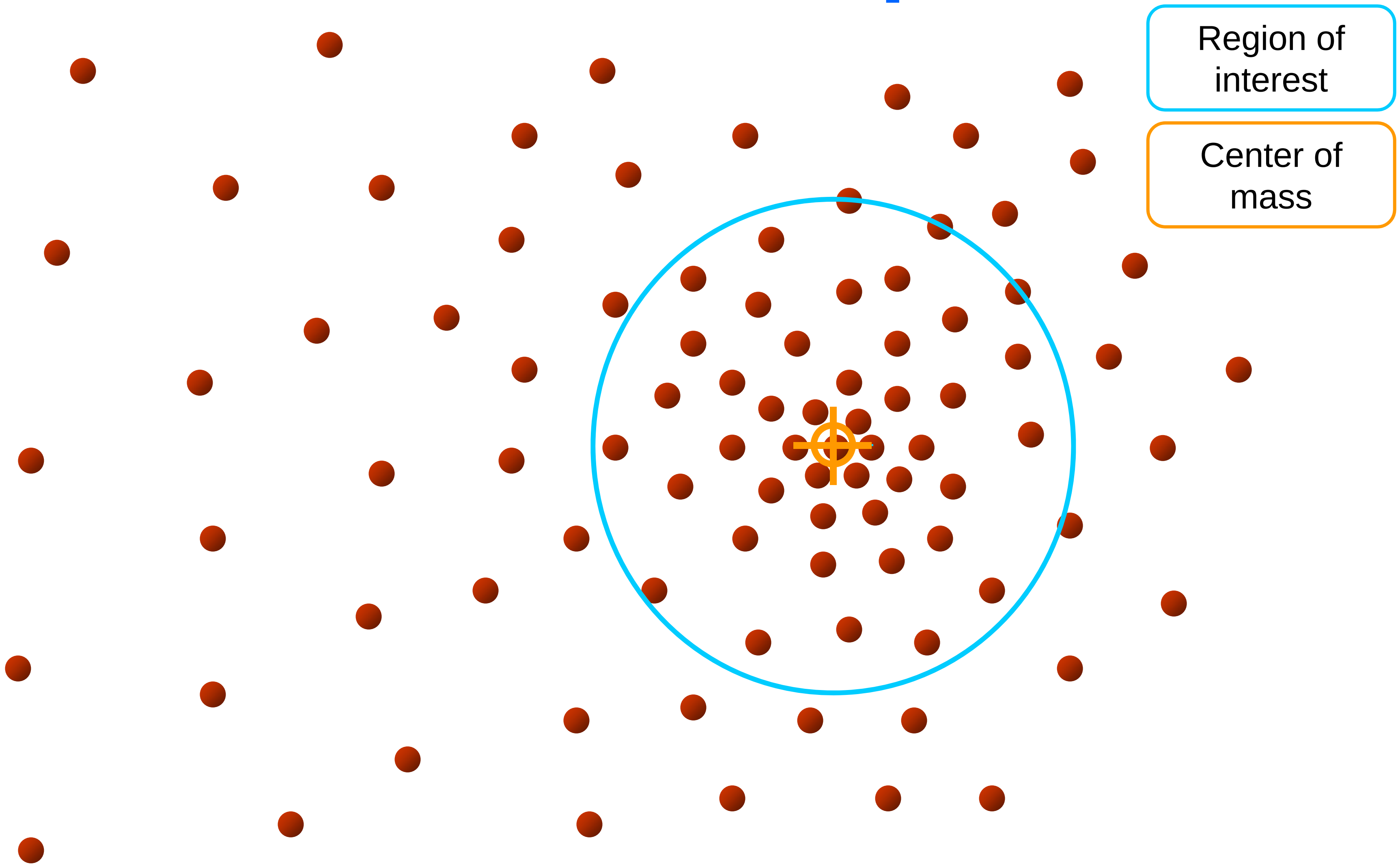
Intuitive Description



Objective : Find the densest region

Distribution of identical billiard balls

Intuitive Description



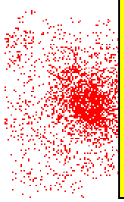
Objective : Find the densest region

Distribution of identical billiard balls

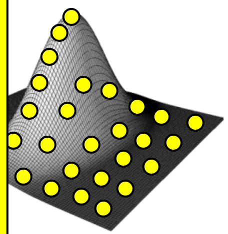
What is Mean Shift ?

A tool for:
Finding modes in a set of data samples, manifesting an underlying probability density function (PDF) in R^N

PDF in feature space
- **Color space**
- **Scale space**
- **Actually any feature space you can conceive**
- **…**

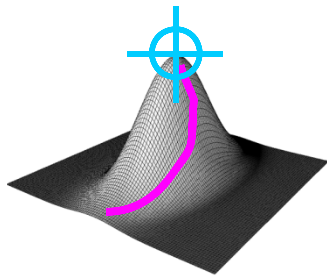


Data



DF Representation

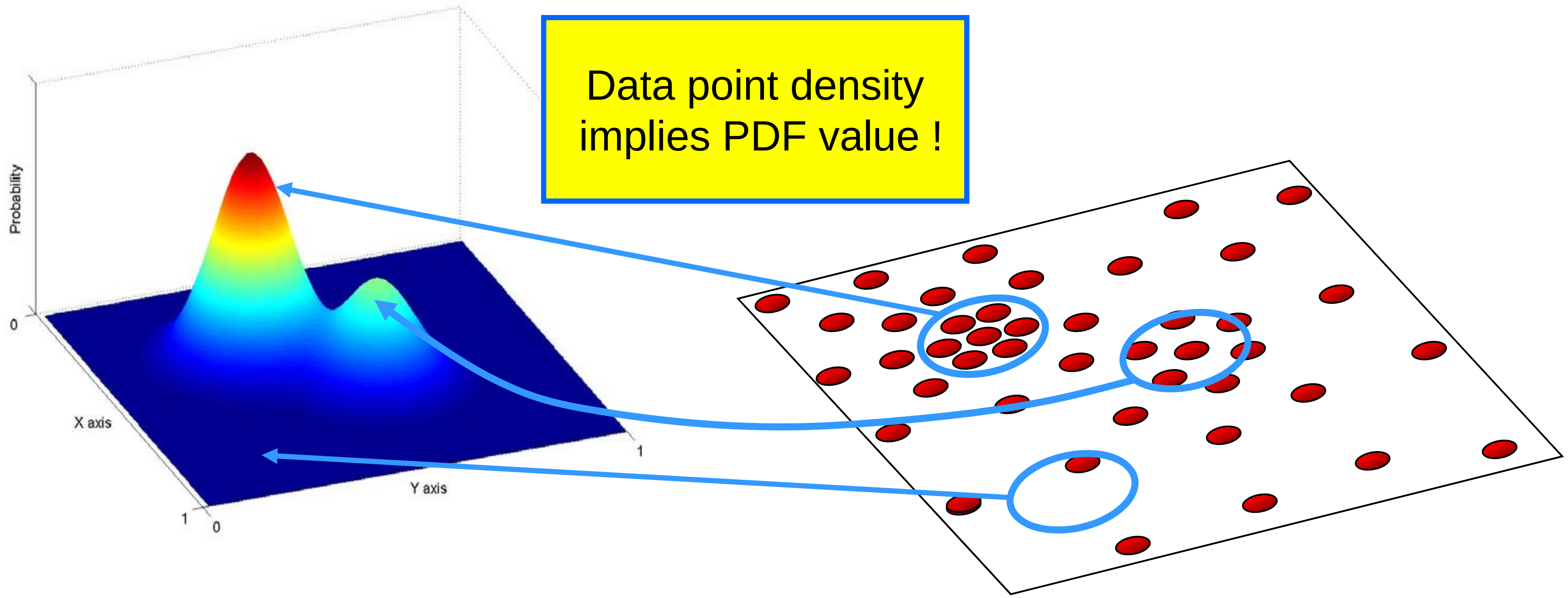
Non-parametric
Density **GRADIENT** Estimation
(Mean Shift)



PDF Analysis

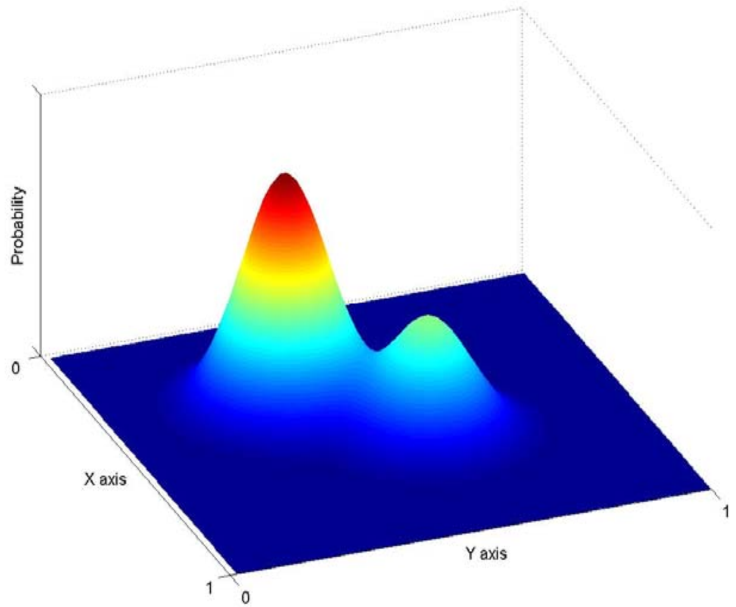
Non-Parametric Density Estimation

<u>Assumption</u> : The data points are sampled from an underlying PDF

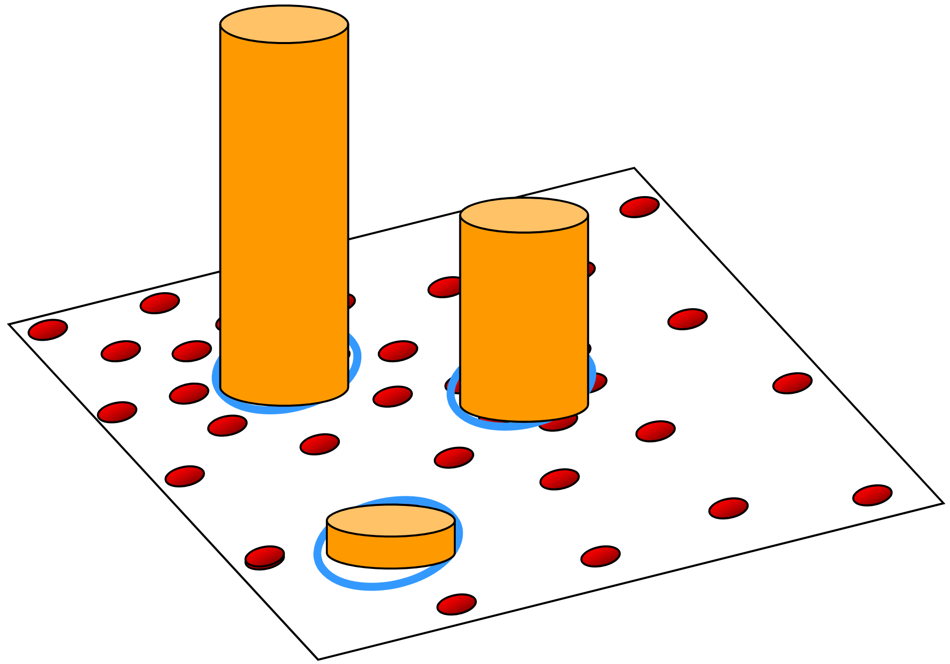


Assumed Underlying PDF **Real Data Samples**

Non-Parametric Density Estimation

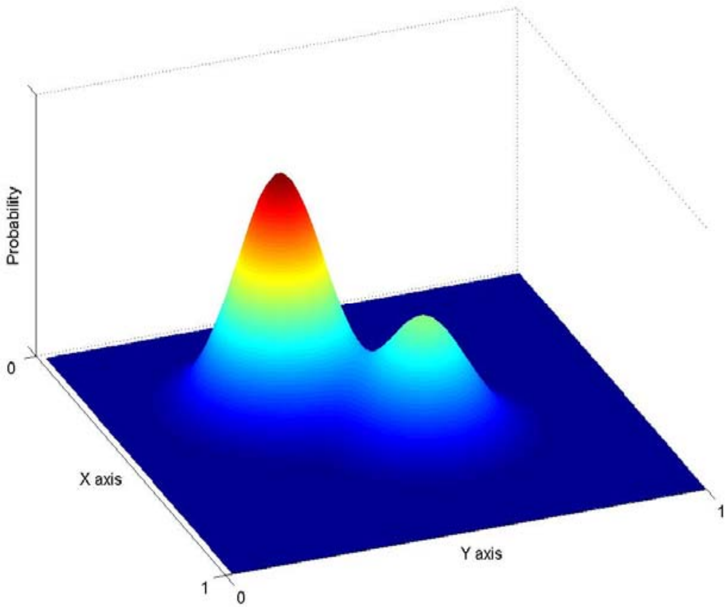


Assumed Underlying PDF

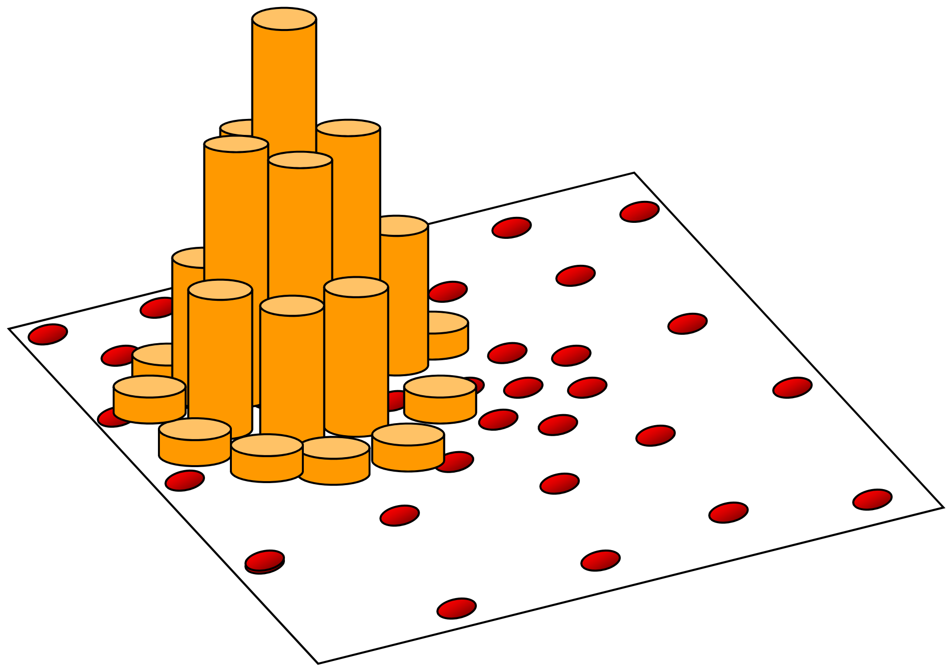


Real Data Samples

Non-Parametric ? Density Estimation



Assumed Underlying PDF



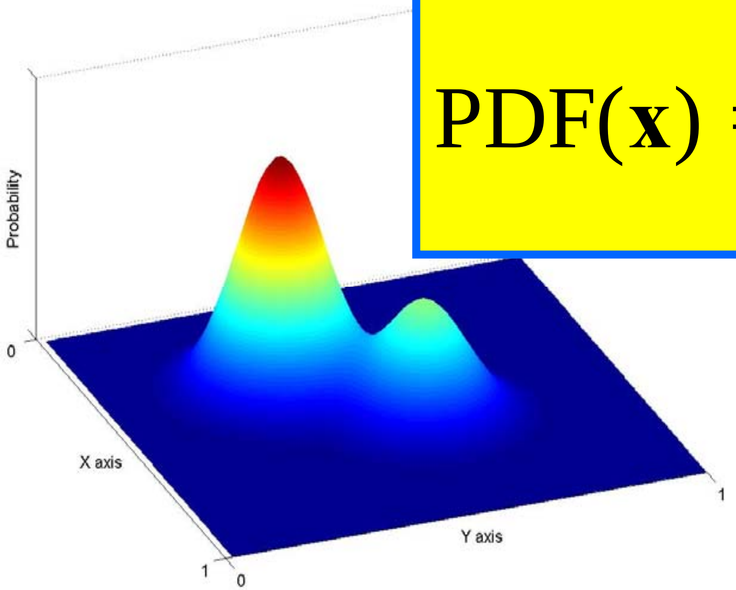
Real Data Samples

Parametric Density Estimation

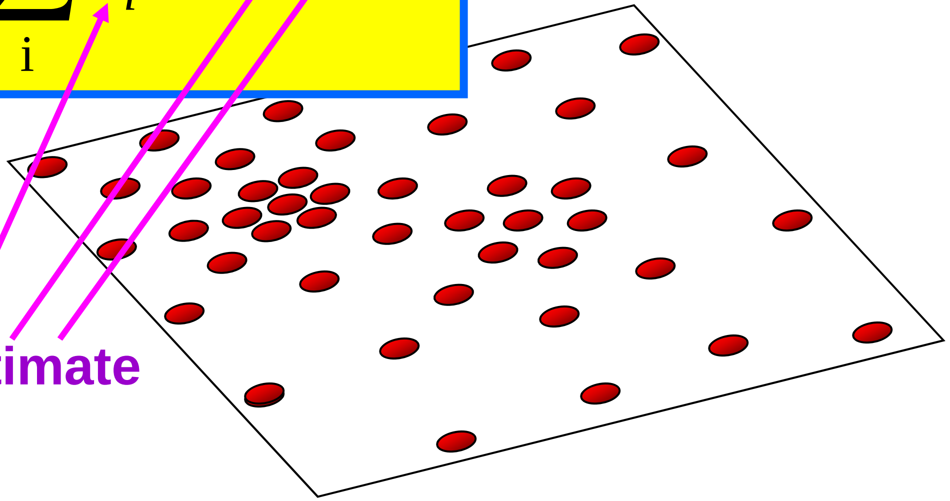
<u>Assumption</u> : The data points are sampled from an underlying PDF

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

Estimate



Assumed Underlying PDF



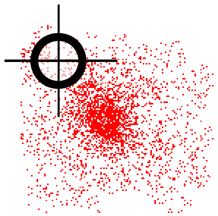
Real Data Samples

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

$$\lim_{\|\mathbf{x}\|\to\infty} \|\mathbf{x}\|^d \, K(\mathbf{x}) = 0$$

- ???

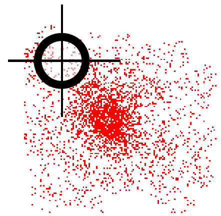
$$\int_{R^d} \mathbf{x}\mathbf{x}^T K(\mathbf{x})d\mathbf{x} = c\mathbf{I}$$

Kernel Density Estimation

Parzen Windows - Function Forms

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

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



Data

In practice one uses the forms:

$$K(\mathbf{x}) = c\prod_{i=1}^{d} k(x_i)$$ or $$K(\mathbf{x}) = ck\left(\|\mathbf{x}\|\right)$$

Same function on each dimension

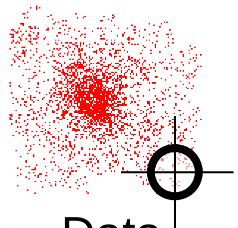
Function of vector length only

Kernel Density Estimation

Various Kernels

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

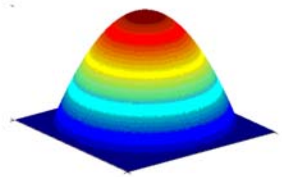
A function of some finite number of data points $x_1 \ldots x_n$



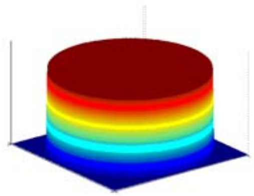
Data

Examples:

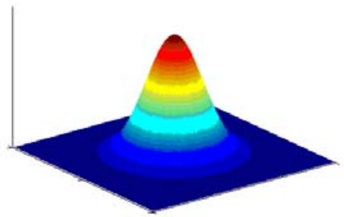
- Epanechnikov Kernel $$K_E(\mathbf{x}) = \begin{cases} c\left(1 - \|\mathbf{x}\|^2\right) & \|\mathbf{x}\| \le 1 \\ 0 & \text{otherwise} \end{cases}$$



- 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

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 :

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

$$g(\mathbf{x}) = -k'(\mathbf{x})$$

Kernel Density Estimation

Computing The Mean Shift

Gradient

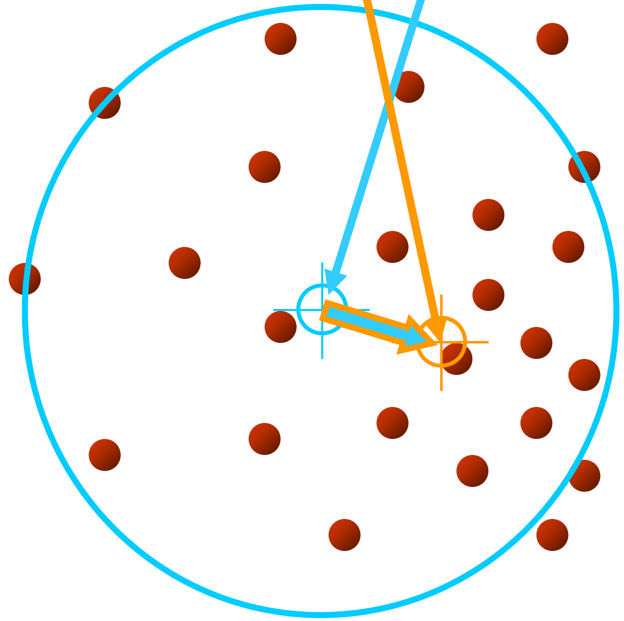
$$\nabla P(\mathbf{x}) = \frac{c}{n}\sum_{i=1}^{n} \nabla k_i = \frac{c}{n}\left[\sum_{i=1}^{n} g_i\right] \cdot \left[\frac{\sum_{i=1}^{n} \mathbf{x}_i g_i}{\sum_{i=1}^{n} g_i} - \mathbf{x}\right]$$

$$g(\mathbf{x}) = -k'(\mathbf{x})$$

Computing The Mean Shift

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

Yet another Kernel
density estimation !



Simple Mean Shift procedure:

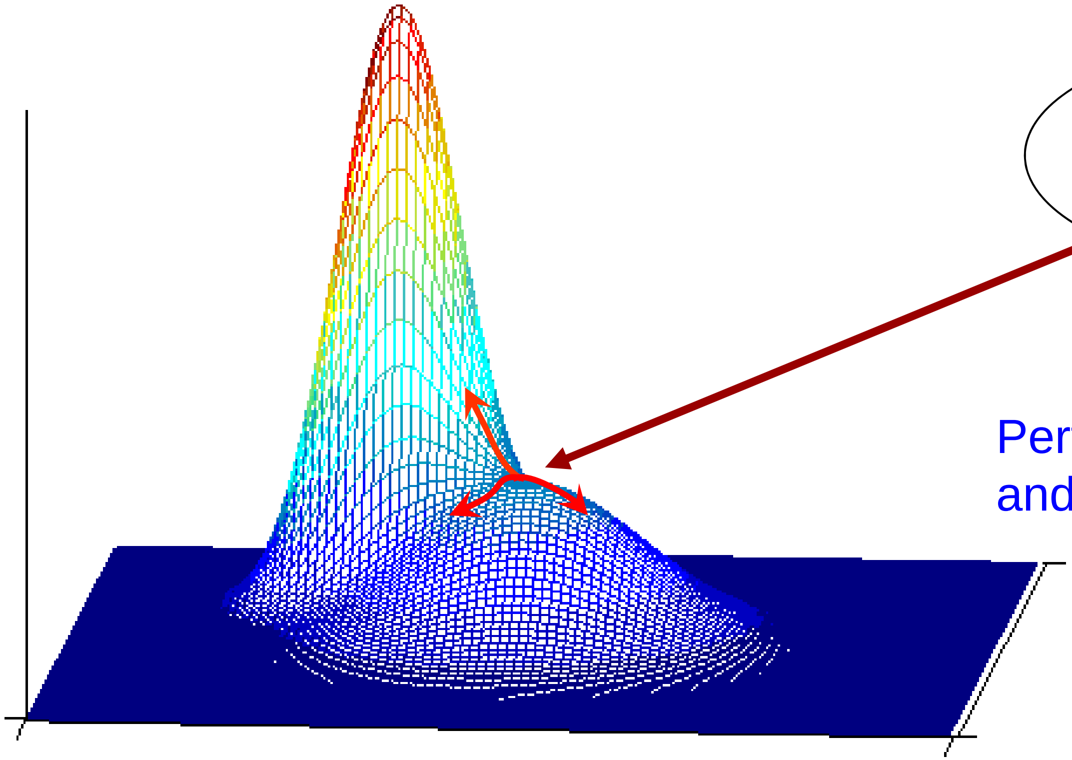
- Compute mean shift vector

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

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

$$g(\mathbf{x}) = -k'(\mathbf{x})$$

Mean Shift Mode Detection



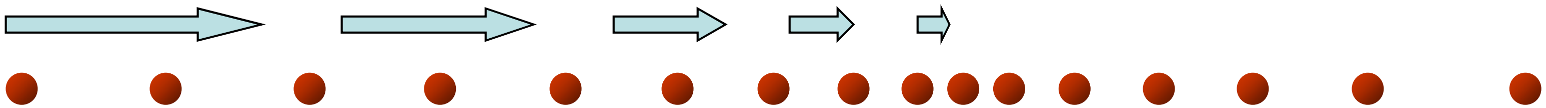
What happens if we reach a saddle point ?

Perturb the mode position and check if we return back

Updated Mean Shift Procedure:

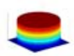
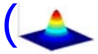
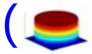
- Find all modes using the Simple Mean Shift Procedure
- Prune modes by perturbing them (find saddle points and plateaus)
- Prune nearby – take highest mode in the window

Mean Shift Properties

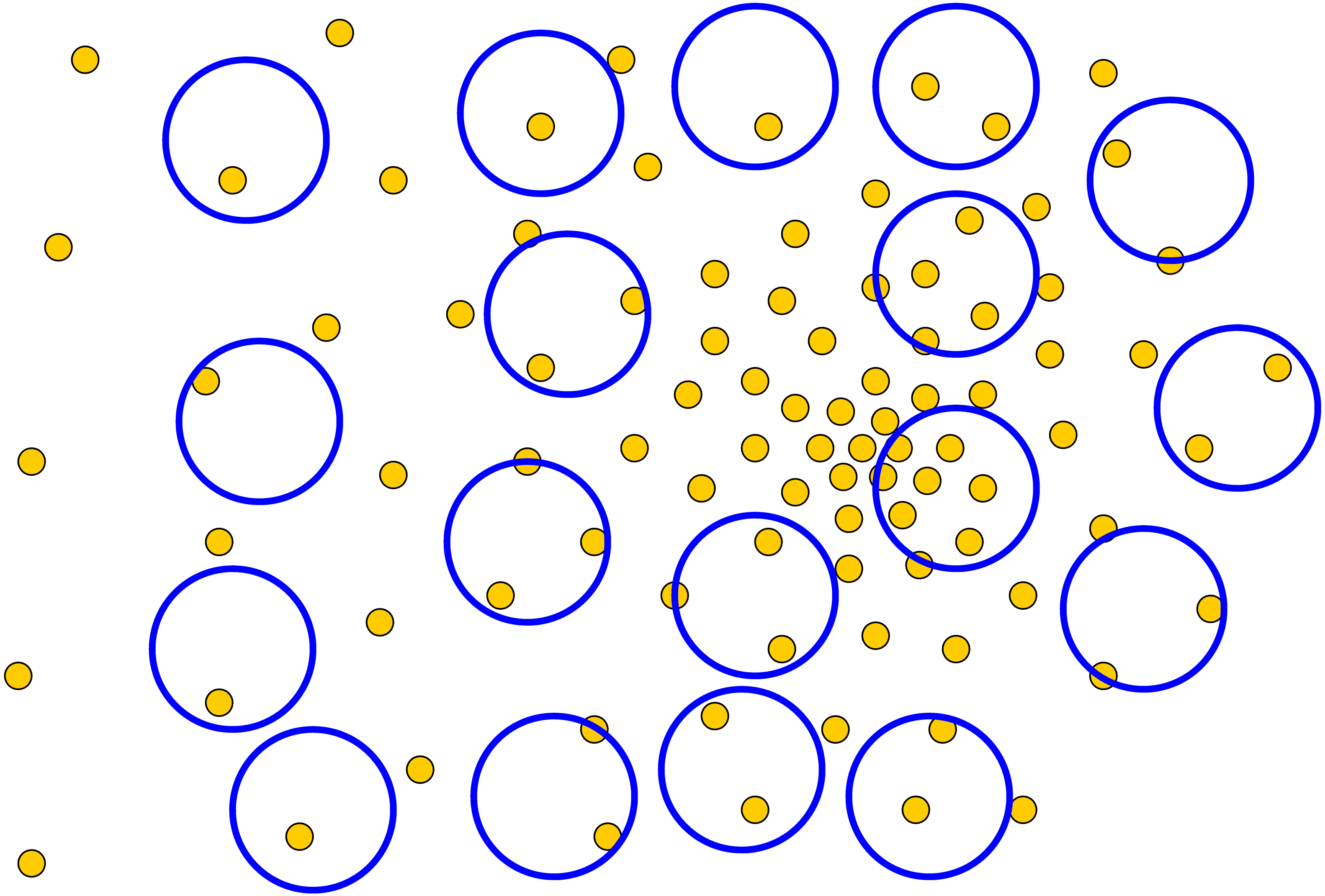


- Automatic convergence speed – the mean shift vector size depends on the gradient itself.
- Near maxima, the steps are small and refined

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 ().

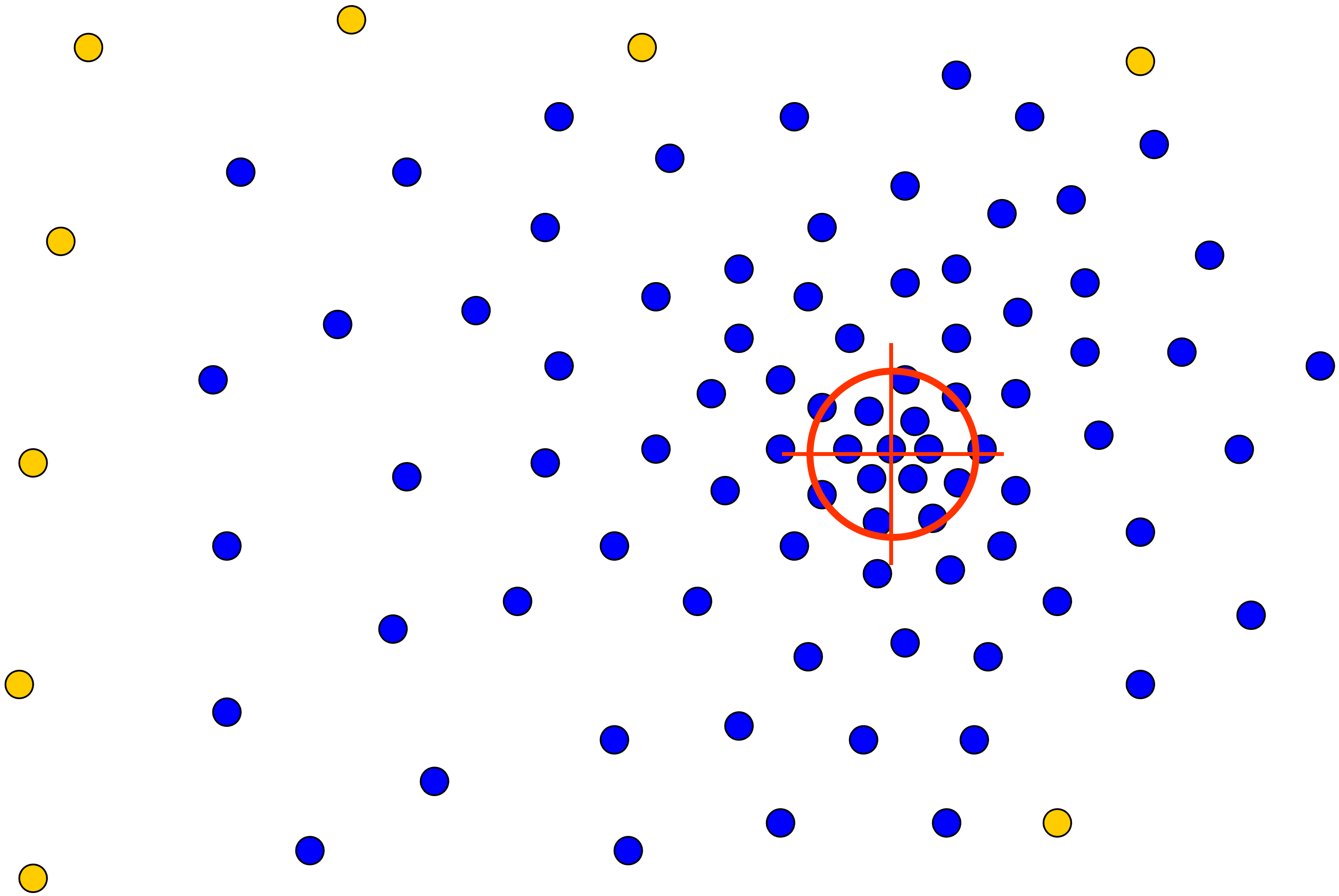
Real Modality Analysis



Tessellate the space with windows

Run the procedure in parallel

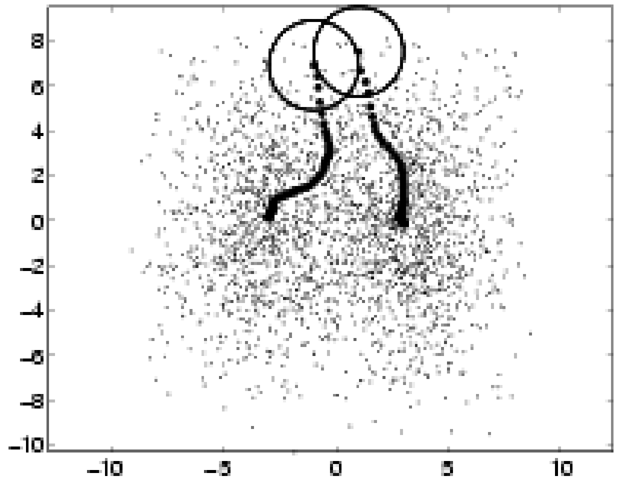
Real Modality Analysis



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

Real Modality Analysis

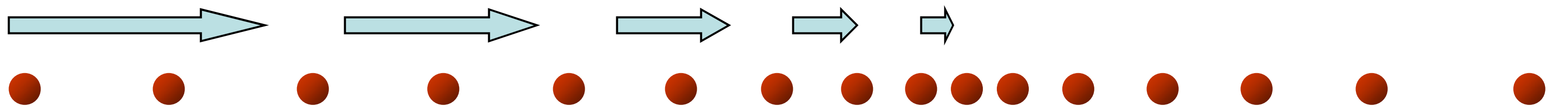
An example



Window tracks signify the steepest ascent directions

Adaptive Mean Shift

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 :

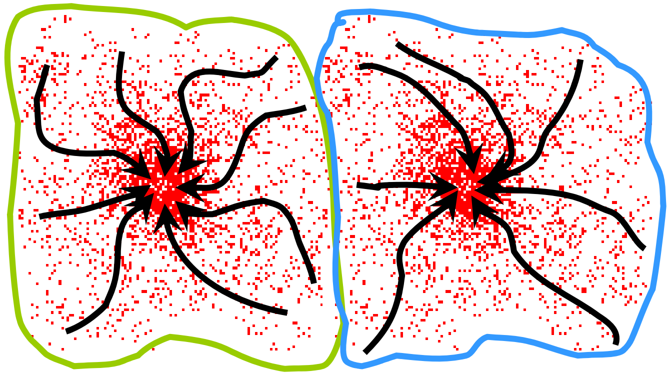
- The window size (bandwidth selection) is not trivial
- Inappropriate window size can cause modes to be merged, or generate additional “shallow” modes ➔ Use adaptive window size

Mean Shift Applications

Clustering

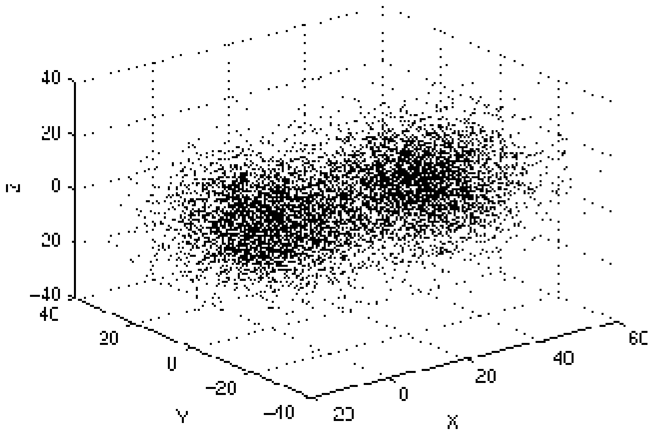
Cluster : All data points in the ***attraction basin*** of a mode

Attraction basin : the region for which all trajectories lead to the same mode



Clustering

Synthetic Examples



Simple Modal Structures

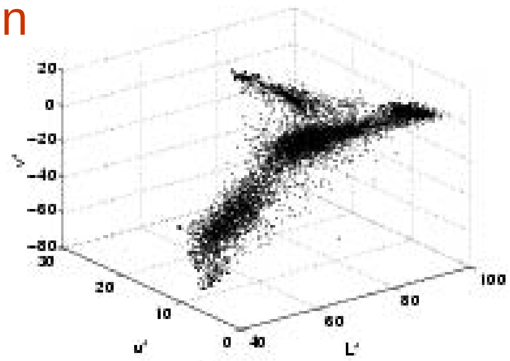
Complex Modal Structures

Clustering

Real Example

Feature space:
L*u*v representation

Initial window
enters

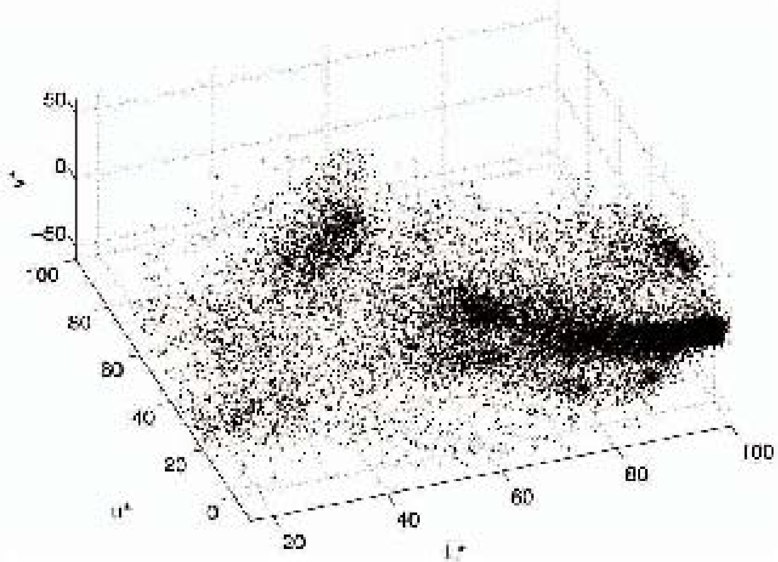


(a)

M

pruning

Clustering
Real Example



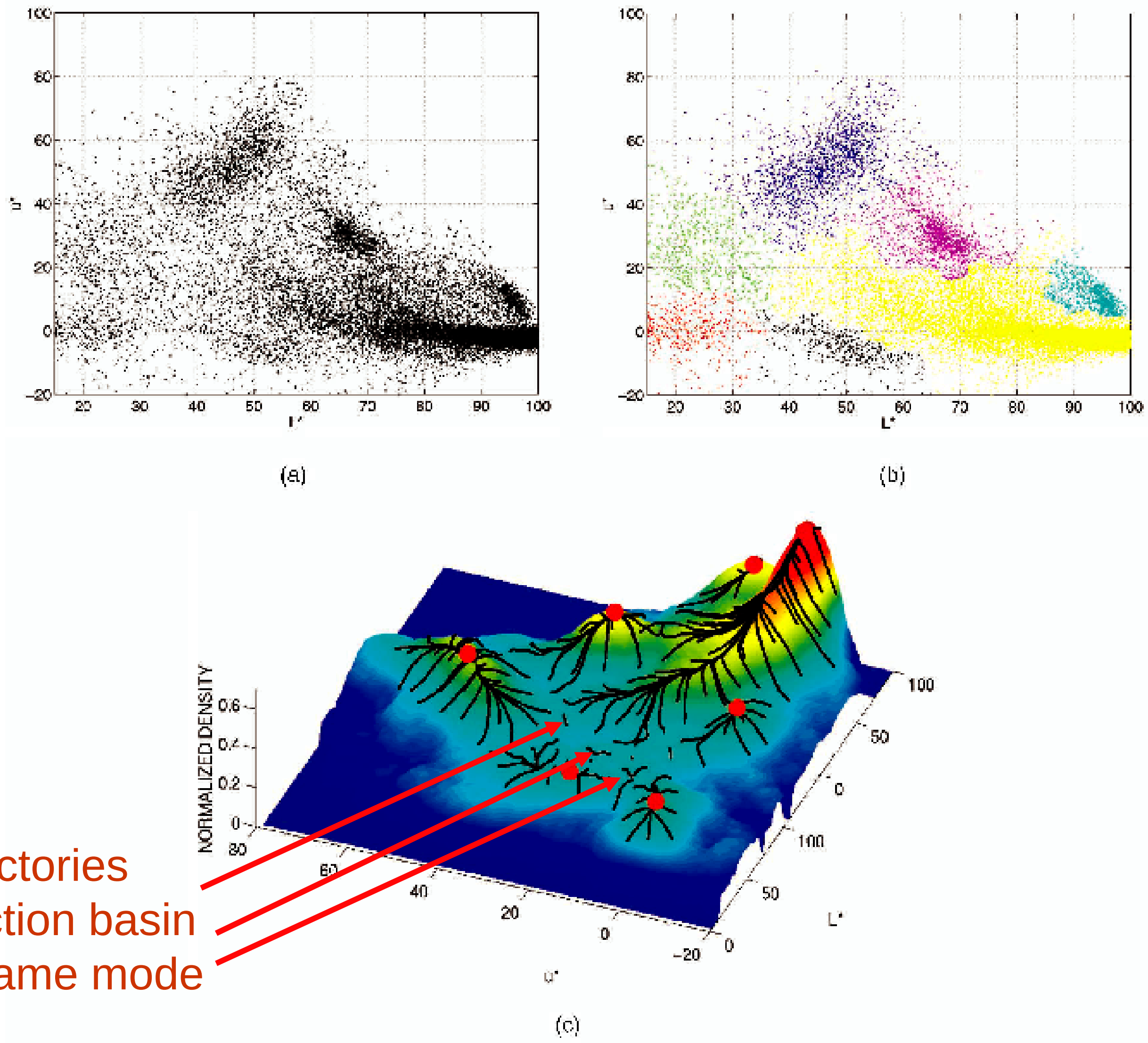
L*u*v space representation

Clustering

Real Example

2D (L*u) space representation

Final clusters



Not all trajectories in the attraction basin reach the same mode

Discontinuity Preserving Smoothing

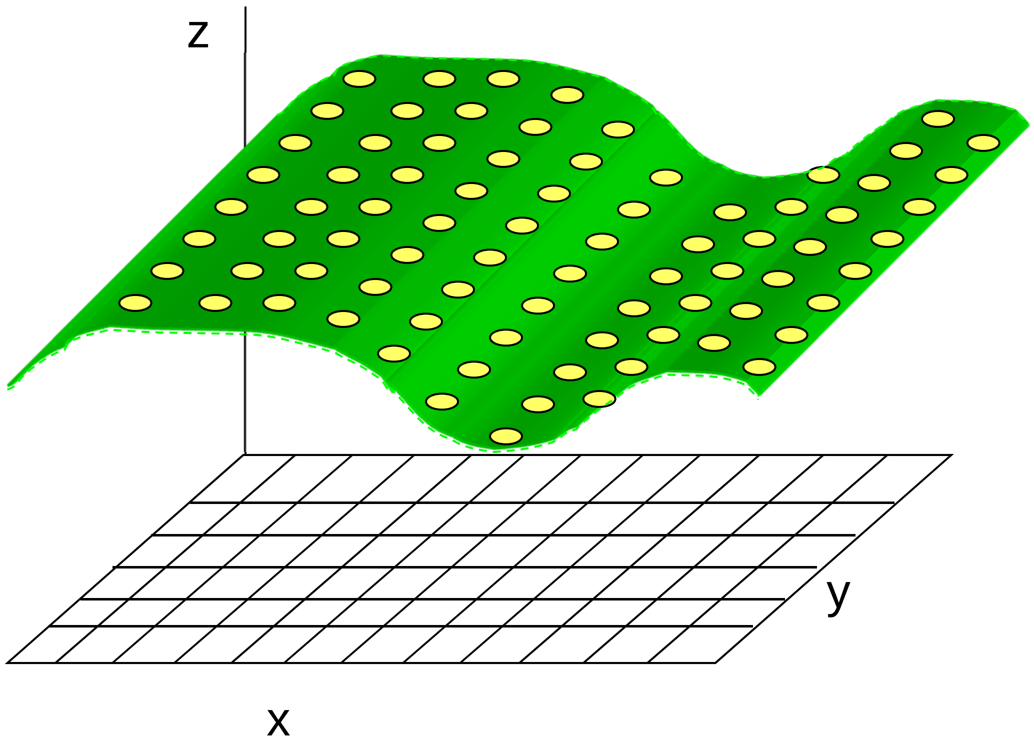
Feature space : Joint domain = spatial coordinates + color space

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

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

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

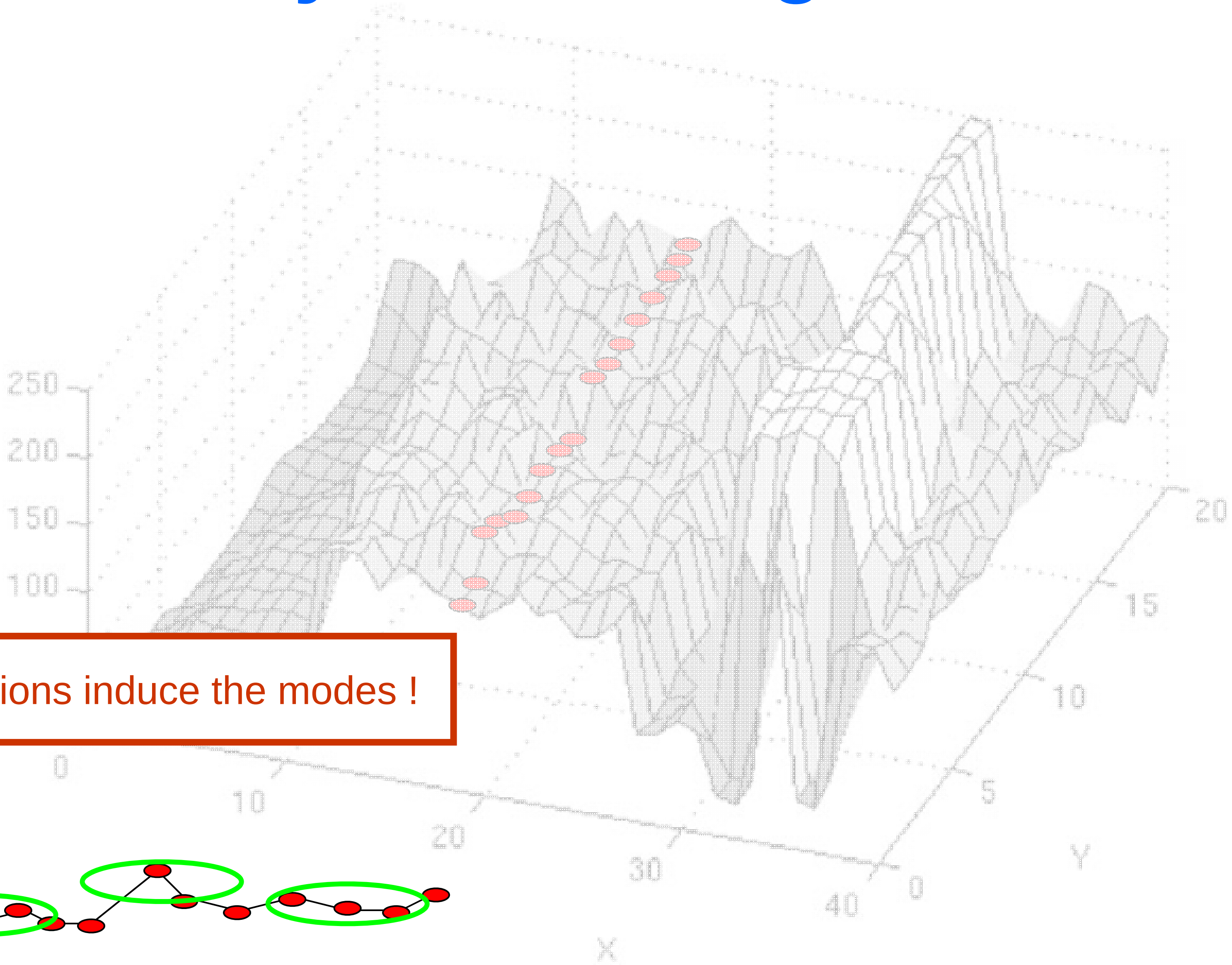
Discontinuity Preserving Smoothing



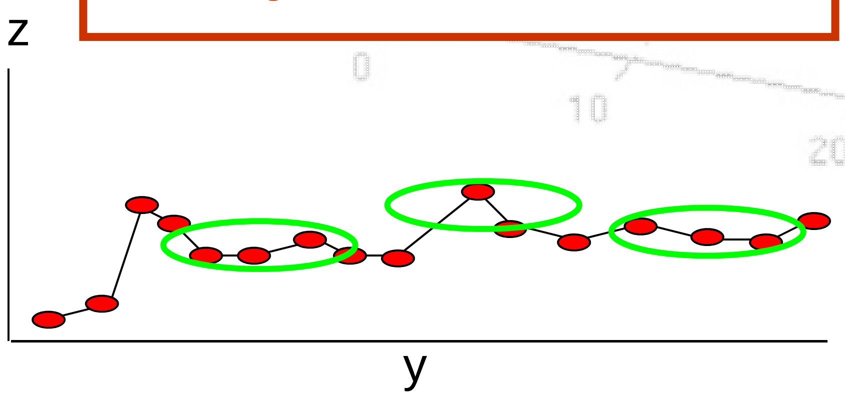
The image gray levels…

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

Discontinuity Preserving Smoothing

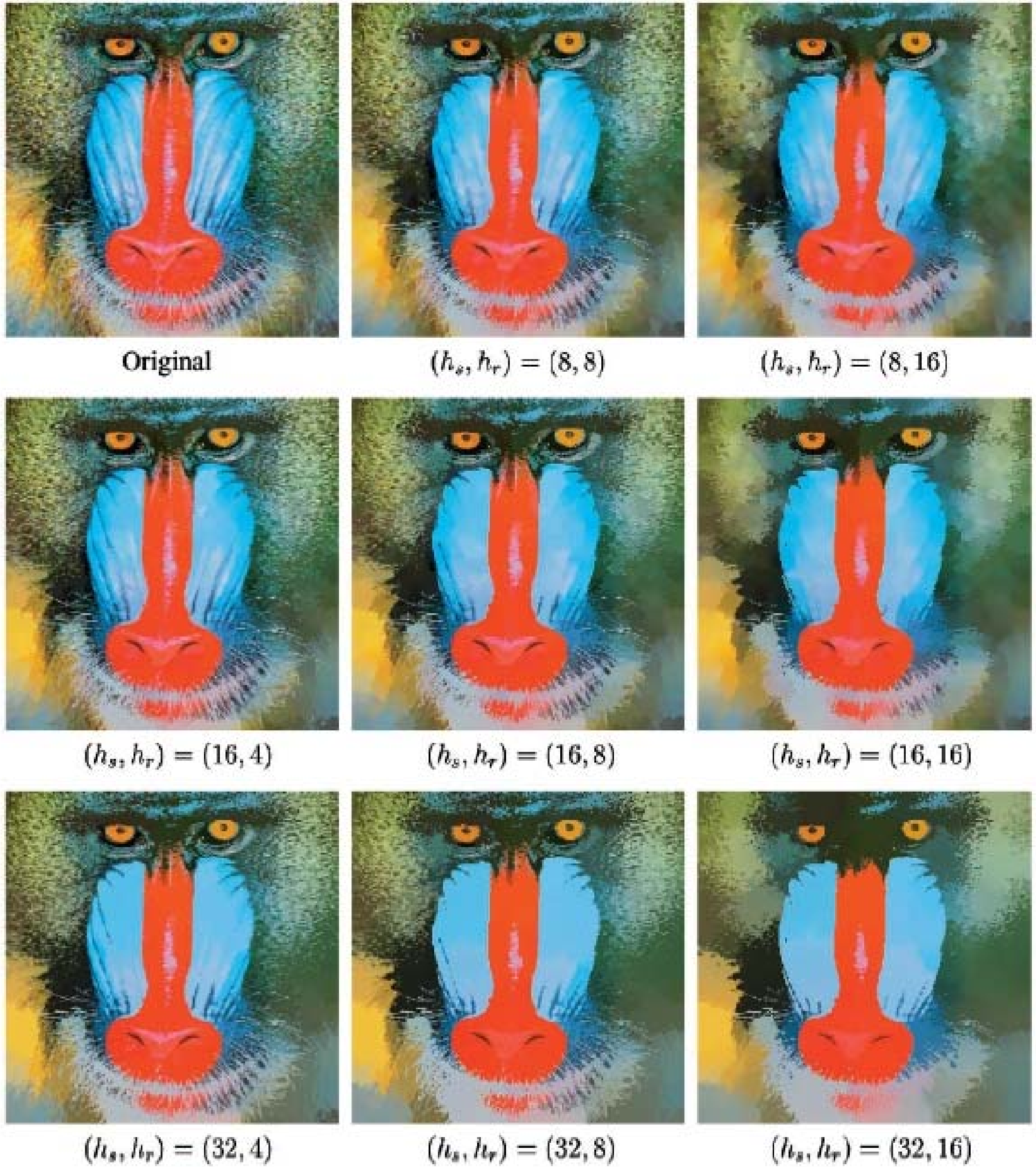


Flat regions induce the modes !



Discontinuity Preserving Smoothing

The effect of window size in spatial and range spaces



Discontinuity Preserving Smoothing

Example

Discontinuity Preserving Smoothing

Example

Segmentation

Segment = Cluster,   or Cluster of Clusters

Algorithm:

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

Segmentation
Example

…when feature space is only gray levels…

Segmentation

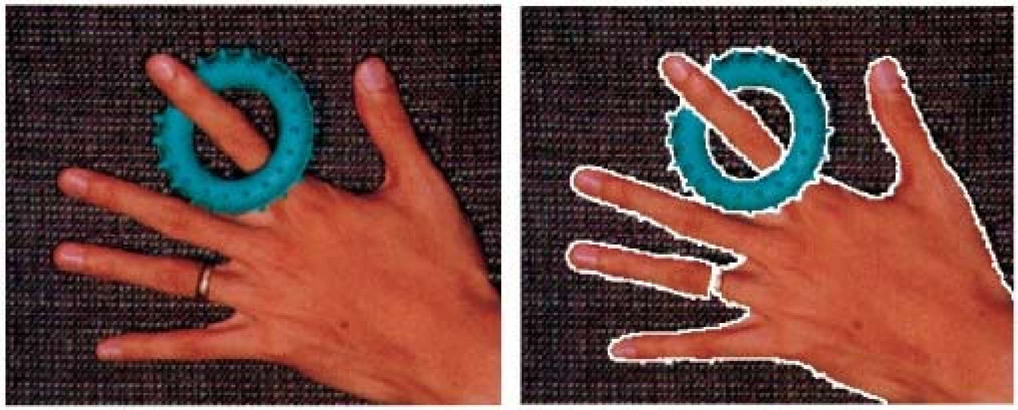
Example

Segmentation

Example

Segmentation

Example

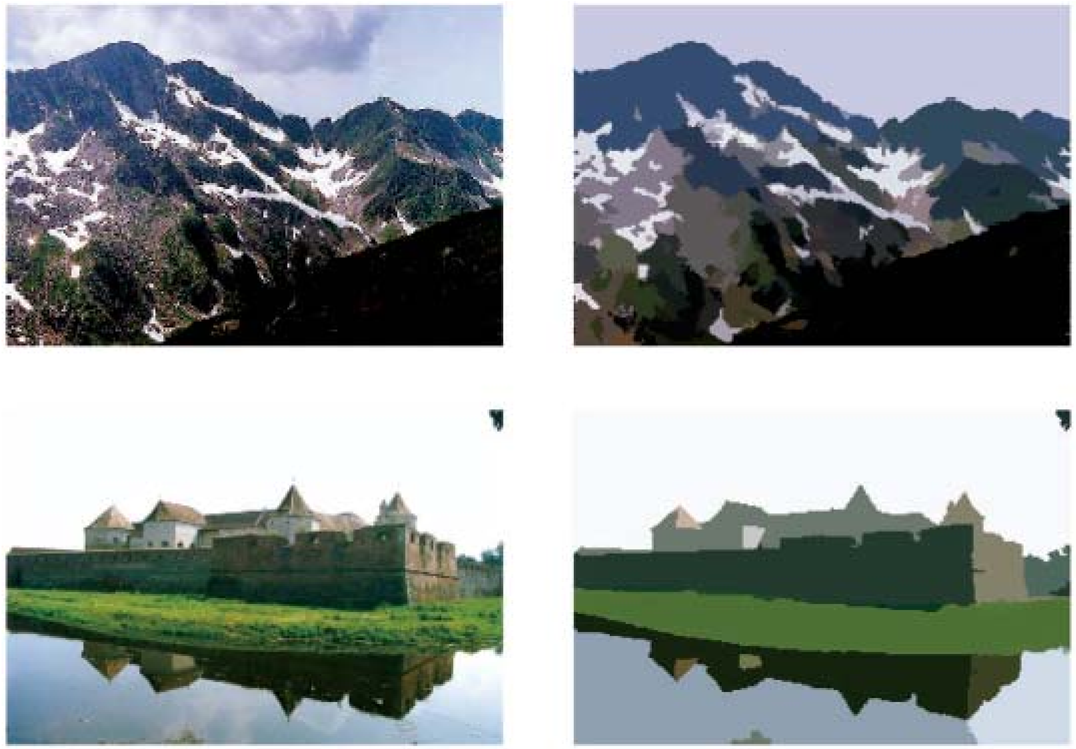


Segmentation

Example

Segmentation

Example



Segmentation

Example