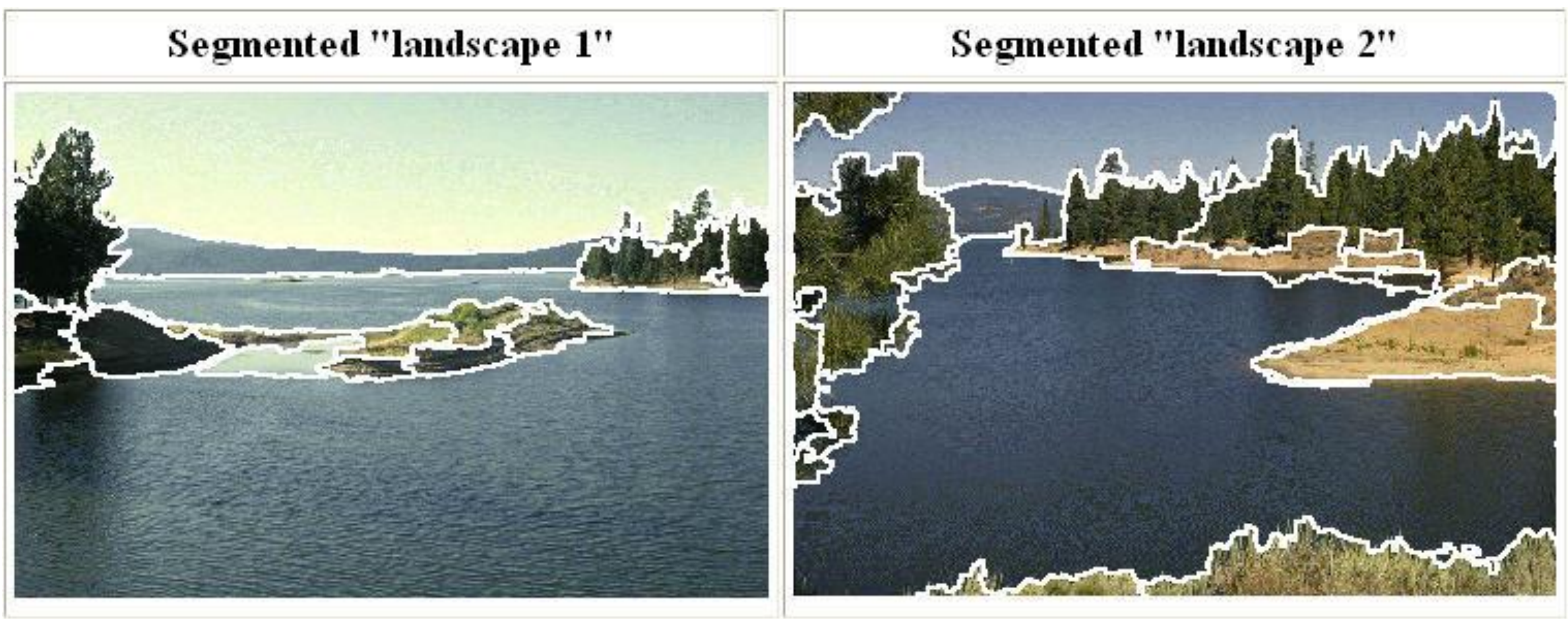
Mean shift segmentation

D. Comaniciu and P. Meer, Mean Shift: A Robust Approach toward Feature Space Analysis,
PAMI 2002, v.24, 603-619.

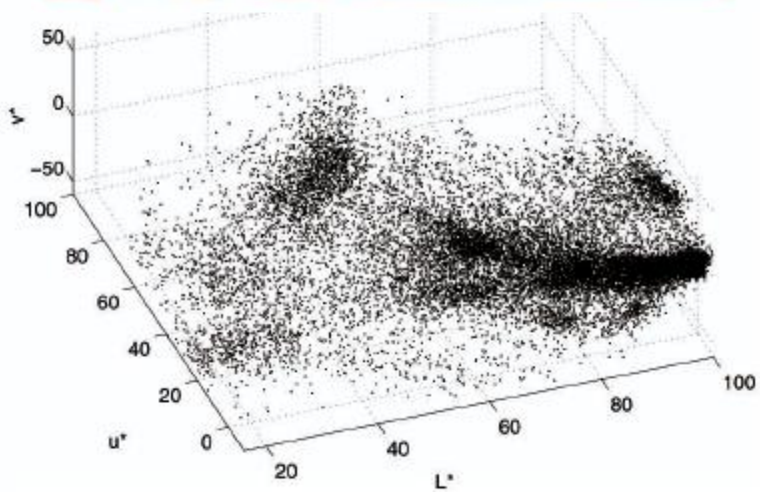
Versatile technique for clustering-based segmentation.

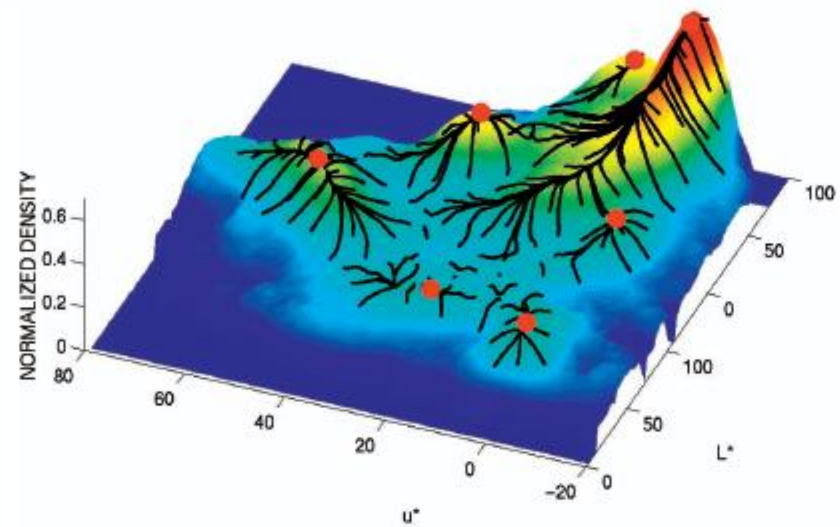
Uses the L*u*v* color space which is also perceptually uniform.
A nonlinear transformation from RGB.

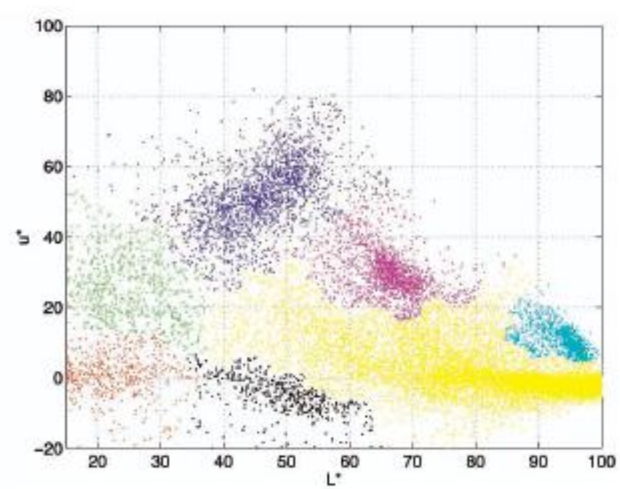


Mean shift algorithm…

…try to find *modes* of this non-parametric density.







Kernel density estimation

Kernel density estimation function

$$\widehat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

K(x) > 0 only for ||x|| <= 1

the bandwidth, h, has to be given by the user.

The kernel is symmetric are depents on x^2.

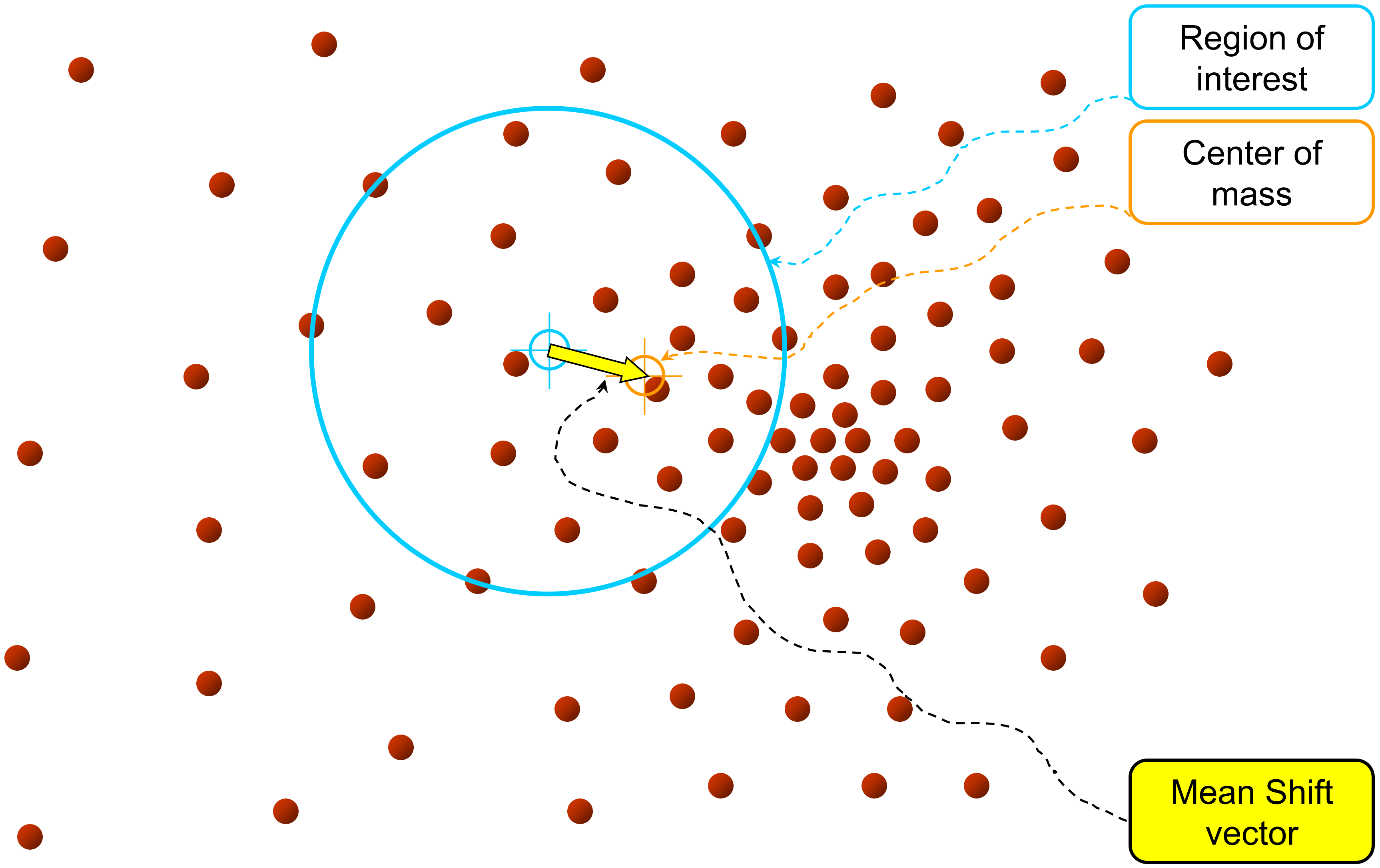
The Epanachikov kernel ~$(1 - ||x||^2)$

and the truncated Gaussian kernel

$$K\left(\frac{x - x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h^2}}$$

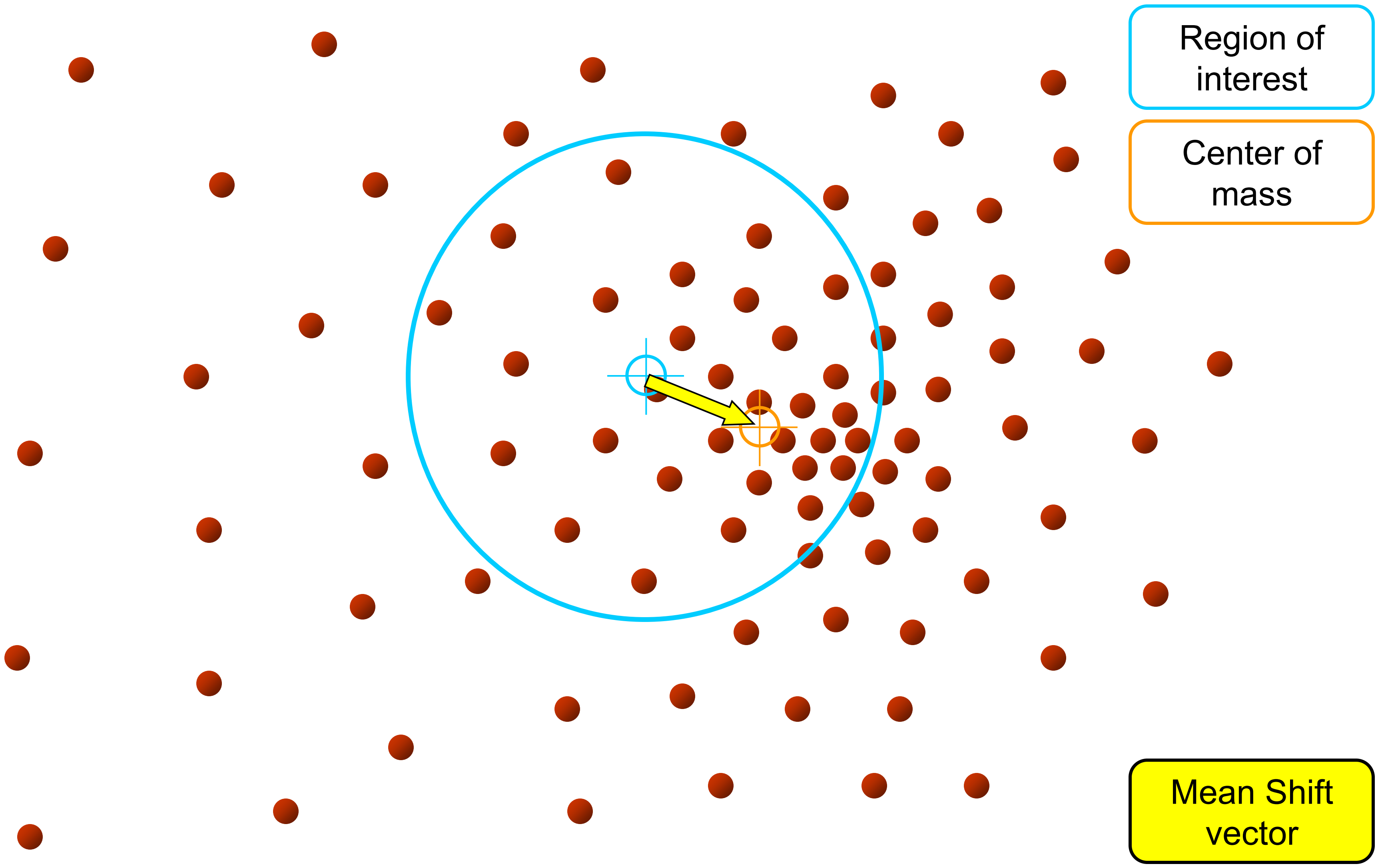
are used.

Mean shift



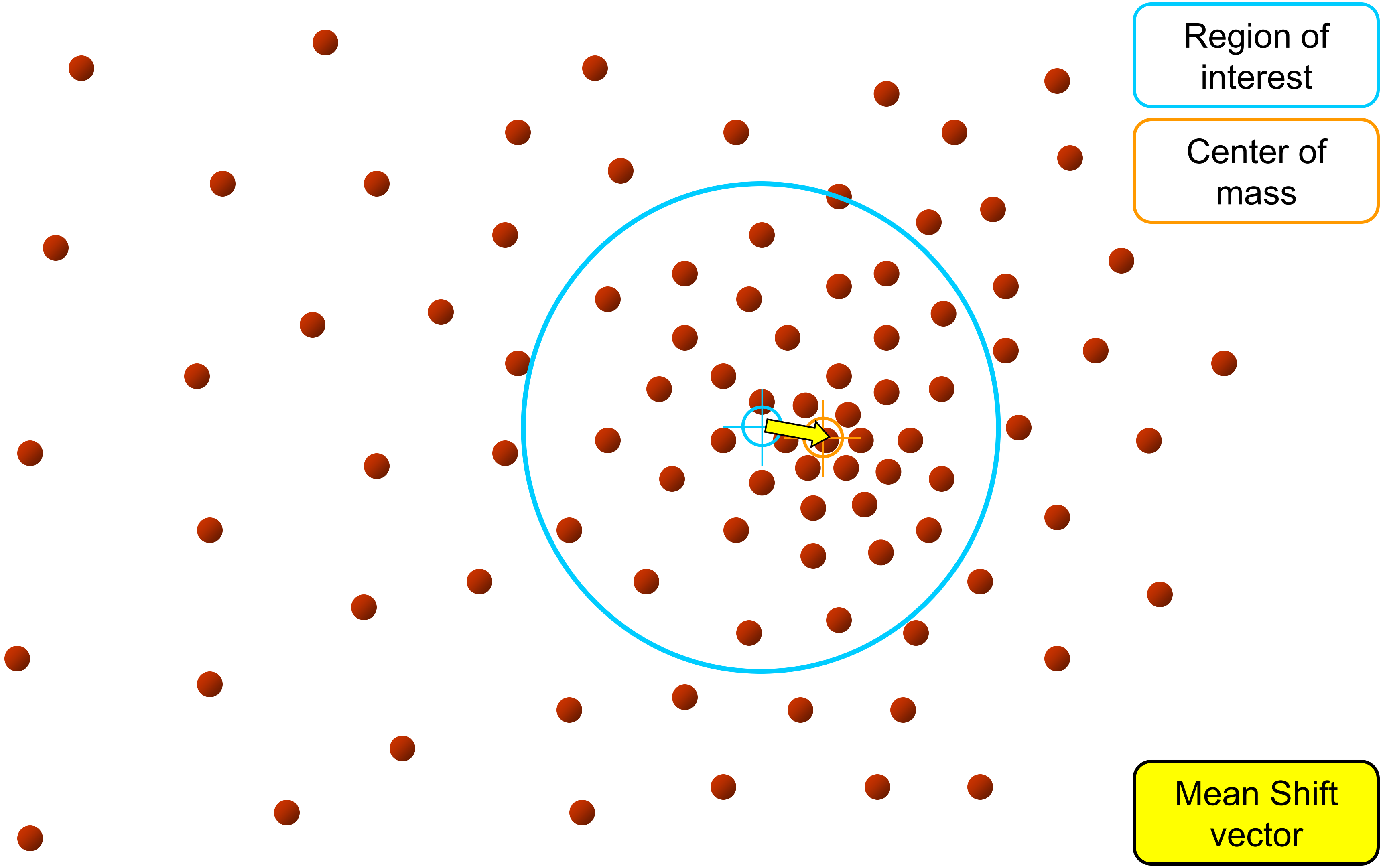
Slide by Y. Ukrainitz & B. Sarel

Mean shift



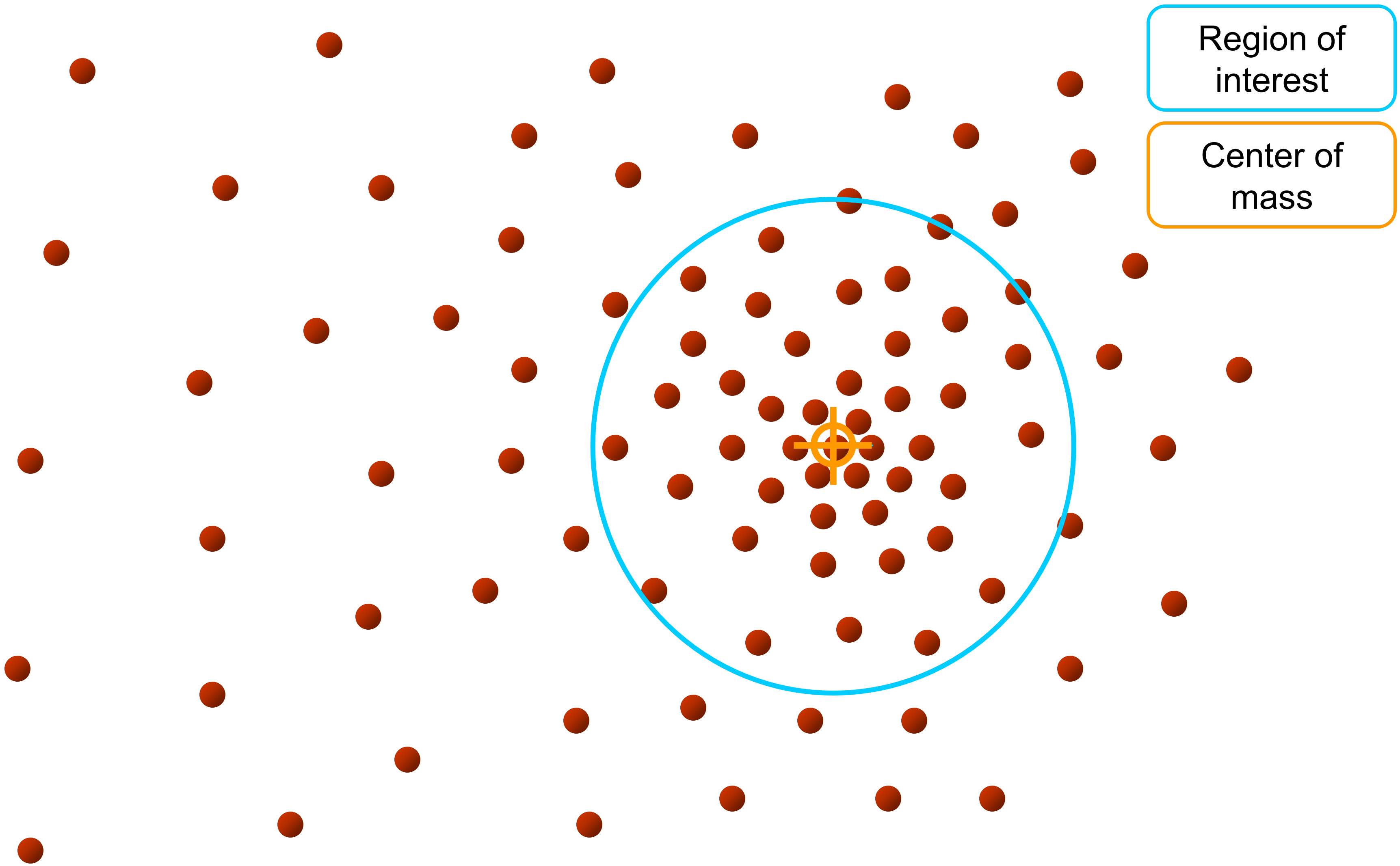
Slide by Y. Ukrainitz & B. Sarel

Mean shift



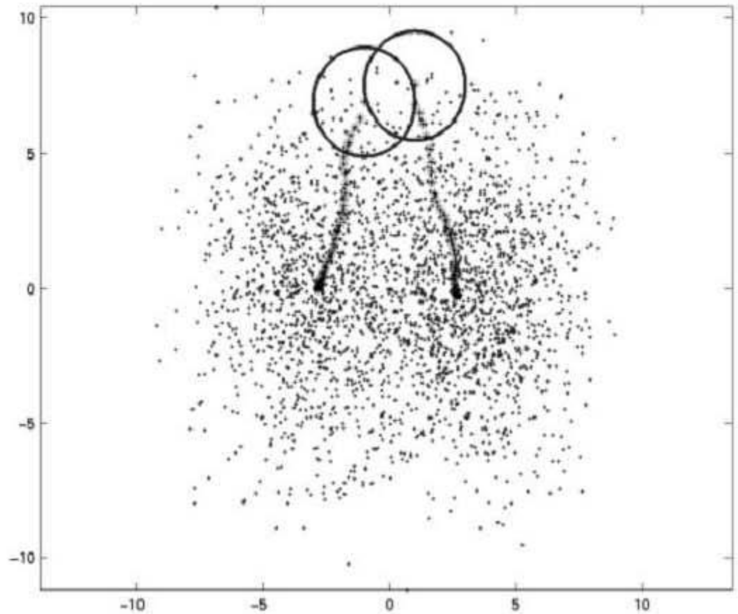
Slide by Y. Ukrainitz & B. Sarel

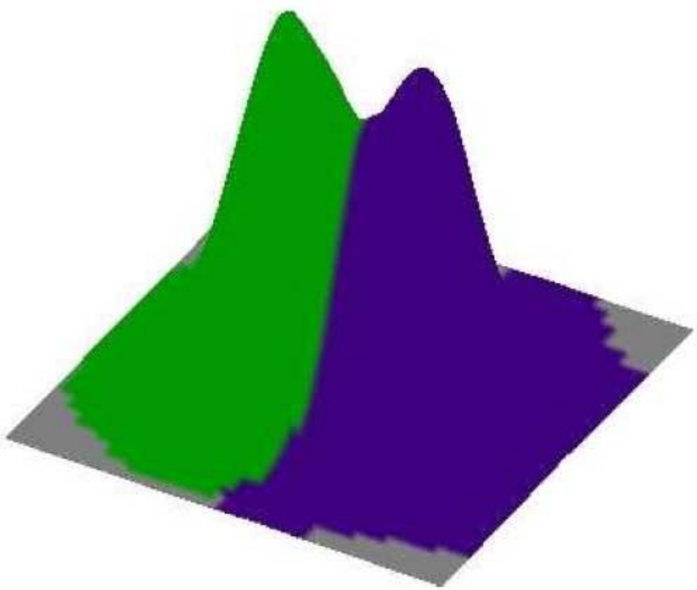
Mean shift



Slide by Y. Ukrainitz & B. Sarel

Two synthetic Gaussian modes, but is not used in mean shift.
The only parameter used is h.





The two kernels converge to different modes in spite starting from overlapping regions.

Computing the Mean Shift

$k(x^2) = K(x)$ profile of the kernel

$g(x) = -k'(x)$

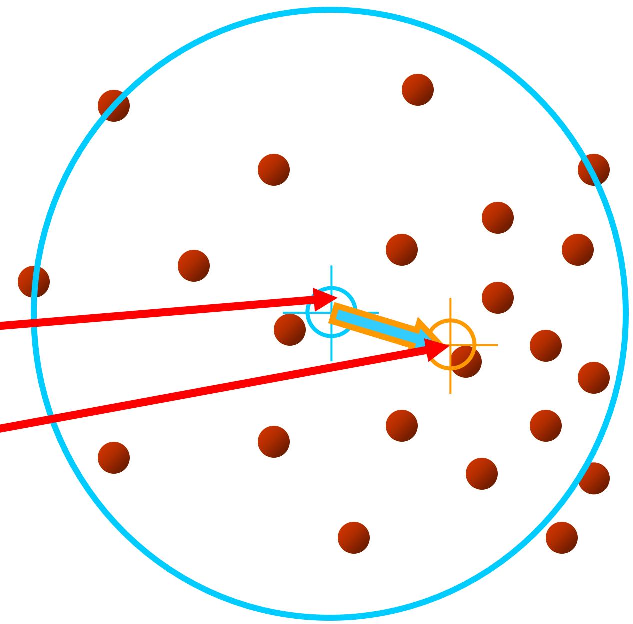
$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

gradient f(x) = 0

in each iteration

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

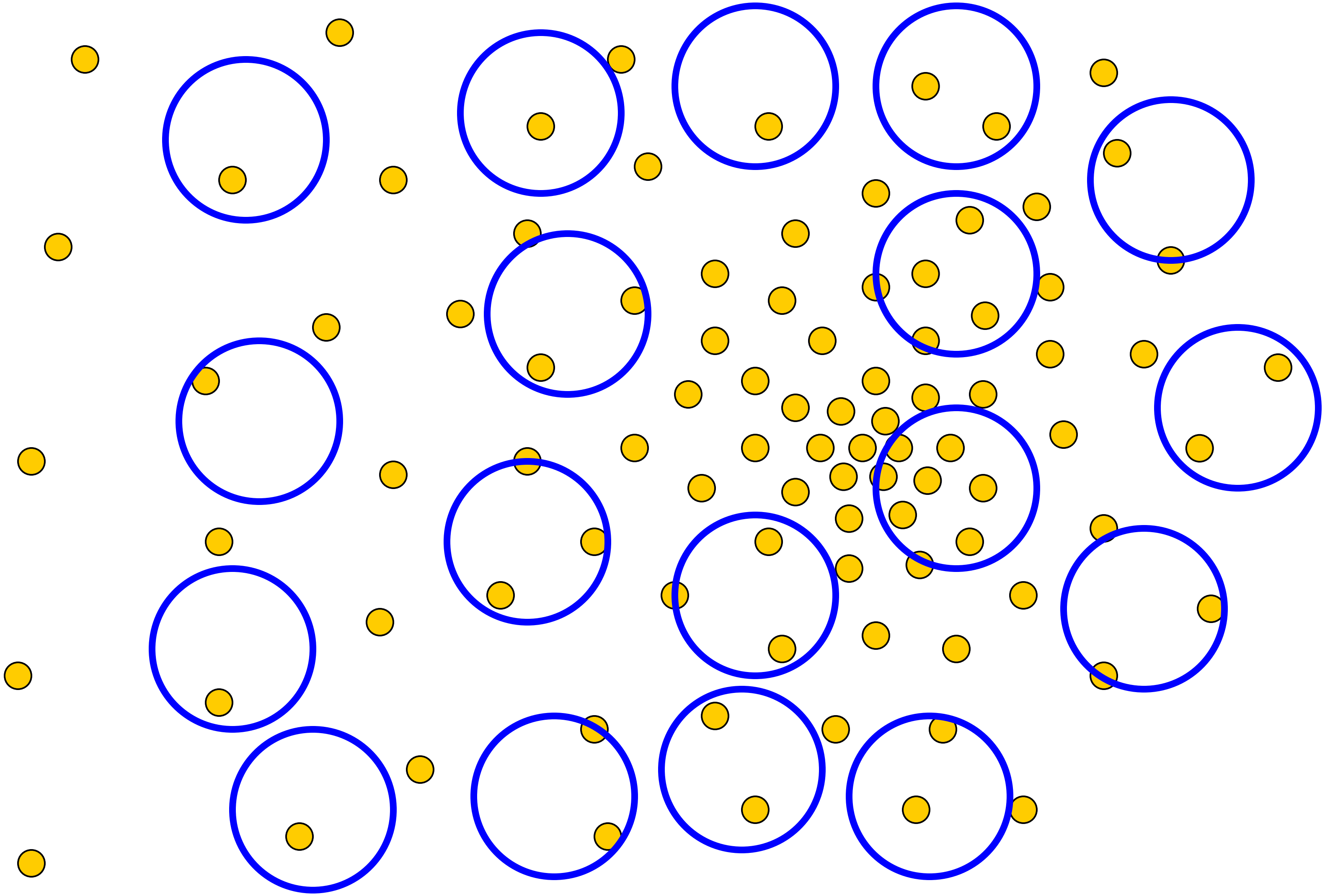
$$\mathbf{m}(\mathbf{x}) = \left[\frac{\sum_{i=1}^{n} \mathbf{x}_i \, g\left(\frac{\left\|\mathbf{x}_{\text{old}} - \mathbf{x}_i\right\|^2}{h}\right)}{\sum_{i=1}^{n} g\left(\frac{\left\|\mathbf{x}_{\text{old}} - \mathbf{x}_i\right\|^2}{h}\right)} - \mathbf{x}_{\text{new}} \right]$$



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

Slide by Y. Ukrainitz & B. Sarel

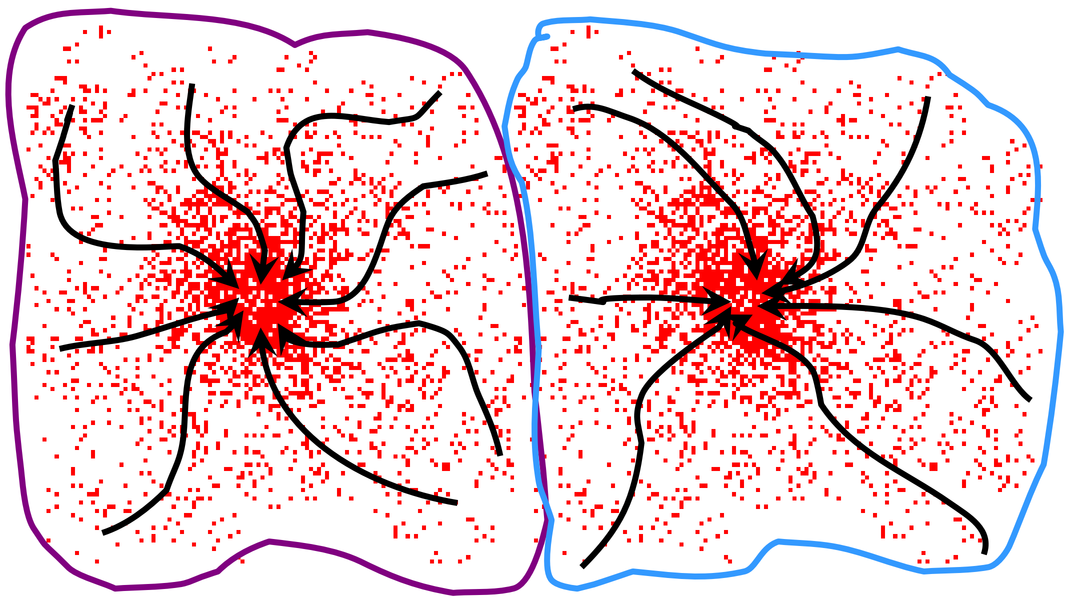
Modality Analysis



- *Tessellate* the space with windows.
- Merge windows that end up *near* the same mode (peak).

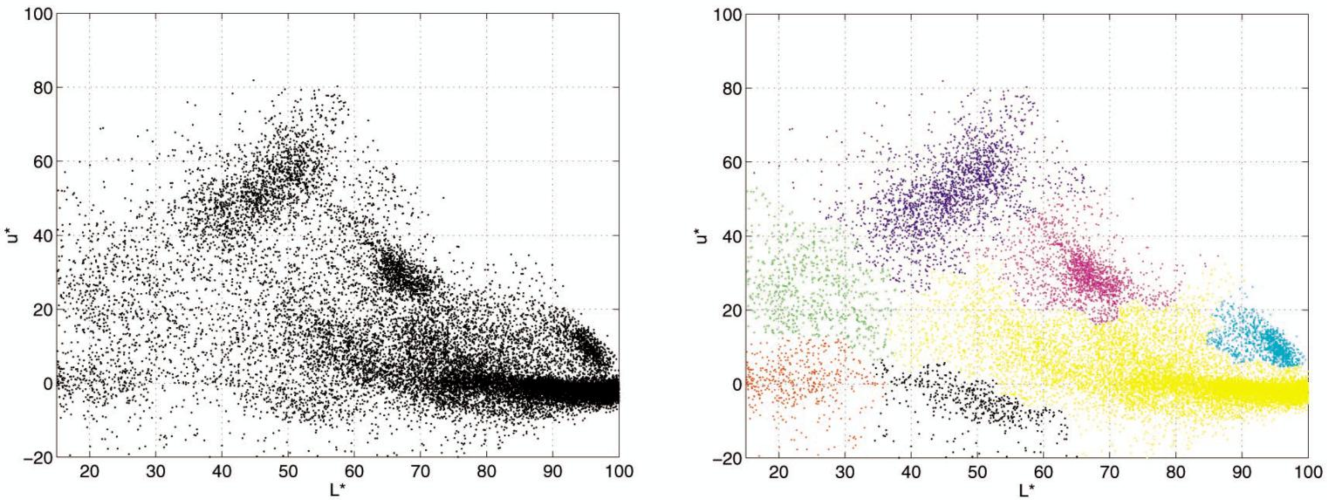
Attraction basin

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

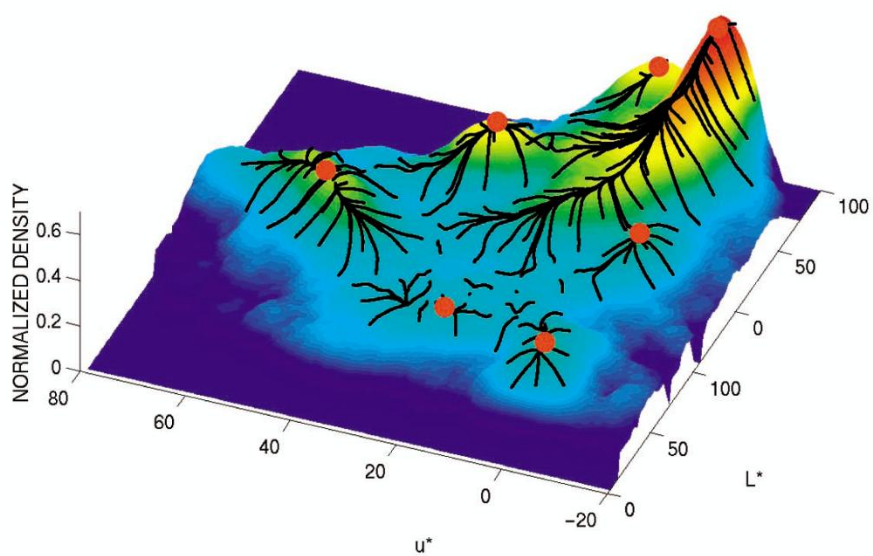


Slide by Y. Ukrainitz & B. Sarel

Example: attraction basins



(a) (b)



zero gradient, g(x)=0
but not a maximum
stationary point

eliminated
by shifting a little bit
the trajectory

A color pixel is represented by a *spatial* and a *range* bandwidth.

a two-dimensional spatial bandwidth: h_s
a three-dimensional range bandwidth: h_r

The user have to give only this two parameters. They are not very strict like in k-means.

$$K_{h_s,h_r}(\mathbf{x}) = \frac{C}{h_s^2 h_r^p} k\left(\left\|\frac{\mathbf{x}^s}{h_s}\right\|^2\right) k\left(\left\|\frac{\mathbf{x}^r}{h_r}\right\|^2\right)$$

2 + 3 = 5 dimensions

Gray level images have one-dimensional range bandwidth.

Example of a window's convergence.

256 x 256

(a)

(b)

Cameraman image. (a) Original. (b) Mean shift filtered $(h_s, h_r) = (8, 4)$.

(8,4)

original
(inverted)

mean shift

filtering

segmentation

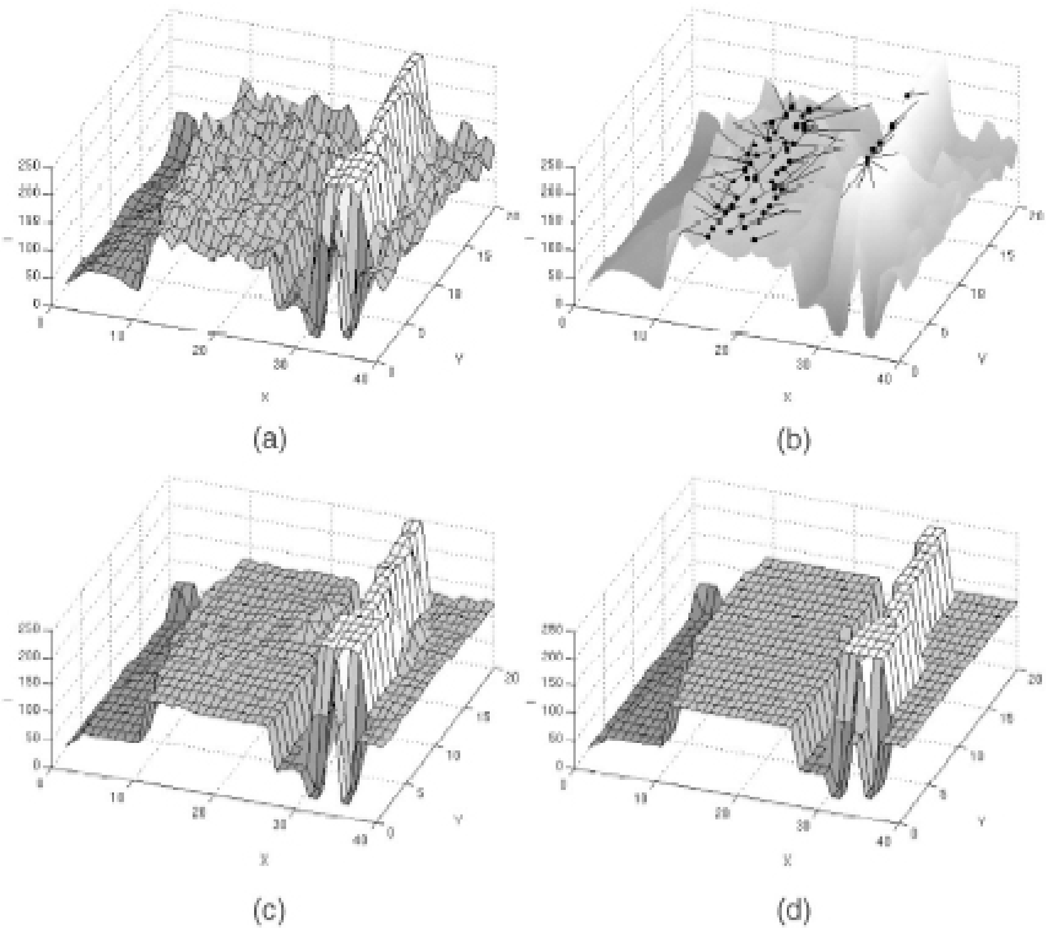


Fig. 4. Visualization of mean shift-based filtering and segmentation for gray-level data. (a) Input. (b) Mean shift paths for the pixels on the plateau and on the line. The black dots are the points of convergence. (c) Filtering result $(h_s, h_r) = (8, 4)$. (d) Segmentation result.

Mean shift clustering

- The mean shift algorithm seeks modes of the given set of points.
    1. Choose kernel and two bandwidths.
    2. For each point:
        a) Center a window on that point.
        b) Compute the mean of the data in the search window.
        c) Center the search window at the new mean location.
        d) Repeat (b,c) until convergence.
    3. Assign points that lead to nearby modes to the same cluster.

In segmentation, the means are both *in spatial and range* and the points always converge *to the nearest mode*.

Mean shift clustering are *two* phases:

filtering, as was described before;

segmentation, unify adjacent clusters if they are closer than h_s in the spatial domain and h_r in the range domain. (Step 3.)

EDISON program can do additional things too, but we will not describe them here.

Mean shift was also used for *tracking of motions* and for *nonlinear spaces* through Riemannian manifolds.

Mean shift segmentation results

512
x
512

(a) (b)

Lake image. (a) Original. (b) Segmented with $(h_s, h_r, M) = (16, 7, 40)$.

(16,7,40)

Eliminate spatial regions containing less than M pixels.

256 x 256

(8,7,20)

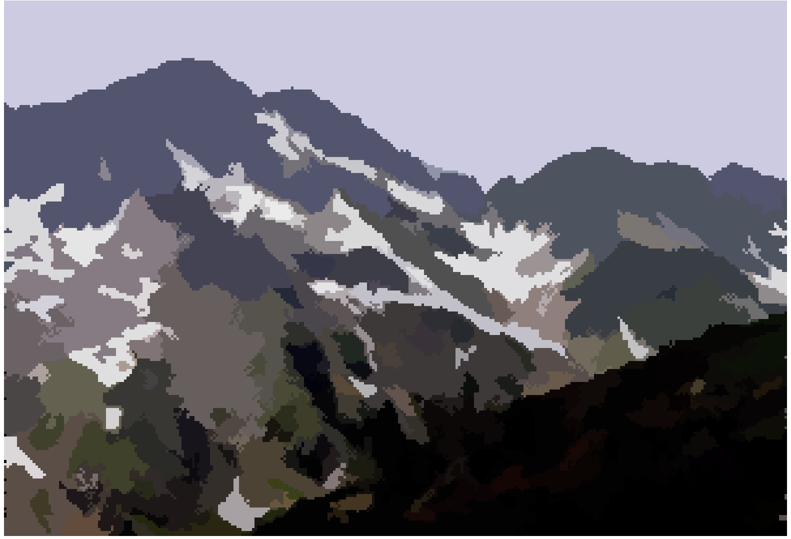
(a) (b) (c)

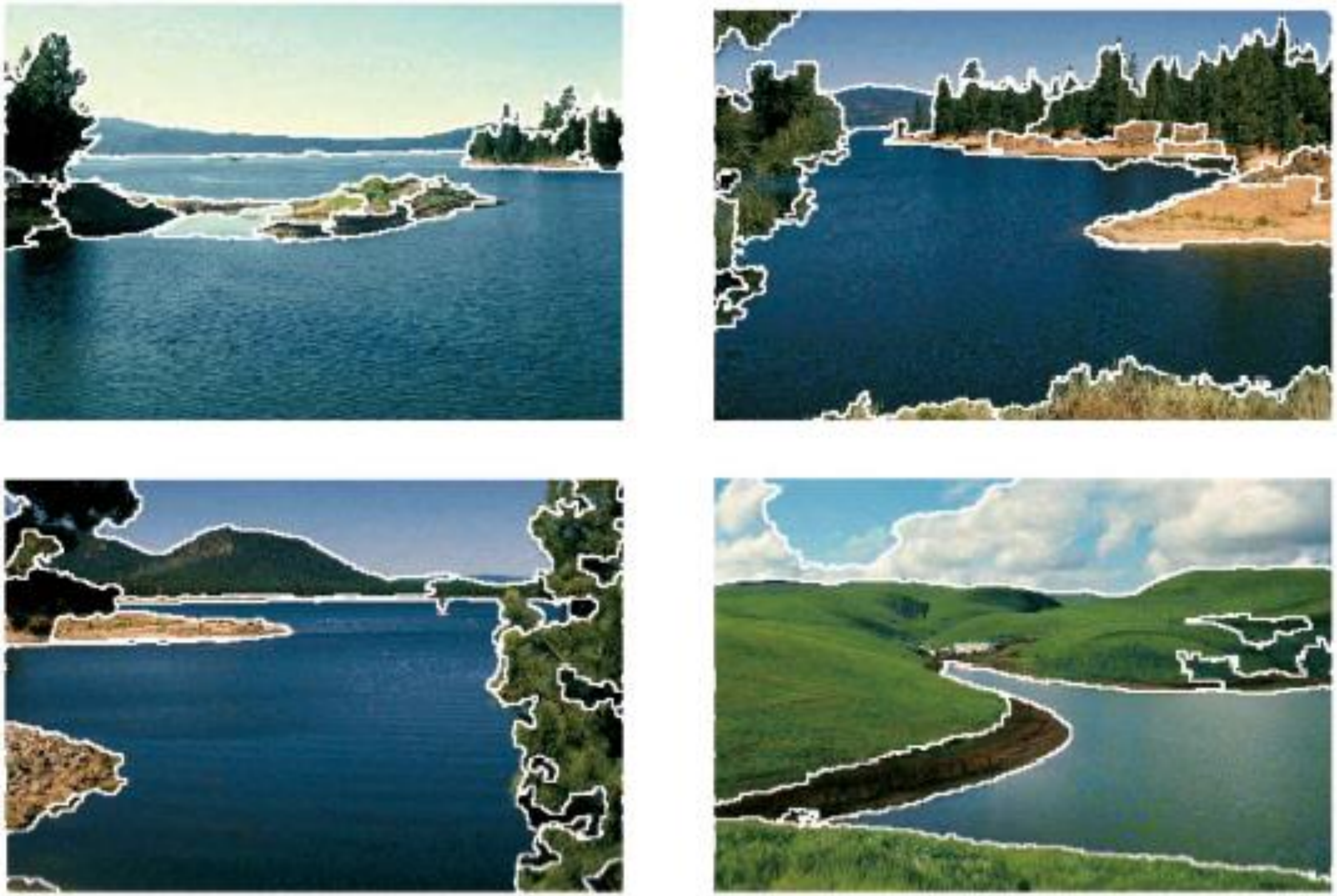
MIT image. (a) Original. (b) Segmented $(h_s, h_r, M) = (8, 7, 20)$. (c) Region boundaries.

The sky changes with location but is still segmented into one.

Sometimes it is possible to take planar surfaces and represent it still as a constant surface.

parameters (8,7,20)





parameters (8,7,100)

Mean shift pros and cons

- Pros
  - Does not assume spherical clusters.
  - Just two parameters (window sizes).
  - Finds variable number of modes, which are *not given.*
  - Robust to outliers and weak nonconstant regions.
- Cons
  - Output depends on window size.
  - Efficient implementation uses on short cuts in the search.
  - Does not scale well directly with dimension of feature space is above ten.