CS 534: Computer Vision
Segmentation III
Statistical Nonparametric Methods for Segmentation

Spring 2004
Ahmed Elgammal
Dept of Computer Science
Rutgers University

Outlines

- Density estimation
- Nonparametric kernel density estimation
- Mean shift
- Mean shift clustering
- Mean shift filtering and segmentation

Statistical Background

Density Estimation: Given a sample $S=\{x_i\}_{i=1..N}$ from a distribution obtain an estimate of the density function $\hat{f}(\cdot)$ at any point.

Parametric : Assume a parametric density family $f(.|\theta)$, (ex. $N(\mu,\sigma^2)$) and obtain the best estimator $\hat{\theta}$ of θ

Advantages:

- Efficient
- Robust to noise: robust estimators can be used

Problem with parametric methods

- An incorrectly specified parametric model has a bias that cannot be removed even by large number of samples.

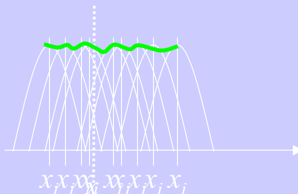
Nonparametric : directly obtain a good estimate $\hat{f}(\cdot)$ of the entire density $f(\cdot)$ from the sample.

Most famous example: Histogram

Kernel Density Estimation

- 1950s + (Fix & Hodges 51, Rosenblatt 56, Parzen 62, Cencov 62)
- Given a set of samples $S=\{x_i\}_{i=1..N}$ we can obtain an estimate for the density at x as:

$$\hat{f}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-x_i}{h}\right) = \frac{1}{N}\sum_{i=1}^{N} K_h(x-x_i)$$



$$\hat{f}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right) = \frac{1}{N}\sum_{i=1}^{N} K_h(x - x_i)$$

where $K_h(t)=K(t/h)/h$ called kernel function (window function)

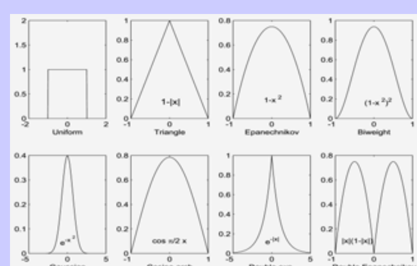
h : scale or bandwidth

K satisfies certain conditions, e.g.:

$$\int K_h(x)dx = 1$$

$$K_h(x) \geq 0$$

Kernel Estimation



- A variety of kernel shapes with different properties.
- Gaussian kernel is typically used for its continuity and differentiability.
- Multivariate case: Kernel Product
  Use same kernel function with different bandwidth h for each dimension.

$$\hat{f}(x) = \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{d} K_{h_j}(x^j - x_i^{\,j})$$

- General form: avoid to store all the samples

$$\hat{f}(x) = \sum_{i=1}^{N} \alpha_i K_h(x - x_i)$$

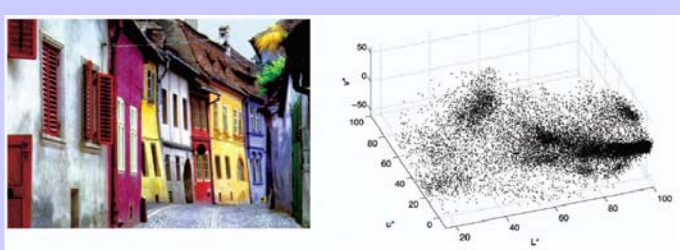
Kernel Density Estimation

Advantages:

- Converge to any density shape with sufficient samples.
  asymptotically the estimate converges to any density.
- No need for model specification.
- Unlike histograms, density estimates are smooth, continuous and differentiable.
- Easily generalize to higher dimensions.
- All other parametric/nonparametric density estimation methods, e.g., histograms, are asymptotically kernel methods.
- In computer vision, the densities are multivariate and multimodal with irregular cluster shapes.

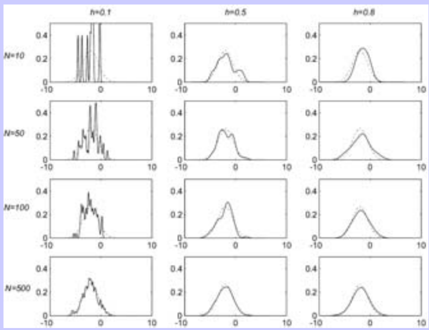
Example: color clusters

- Cluster shapes are irregular
- Cluster boundaries are not well defined.

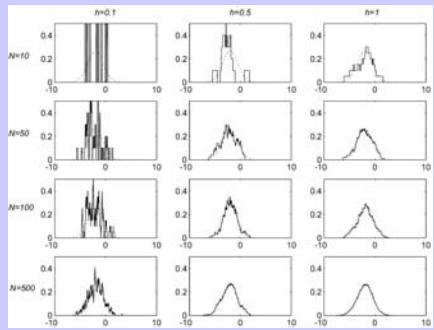


From D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis,"

Conversion - KDE

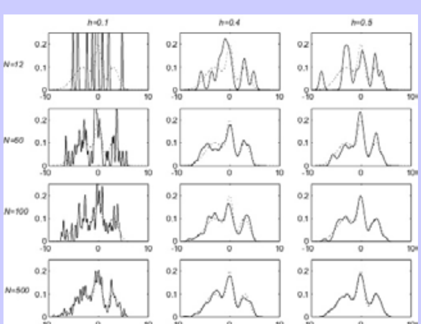


Estimation using Gaussian Kernel

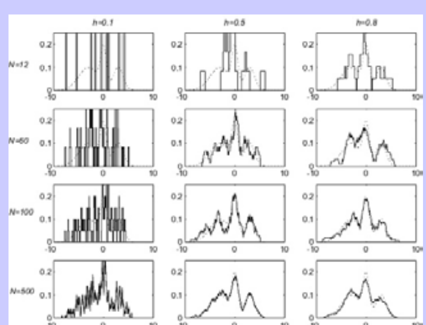


Estimation using Uniform Kernel

Conversion - KDE



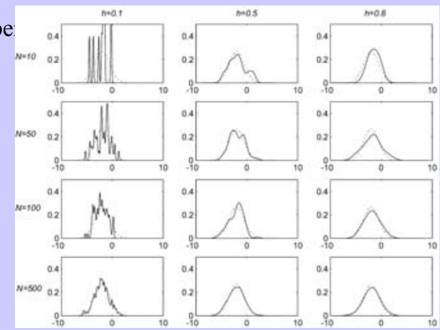
Estimation using Gaussian Kernel



Estimation using Uniform Kernel

Scale selection

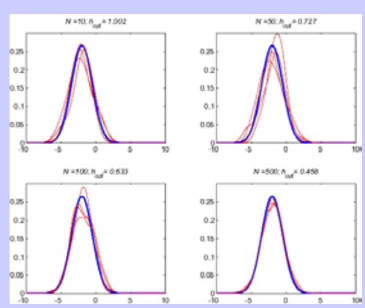
- Important problem. Large literature.
- Small *h* results in ragged densities.
- Large *h* results in over smoothing.
- Best choice for h depends on the number of samples:
  - small *n*, wide kernels
  - large *n*, Narrow kernels
  - $\lim_{n\to\infty} h(n) = 0$



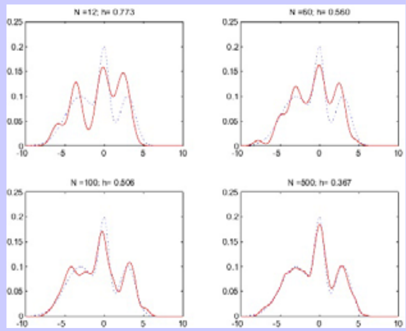
Optimal scale

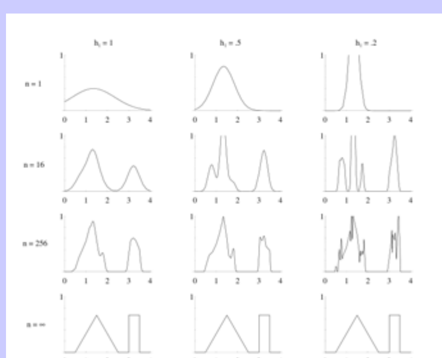
- Optimal kernel and optimal scale can be achieved by minimizing the mean integrated square error – if we know the density !
- Normal reference rule:

$$h^{opt} = (4/3)^{1/5}\sigma \cdot n^{-1/5} \approx 1.06\hat{\sigma} \cdot n^{-1/5}$$



Scale selection





From R. O. Duda, P. E. Hart, and D. G. Stork. "*Pattern Classification*" Wiley, New York, 2nd edition, 2000

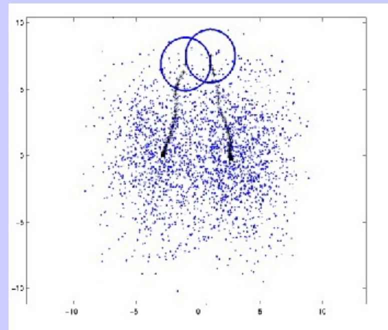
Mean Shift

- Given a sample $S=\{s_i : s_i \in R^n\}$ and a kernel K, the sample mean using K at point x:

$$m(x) = \frac{\sum_i s_i K(s_i - x)}{\sum_i K(s_i - x)}$$

- Iteration of the form $x \leftarrow m(x)$ will lead to the density local mode
- Let x is the center of the window

  Iterate until conversion.
  - Compute the sample mean $m(x)$ from the samples inside the window.
  - Replace x with $m(x)$



Mean Shift

- Given a sample $S=\{s_i : s_i \in R^n\}$ and a kernel K, the sample mean using K at point x:

$$m(x) = \frac{\sum_i s_i K(s_i - x)}{\sum_i K(s_i - x)}$$

- Fukunaga and Hostler 1975 introduced the mean shift as the difference $m(x)-x$ using a flat kernel.
- Iteration of the form $x \leftarrow m(x)$ will lead to the density mode
- Cheng 1995 generalized the definition using general kernels and weighted data

$$m(x) = \frac{\sum_i s_i K(s_i - x) w(s_i)}{\sum_i K(s_i - x) w(s_i)}$$

- Recently popularized by D. Comaniciu and P. Meer 99+
- Applications: Clustering[Cheng,Fu 85], image filtering, segmentation[Meer 99] and tracking [Meer 00].

Mean Shift

- Iterations of the form x ← m(x) are called mean shift algorithm.
- If K is a Gaussian (e.g.) and the density estimate using K is

$$\hat{P}(x) = C \sum_i K(x - s_i) w(s_i)$$

- Using Gaussian Kernel $K_\sigma(x)$, the derivative is $K'_\sigma(x) = -\frac{x}{\sigma^2} K_\sigma(x)$

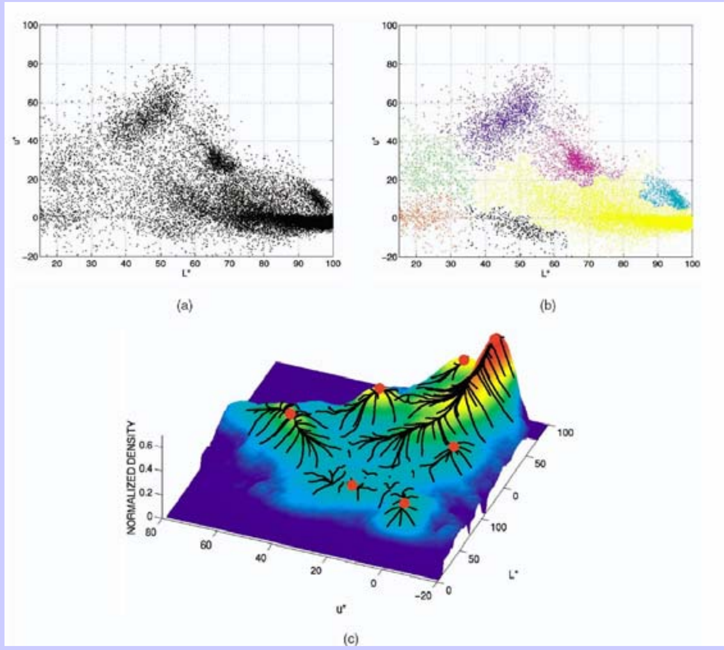
we can show that:

$$\frac{\nabla \hat{P}(x)}{\hat{P}(x)} = m(x) - x$$

- the mean shift is in the gradient direction of the density estimate.

Mean Shift

- The mean shift is in the gradient direction of the density estimate.
- Successive iterations would converge to a local maxima of the density, i.e., a stationary point: $m(x)=x$.
- Mean shift is a steepest-ascent like procedure with variable size steps that leads to fast convergence “well-adjusted steepest ascent”.



From D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis,"

Mean shift and Image Filtering

Discontinuity preserving smoothing

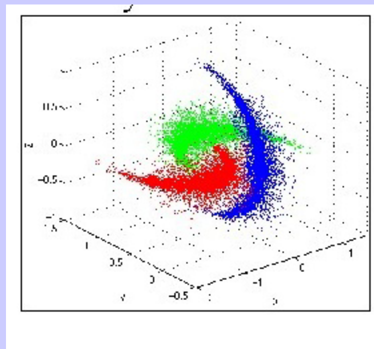
- Recall, average or Gaussian filters blur images and do not preserve region boundaries.

Mean shift application:

- Represent each pixel x as spatial location x^s and range x^r (color, intensity)
- Look for modes in the joint spatial-range space
- Use a product of two kernels: a spatial kernel with bandwidth h_s and a range kernel with bandwidth h_r

$$K_{h_s,h_r} = k_{h_s}(x^s)k_{h_r}(x^r)$$

- Algorithm:
  For each pixel $x_i=(x_i^s,x_i^r)$
  - apply mean shift until conversion. Let the conversion point be (y_i^s,y_i^r)
  - Assign $z_i = (x_i^s,y_i^r)$ as filter output
- Results: see the paper.



Mean Shift and Segmentation

- Similar to filtering but group clusters from the filtered image: group together all z_i which are closer than h_s in the spatial domain and closer than h_r in the range domain.

Sources

- R. O. Duda, P. E. Hart, and D. G. Stork. "*Pattern Classification*." Wiley, New York, 2nd edition, 2000
- Y. Cheng "Mean Shift, Mode Seeking, and Clustering" IEEE Trans. Pattern Anal. Mach. Intell. Volume 17 , Issue 8 (August 1995)
- D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 5, pp. 603--619, 2002.
- Slides by D. Comaniciu