Mean Shift Tracking

CS4243 Computer Vision and Pattern Recognition

Leow Wee Kheng

Department of Computer Science
School of Computing
National University of Singapore

Mean Shift

Mean Shift [Che98, FH75, Sil86]

- An algorithm that iteratively shifts a data point to the average of data points in its neighborhood.
- Similar to clustering.
- Useful for clustering, mode seeking, probability density estimation, tracking, etc.

- Consider a set S of n data points $\mathbf{x}_i$ in d-D Euclidean space X.
- Let $K(\mathbf{x})$ denote a kernel function that indicates how much $\mathbf{x}$ contributes to the estimation of the mean.
- Then, the sample mean $\mathbf{m}$ at $\mathbf{x}$ with kernel K is given by

$$\mathbf{m}(\mathbf{x}) = \frac{\displaystyle\sum_{i=1}^{n} K(\mathbf{x} - \mathbf{x}_i)\,\mathbf{x}_i}{\displaystyle\sum_{i=1}^{n} K(\mathbf{x} - \mathbf{x}_i)} \tag{1}$$

- The difference $\mathbf{m}(\mathbf{x}) - \mathbf{x}$ is called mean shift.
- Mean shift algorithm: iteratively move date point to its mean.
- In each iteration, $\mathbf{x} \leftarrow \mathbf{m}(\mathbf{x})$.
- The algorithm stops when $\mathbf{m}(\mathbf{x}) = \mathbf{x}$.

- The sequence $\mathbf{x}, \mathbf{m}(\mathbf{x}), \mathbf{m}(\mathbf{m}(\mathbf{x})), \ldots$ is called the trajectory of $\mathbf{x}$.
- If sample means are computed at multiple points, then at each iteration, update is done simultaneously to all these points.

Kernel

- Typically, kernel K is a function of $\|\mathbf{x}\|^2$:

$$K(\mathbf{x}) = k(\|\mathbf{x}\|^2) \tag{2}$$

- k is called the profile of K.

Properties of Profile:

1. k is nonnegative.
2. k is nonincreasing: $k(x) \geq k(y)$ if $x < y$.
3. k is piecewise continuous and

$$\int_0^\infty k(x)dx < \infty \tag{3}$$

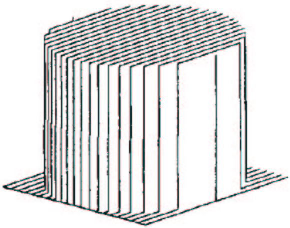
Examples of kernels [Che98]:

- Flat kernel:

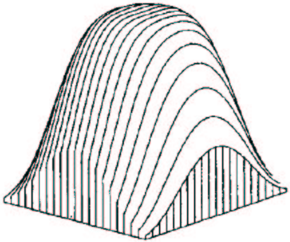
$$K(\mathbf{x}) = \begin{cases} 1 & \text{if } \|\mathbf{x}\| \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

- Gaussian kernel:

$$K(\mathbf{x}) = \exp(-\|\mathbf{x}\|^2) \tag{5}$$



(a) Flat kernel



(b) Gaussian kernel

Density Estimation

- Kernel density estimation (Parzen window technique) is a popular method for estimating probability density [CRM00, CRM02, DH73].
- For a set of n data points $\mathbf{x}_i$ in d-D space, the kernel density estimate with kernel $K(\mathbf{x})$ (profile $k(x)$) and radius h is

$$\begin{aligned} \tilde{f}_K(\mathbf{x}) &= \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right) \\ &= \frac{1}{nh^d}\sum_{i=1}^{n} k\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right) \end{aligned} \tag{6}$$

- The quality of kernel density estimator is measured by the mean square error between the actual density and the estimate.

- The mean square error is minimized by the Epanechnikov kernel:

$$K_E(\mathbf{x}) = \begin{cases} \dfrac{1}{2C_d}(d+2)(1-\|\mathbf{x}\|^2) & \text{if } \|\mathbf{x}\| \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where C_d is the volume of the unit d-D sphere, with profile

$$k_E(x) = \begin{cases} \dfrac{1}{2C_d}(d+2)(1-x) & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x > 1 \end{cases} \tag{8}$$

- A more commonly used kernel is Gaussian

$$K(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right) \tag{9}$$

with profile

$$k(x) = \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}x\right) \tag{10}$$

- Define another kernel $G(\mathbf{x}) = g(\|\mathbf{x}\|^2)$ such that

$$g(x) = -k'(x) = -\frac{dk(x)}{dx} \tag{11}$$

Important Result

Mean shift with kernel G moves $\mathbf{x}$ along the direction of the gradient of density estimate $\tilde{f}$ with kernel K.

- Define density estimate with kernel K.
- But, perform mean shift with kernel G.
- Then, mean shift performs gradient ascent on density estimate.

Proof:

- Define $\tilde{f}$ with kernel G

$$\tilde{f}_G(\mathbf{x}) \equiv \frac{C}{nh^d}\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right) \tag{12}$$

where C is a normalization constant.

- Mean shift $\mathbf{M}$ with kernel G is

$$\mathbf{M}_G(\mathbf{x}) \equiv \frac{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)\mathbf{x}_i}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x} \tag{13}$$

- Estimate of density gradient is the gradient of density estimate

$$\begin{aligned}
\tilde{\nabla} f_K(\mathbf{x}) &\equiv \nabla \tilde{f}_K(\mathbf{x}) \\
&= \frac{2}{nh^{d+2}} \sum_{i=1}^{n} (\mathbf{x} - \mathbf{x}_i)\, k' \left(\left\| \frac{\mathbf{x} - \mathbf{x}_i}{h} \right\|^2 \right) \\
&= \frac{2}{nh^{d+2}} \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{x})\, g \left(\left\| \frac{\mathbf{x} - \mathbf{x}_i}{h} \right\|^2 \right) \\
&= \frac{2}{Ch^2} \tilde{f}_G(\mathbf{x}) \mathbf{M}_G(\mathbf{x})
\end{aligned} \tag{14}$$

- Then,

$$\mathbf{M}_G(\mathbf{x}) = \frac{Ch^2}{2} \frac{\tilde{\nabla} f_K(\mathbf{x})}{\tilde{f}_G(\mathbf{x})} \tag{15}$$

- Thus, $\mathbf{M}_G$ is an estimate of the normalized gradient of f_K.
- So, can use mean shift (Eq. 13) to obtain estimate of f_K.

Mean Shift Tracking

Basic Ideas [CRM00]:

- Model object using color probability density.
- Track target object in video by matching color density.
- Use mean shift to estimate color density and target location.

Object Model

- Let $\mathbf{x}_i$, $i = 1, \ldots, n$, denote pixel locations of model centered at $\mathbf{0}$.
- Represent color distribution by discrete m-bin color histogram.
- Let $b(\mathbf{x}_i)$ denote the color bin of the color at $\mathbf{x}_i$.
- Assume size of model is normalized; so, kernel radius $h = 1$.
- Then, the probability q of color u in the model is

$$q_u = C \sum_{i=1}^{n} k(\|\mathbf{x}_i\|^2)\, \delta(b(\mathbf{x}_i) - u) \tag{16}$$

- C is the normalization constant

$$C = \left[\sum_{i=1}^{n} k(\|\mathbf{x}_i\|^2) \right]^{-1} \tag{17}$$

- Kernel profile k weights contribution by distance to centroid.

- δ is the Kronecker delta function

$$\delta(a) = \begin{cases} 1 & \text{if } a = 0 \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

That is, contribute $k(\|\mathbf{x}_i\|^2)$ to q_u if $b(\mathbf{x}_i) = u$.

Target Candidate

- Let $\mathbf{y}_i$, $i = 1, \ldots, n_h$, denote pixel locations of target centered at $\mathbf{y}$.
- Then, the probability p of color u in the target is

$$p_u(\mathbf{y}) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y}-\mathbf{y}_i}{h}\right\|^2\right) \delta(b(\mathbf{y}_i) - u) \tag{19}$$

- C_h is the normalization constant

$$C_h = \left[\sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y}-\mathbf{y}_i}{h}\right\|^2\right)\right]^{-1} \tag{20}$$

Color Density Matching

- Use Bhattacharyya coefficient ρ

$$\rho(p(\mathbf{y}), q) = \sum_{u=1}^{m} \sqrt{p_u(\mathbf{y})\, q_u} \qquad (21)$$

- ρ is the cosine of vectors $(\sqrt{p_1}, \ldots, \sqrt{p_m})^\top$ and $(\sqrt{q_1}, \ldots, \sqrt{q_m})^\top$.
- Large ρ means good color match.
- For each image frame, find $\mathbf{y}$ that maximizes ρ.
- This $\mathbf{y}$ is the location of the target.

Tracking Algorithm

Given $\{q_u\}$ of model and location $\mathbf{y}$ of target in previous frame:

1. Initialize location of target in current frame as $\mathbf{y}$.
2. Compute $\{p_u(\mathbf{y})\}$ and $\rho(p(\mathbf{y}), q)$.
3. Apply mean shift: Compute new location $\mathbf{z}$ as

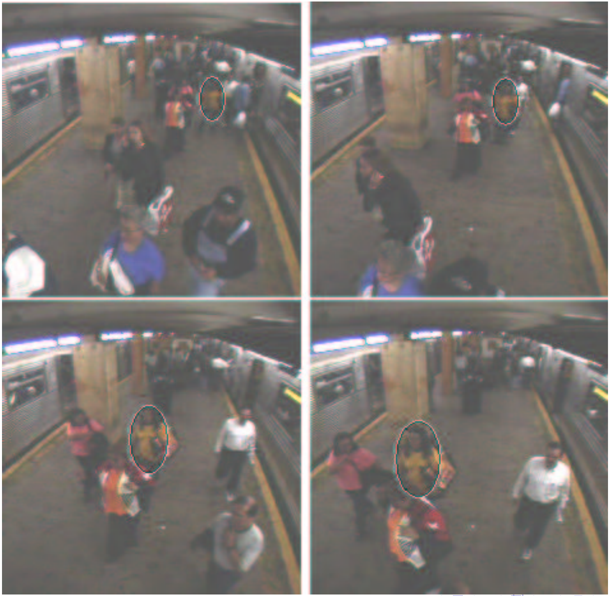
$$\mathbf{z} = \frac{\displaystyle\sum_{i=1}^{n_h} g\left(\left\|\frac{\mathbf{y}-\mathbf{y}_i}{h}\right\|^2\right)\mathbf{y}_i}{\displaystyle\sum_{i=1}^{n_h} g\left(\left\|\frac{\mathbf{y}-\mathbf{y}_i}{h}\right\|^2\right)} \tag{22}$$

4. Compute $\{p_u(\mathbf{z})\}$ and $\rho(p(\mathbf{z}), q)$.
5. While $\rho(p(\mathbf{z}), q) < \rho(p(\mathbf{y}), q)$, do $\mathbf{z} \leftarrow \frac{1}{2}(\mathbf{y}+\mathbf{z})$.
6. If $\|\mathbf{z}-\mathbf{y}\|$ is small enough, stop. Else, set $\mathbf{y} \leftarrow \mathbf{z}$ and goto (1).

- Step 3: In practice, a window of pixels $\mathbf{y}_i$ is considered.
  Size of window is related to h.
- Step 5 is used to validate the target's new location.
  Can stop Step 5 if $\mathbf{y}$ and $\mathbf{z}$ round off to the same pixel.
- Tests show that Step 5 is needed only 0.1% of the time.
- Step 6: can stop algorithm if $\mathbf{y}$ and $\mathbf{z}$ round off to the same pixel.
- To track object that changes size, varies radius h
  (see [CRM00] for details).

Example 1: Track football player no. 78 [CRM00].

Example 2: Track a passenger in train station [CRM00].



Reference I

- Y. Cheng.
  Mean shift, mode seeking, and clustering.
  IEEE Trans. on Pattern Analysis and Machine Intelligence, 17(8):790–799, 1998.

- D. Comaniciu, V. Ramesh, and P. Meer.
  Real-time tracking of non-rigid objects using mean shift.
  In *IEEE Proc. on Computer Vision and Pattern Recognition*, pages 673–678, 2000.

- D. Comaniciu, V. Ramesh, and P. Meer.
  Mean shift: A robust approach towards feature space analysis.
  IEEE Trans. on Pattern Analysis and Machine Intelligence, 24(5):603–619, 2002.

Reference II

- R. O. Duda and P. E. Hart.
  Pattern Classification and Scene Analysis.
  Wiley, 1973.

- K. Fukunaga and L. D. Hostetler.
  The estimation of the gradient of a density function, with applications in pattern recognition.
  IEEE Trans. on Information Theory, 21:32–40, 1975.

- B. W. Silverman.
  Density Estimation for Statistics and Data Analysis.
  Chapman and Hall, 1986.