Mean Shift Tracker

Lecturer:

Jiří Matas

Authors:

Ondřej Drbohlav and Jiří Matas

Center for Machine Perception

Czech Technical University in Prague

http://cmp.felk.cvut.cz

Lecture date: 3.4.2017

Mean Shift Tracking

slide credit:
Patrick Perez

- Global description of tracked region: histogram
- Reference histogram with B bins

$$\mathbf{q}^* = (q_u^*)_{u=1\cdots B}$$

set at track initialization

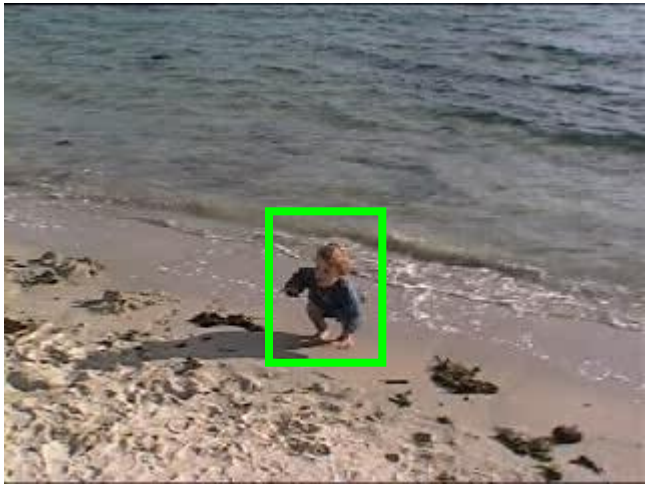
- Candidate histogram at current instant

$$\mathbf{q}(\mathbf{x}) = (q_u(\mathbf{x}))_{u=1\cdots B}$$

gathered in region $R(\mathbf{x})$ of current image.

- At each instant

$$\hat{\mathbf{x}}_{t+1} = \arg\min_{\mathbf{x}} \text{dist}(\mathbf{q}^*, \mathbf{q}(\mathbf{x}))$$



- searched around $\hat{\mathbf{x}}_t$
- iterative search initialized with $\hat{\mathbf{x}}_t$: meanshift-like iteration

Mean Shift Tracking

slide credit:
Patrick Perez

- Global description of tracked region: histogram
- Reference histogram with B bins

$$\mathbf{q}^* = (q_u^*)_{u=1\cdots B}$$

set at track initialization

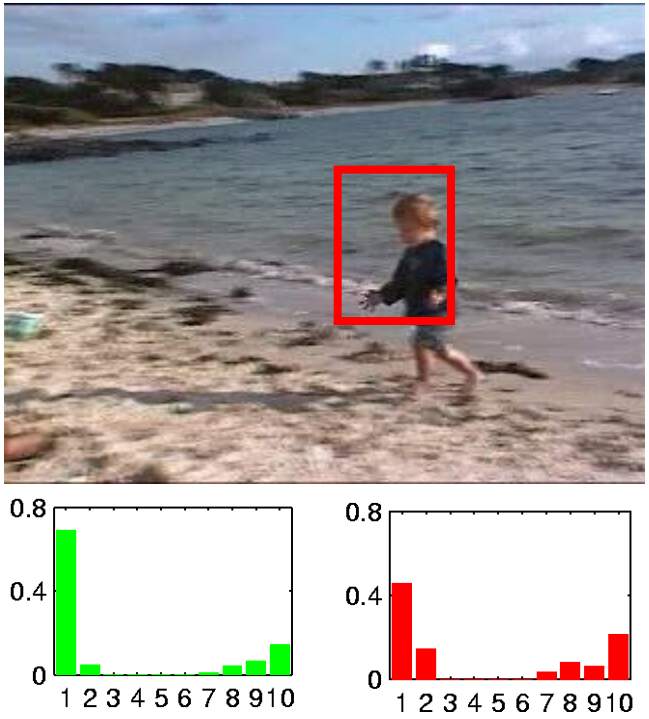
- Candidate histogram at current instant

$$\mathbf{q}(\mathbf{x}) = (q_u(\mathbf{x}))_{u=1\cdots B}$$

gathered in region $R(\mathbf{x})$ of current image.

- At each instant

$$\hat{\mathbf{x}}_{t+1} = \arg\min_{\mathbf{x}} \text{dist}(\mathbf{q}^*, \mathbf{q}(\mathbf{x}))$$



- searched around $\hat{\mathbf{x}}_t$
- iterative search initialized with $\hat{\mathbf{x}}_t$: meanshift-like iteration

Mean Shift Tracking

slide credit:
Patrick Perez

- Global description of tracked region: histogram
- Reference histogram with B bins

$$\mathbf{q}^* = (q_u^*)_{u=1\cdots B}$$

set at track initialization

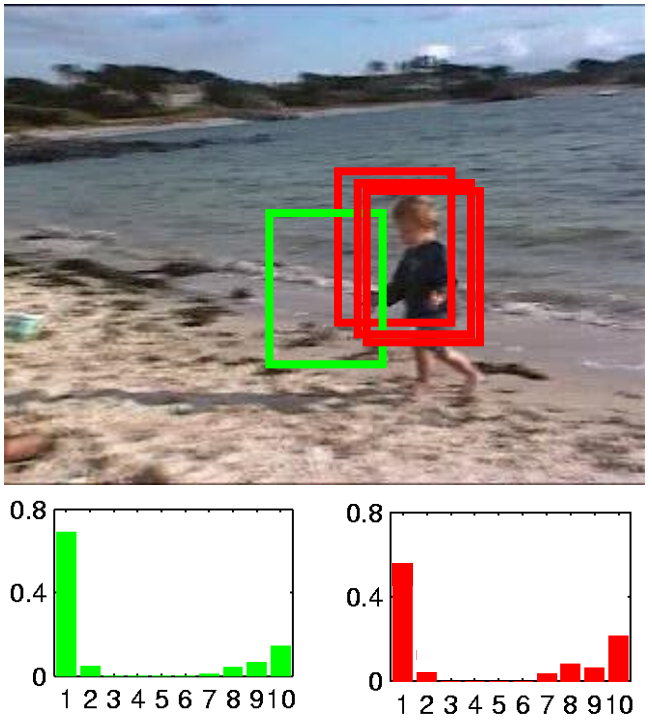
- Candidate histogram at current instant

$$\mathbf{q}(\mathbf{x}) = (q_u(\mathbf{x}))_{u=1\cdots B}$$

gathered in region $R(\mathbf{x})$ of current image.

- At each instant

$$\hat{\mathbf{x}}_{t+1} = \arg\min_{\mathbf{x}} \text{dist}(\mathbf{q}^*, \mathbf{q}(\mathbf{x}))$$



- searched around $\hat{\mathbf{x}}_t$
- iterative search initialized with $\hat{\mathbf{x}}_t$: meanshift-like iteration

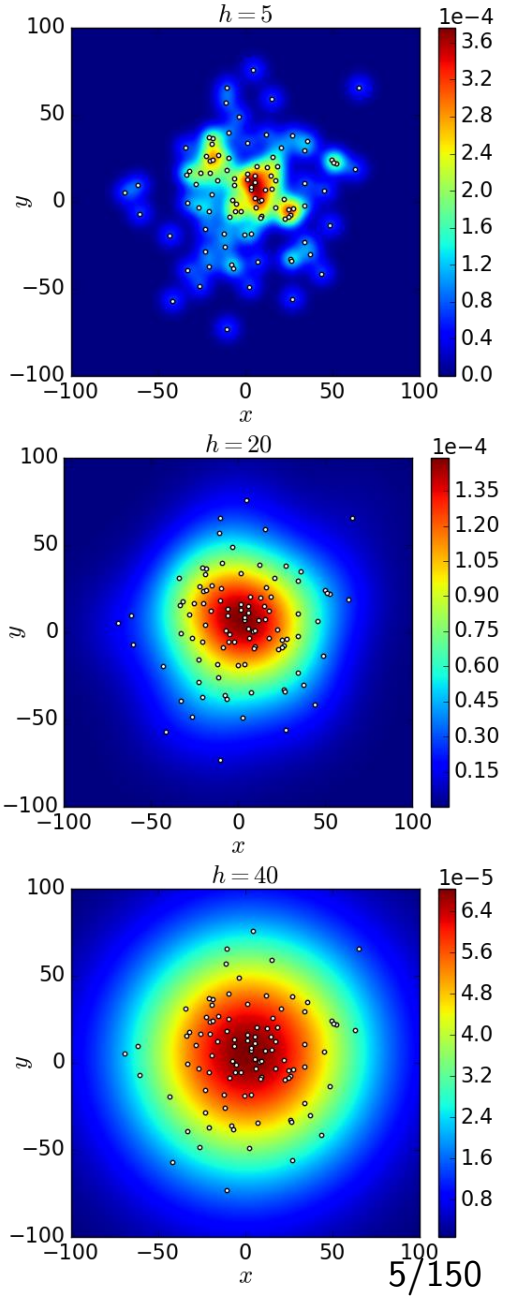
Mean Shift: Kernel Density Estimation

Given $\{\mathbf{x}_i,\, i = 1, ..., N\}$ the set of points in $\mathbb{R}^d$, the *multivariate kernel density estimator* with kernel $K(\mathbf{x})$ and bandwith h is computed as

$$p(\mathbf{x}) = \frac{1}{Nh^d} \sum_{i=1}^{N} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right) . \qquad (1)$$

Example: Normal kernel K_N:

$$K_N(\mathbf{x}) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right) \qquad (2)$$



Kernel density estimation for different bandwiths *h*.

Data points (*N*=100) are sampled from normal distribution with std=30.

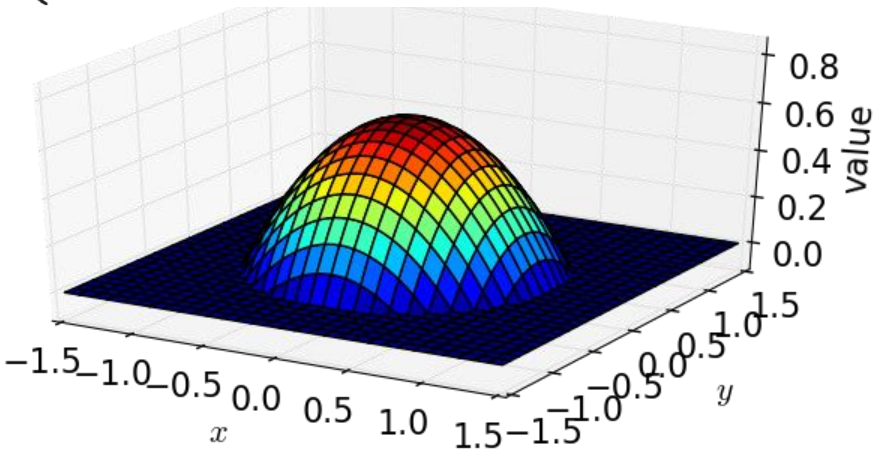
Mean Shift: Kernel Density Estimation

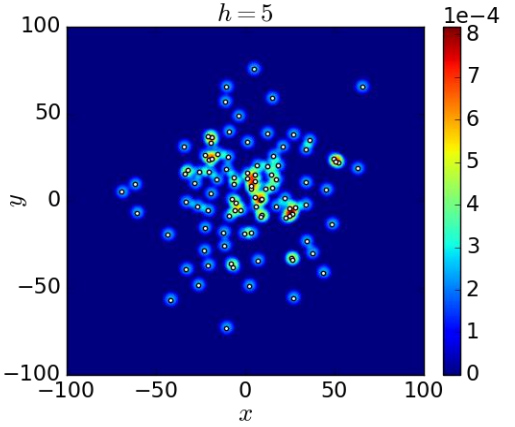
Given $\{\mathbf{x}_i, i = 1, ..., N\}$ the set of points in $\mathbb{R}^d$, the *multivariate kernel density estimator* with kernel $K(\mathbf{x})$ and bandwith h is computed as

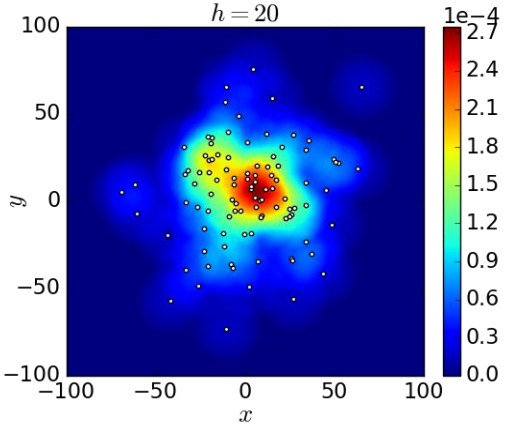
$$p(\mathbf{x}) = \frac{1}{Nh^d} \sum_{i=1}^{N} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right) . \qquad (1)$$

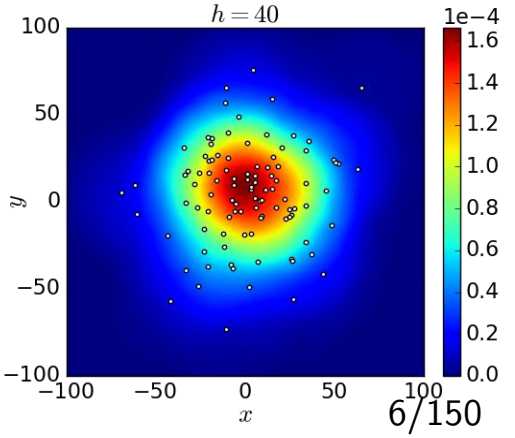
Example: *Epanechnikov* kernel K_E (c_d: volume of d-dimensional sphere):

$$K_E(\mathbf{x}) = \begin{cases} \frac{1}{2} c_d^{-1} (d+2)(1 - \|\mathbf{x}\|^2) & \text{if } \|x\| < 1 \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$









Mean Shift: Kernel Profile (isotropic kernels)

Let us rewrite the computation of KDE

$$p(\mathbf{x}) = \frac{1}{Nh^d} \sum_{i=1}^{N} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right) \tag{1}$$

using the **kernel profile** $k(z) = k(\|\mathbf{x}\|^2) = K(\mathbf{x})$:

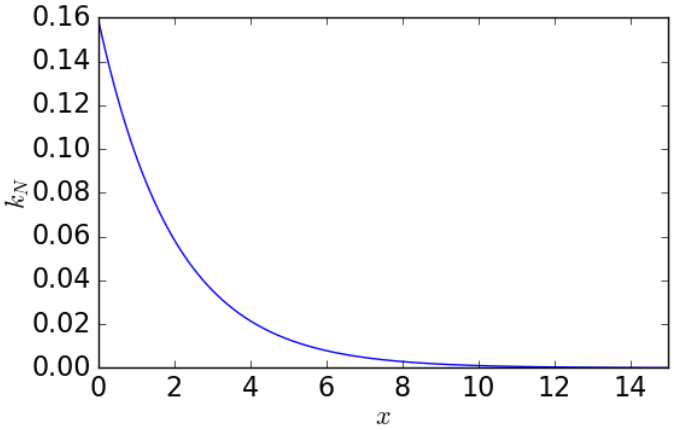
$$p(\mathbf{x}) = \frac{1}{Nh^d} \sum_{i=1}^{N} k\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right), \tag{2}$$

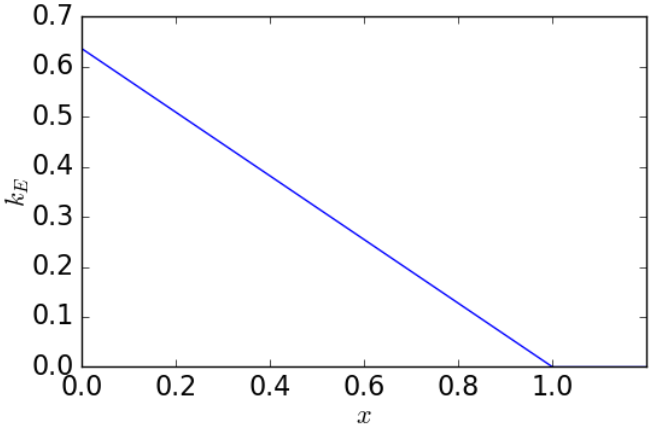
as this will simplify further derivations. The normal and Epanechnikov profiles are then:

$$k_N(x) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2}x\right), \tag{3}$$

$$k_E(x) = \begin{cases} \frac{1}{2}c_d^{-1}(d+2)(1-x) & \text{if } x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Normal and Epanechnikov kernel profiles:





Mean Shift: Mode Seeking

Task: Given $\{\mathbf{x}_i,\, i = 1, ..., N\}$ the set of points in $\mathbb{R}^d$, a kernel with profile $k(x)$ and bandwith parameter h, find the local mode of distribution $p(\mathbf{x})$ estimated using KDE, starting from initial mode estimate $\mathbf{y}_0$. This will employ grad. ascent.

$$p(\mathbf{x}) = \frac{1}{Nh^d} \sum_{i=1}^{N} k\left(\left\| \frac{\mathbf{x} - \mathbf{x}_i}{h} \right\|^2 \right), \tag{1}$$

$$\nabla p(\mathbf{x}) = \frac{2}{Nh^{d+2}} \sum_{i=1}^{N} k'\left(\left\| \frac{\mathbf{x} - \mathbf{x}_i}{h} \right\|^2 \right) (\mathbf{x} - \mathbf{x}_i) = \frac{2}{Nh^{d+2}} \times$$

$$\sum_{i=1}^{N} k'\left(\left\| \frac{\mathbf{x} - \mathbf{x}_i}{h} \right\|^2 \right) \left(\mathbf{x} - \frac{\sum_{m=1}^{N} \mathbf{x}_m k'\left(\left\| \frac{\mathbf{x} - \mathbf{x}_m}{h} \right\|^2 \right)}{\sum_{m=1}^{N} k'\left(\left\| \frac{\mathbf{x} - \mathbf{x}_m}{h} \right\|^2 \right)} \right). \tag{2}$$

For discussed kernel profiles, $k'(x)$ is negative (or zero); define $g(x) = -k'(x)$. Then the last equation is rewritten as

$$\nabla p(\mathbf{x}) = \frac{2}{Nh^{d+2}} \sum_{i-1}^{N} g\left(\left\| \frac{\mathbf{x} - \mathbf{x}_i}{h} \right\|^2 \right) \left(\frac{\sum_{m=1}^{N} \mathbf{x}_m g\left(\left\| \frac{\mathbf{x} - \mathbf{x}_m}{h} \right\|^2 \right)}{\sum_{m=1}^{N} g\left(\left\| \frac{\mathbf{x} - \mathbf{x}_m}{h} \right\|^2 \right)} - \mathbf{x} \right). \tag{3}$$

Mean Shift: Mode Seeking

const > 0 (kernel-weighted) sample **mean**

$$\nabla p(\mathbf{x}) = \frac{2}{Nh^{d+2}} \sum_{i=1}^{N} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right) \left(\frac{\sum_{m=1}^{N} \mathbf{x}_m g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_m}{h}\right\|^2\right)}{\sum_{m=1}^{N} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_m}{h}\right\|^2\right)} - \mathbf{x}\right) . \quad (1)$$

mean shift vector $M_{h,G}(\mathbf{x})$

Note that $g(x) = -k'(x)$ defines a kernel $G(\mathbf{x}) = Cg(\|\mathbf{x}\|^2)$
(C is a norm. constant)

- For Epanechnikov profile, $g(x)$ profile is a constant and $G(\mathbf{x})$ is thus a uniform kernel.
- For normal profile, $g(x)$ is again a normal profile and $G(\mathbf{x})$ is thus equal to $K(\mathbf{x})$.

Mean Shift: Mode Seeking

$$\nabla p(\mathbf{x}) = \underbrace{\frac{2}{Nh^{d+2}}}_{\text{const} > 0} \sum_{i=1}^{N} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right) \underbrace{\left(\frac{\sum_{m=1}^{N} \mathbf{x}_m g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_m}{h}\right\|^2\right)}{\sum_{m=1}^{N} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_m}{h}\right\|^2\right)} - \mathbf{x}\right)}_{\text{mean shift vector } M_{h,G}(\mathbf{x})} . \quad (1)$$

(kernel-weighted) sample **mean**

The **Mean Shift** procedure for mode seeking:

- Given $\mathbf{y}_1$,
- Compute

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^{N} \mathbf{x}_i g\left(\left\|\frac{\mathbf{y_j}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{N} g\left(\left\|\frac{\mathbf{y_j}-\mathbf{x}_i}{h}\right\|^2\right)}, \quad j = 1, 2, \ldots \quad (1)$$

- Proof of convergence: [CRM] Comaniciu, Ramesh and Meer: *Real-Time Tracking of Non-Rigid Objects using Mean Shift*. In CVPR 2000.

Mean Shift Tracking

- Let $\mathbf{x}_i$ $(i = 1, 2, ..., N)$ be pixel locations and $b(\mathbf{x}_i)$ be a function which assigns the histogram bin index to pixel location $\mathbf{x}_i$.
- Let the center of template be at $\mathbf{c}$ and its pixel set be S_q.
- The histogram description q_u of the template is computed as:

$$q_u = C_q \sum_{\substack{\mathbf{x}_i \in S_q \\ b(\mathbf{x}_i)=u}} k\left(\left\|\frac{\mathbf{c}-\mathbf{x}_i}{h_q}\right\|^2\right), \quad u = 1, 2, ..., B \tag{1}$$

  thus votes are weighted by a kernel profile k with bandwith h_q (C_q is a norm. constant. and B is the number of bins.)
- The histogram description p_u of target candidate centered at $\mathbf{y}$ and with extent comprising pixels S_p is computed in the same way:

$$p_u = C_p \sum_{\substack{\mathbf{x}_i \in S_p \\ b(\mathbf{x}_i)=u}} k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h_p}\right\|^2\right), \quad u = 1, 2, ..., B \tag{2}$$

Mean Shift Tracking

The similarity between template and candidate histograms $\mathbf{q} = (q_1, q_2, ..., q_B)$ and $\mathbf{p} = (p_1, p_2, ..., p_B)$ is measured using the Bhattacharyya coefficient:

$$\rho(\mathbf{p}(\mathbf{y}), \mathbf{q}) = \sum_{u=1}^{B} \sqrt{p_u(\mathbf{y}) q_u}\,. \tag{1}$$

Note that if $\mathbf{p} = \mathbf{q}$ then $\rho = 1$.

Let us make Taylor expansion around the current histogram values $\mathbf{p}(\mathbf{y}_0)$ computed at position $\mathbf{y}_0$, thus at $\mathbf{p}(\mathbf{y}) = \mathbf{p}(\mathbf{y}_0) + \Delta\mathbf{p}$:

$$\begin{aligned}
\rho(\mathbf{p}(\mathbf{y}_0) + \Delta\mathbf{p}, \mathbf{q}) &= \sum_{u=1}^{B} \sqrt{(p_u(\mathbf{y}_0) + \Delta p_u) q_u} \\
&\approx \sum_{u=1}^{B} \sqrt{p_u(\mathbf{y}_0) q_u} + \frac{1}{2} \sum_{u=1}^{B} \frac{1}{\sqrt{p_u(\mathbf{y}_0) q_u}} q_u \Delta p_u \qquad (2) \\
&= \frac{1}{2} \sum_{u=1}^{B} \sqrt{p_u(\mathbf{y}_0) q_u} + \frac{1}{2} \sum_{u=1}^{B} \sqrt{\frac{q_u}{p_u(\mathbf{y_0})}} p_u(\mathbf{y}) \qquad (3)
\end{aligned}$$

Mean Shift Tracking

For $\mathbf{p}(\mathbf{y}) = \mathbf{p}(\mathbf{y}_0) + \Delta\mathbf{p}$:

$$\rho(\mathbf{p}(\mathbf{y}), \mathbf{q}) = \frac{1}{2}\sum_{u=1}^{B}\sqrt{p_u(\mathbf{y}_0)q_u} + \frac{1}{2}\sum_{u=1}^{B}\sqrt{\frac{q_u}{p_u(\mathbf{y_0})}}p_u(\mathbf{y}) \tag{1}$$

Recall that

$$p_u = C_p \sum_{\substack{\mathbf{x}_i \in S_p \\ b(\mathbf{x}_i)=u}} k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h_p}\right\|^2\right), \quad u = 1, 2, ..., B\,. \tag{2}$$

Inserting (2) to (1) we get

$$\rho(\mathbf{p}(\mathbf{y}), \mathbf{q}) = \frac{1}{2}\sum_{u=1}^{B}\sqrt{p_u(\mathbf{y}_0)q_u} + \frac{C_p}{2}\sum_{u=1}^{B}\sum_{\substack{\mathbf{x}_i \in S_p \\ b(\mathbf{x}_i)=u}}\sqrt{\frac{q_u}{p_u(\mathbf{y_0})}}k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h_p}\right\|^2\right)$$

$$= \underbrace{\frac{1}{2}\sum_{u=1}^{B}\sqrt{p_u(\mathbf{y}_0)q_u}}_{\text{const}} + \frac{C_p}{2}\sum_{\mathbf{x}_i \in S_p} w_i k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h_p}\right\|^2\right), \quad w_i = \sqrt{\frac{q_{b(\mathbf{x}_i)}}{p_{b(\mathbf{x}_i)}(\mathbf{y_0})}} \tag{3}$$

KDE estimation!

In order to maximize $\rho(\mathbf{p}(\mathbf{y}), \mathbf{q})$

$$\rho(\mathbf{p}(\mathbf{y}), \mathbf{q}) = \text{const}(\mathbf{y}_0) + \frac{C_p}{2} \sum_{\mathbf{x}_i \in S_p} w_i k\left(\left\| \frac{\mathbf{y} - \mathbf{x}_i}{h_p} \right\|^2 \right), \tag{1}$$

$$\text{where } w_i = \sqrt{\frac{q_{b(\mathbf{x}_i)}}{p_{b(\mathbf{x}_i)}(\mathbf{y_0})}}, \tag{2}$$

we need to find the mode of prob. density estimated using KDE with kernel profile k and bandwith h_p, where the individual pixels of target candidate are weighted by weights w_i. This can be done using the Mean Shift procedure, with kernel profile $g(x) = -k'(x)$.

Mean Shift Tracking

This leads to the Mean Shift tracking algorithm:

1. Given initial estimate for candidate position $\mathbf{y}_0$, compute $\rho(\mathbf{p}(\mathbf{y}_0), \mathbf{q})$.
2. Compute weights w_i according to (2) on the previous slide.
3. Update the new location $\mathbf{y}_1$ of target candidate:

$$\mathbf{y}_1 = \frac{\sum_{\mathbf{x}_i \in S_p} \mathbf{x}_i w_i g\left(\left\| \frac{\mathbf{y}_0 - \mathbf{x}_i}{h_p} \right\|^2\right)}{\sum_{\mathbf{x}_i \in S_p} w_i g\left(\left\| \frac{\mathbf{y}_0 - \mathbf{x}_i}{h_p} \right\|^2\right)} \tag{1}$$

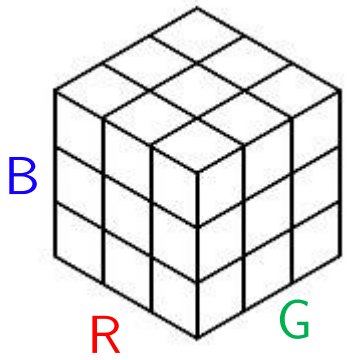
   Compute $\mathbf{p}(\mathbf{y}_1)$ and $\rho(\mathbf{p}(\mathbf{y}_1), \mathbf{q})$

4. While $\rho(\mathbf{p}(\mathbf{y}_1), \mathbf{q}) < \rho(\mathbf{p}(\mathbf{y}_0), \mathbf{q})$ do: $\mathbf{y}_1 \leftarrow \frac{1}{2}(\mathbf{y}_0 + \mathbf{y}_1)$
5. If $\|\mathbf{y_1} - \mathbf{y_0}\|$, stop. Otherwise, $\mathbf{y}_0 \leftarrow \mathbf{y}_1$ and goto step 1.

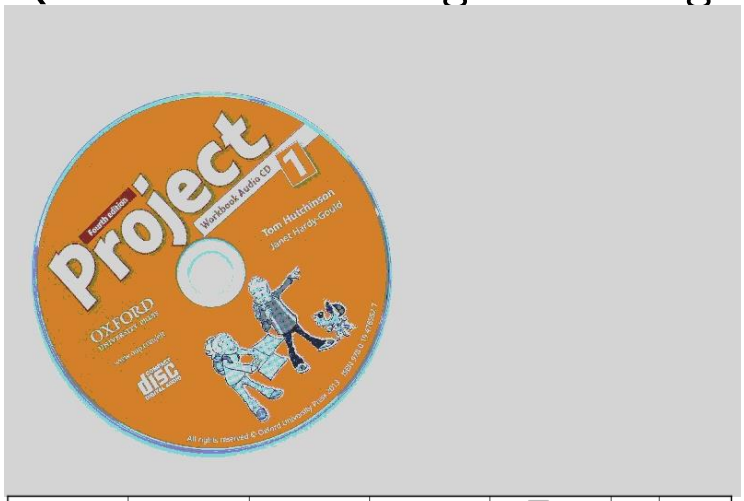
Mean Shift: Example

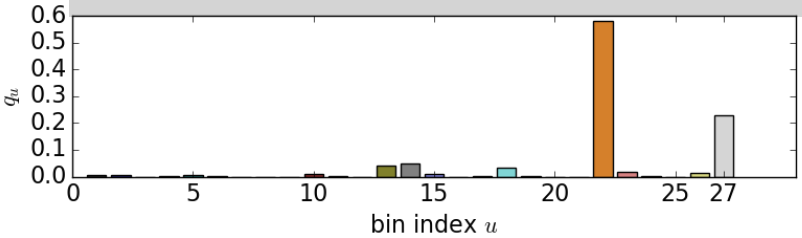
Template image patch:

Description of template:
Color histogram, 3x3x3 (27 bins)



Quantized according to binning:

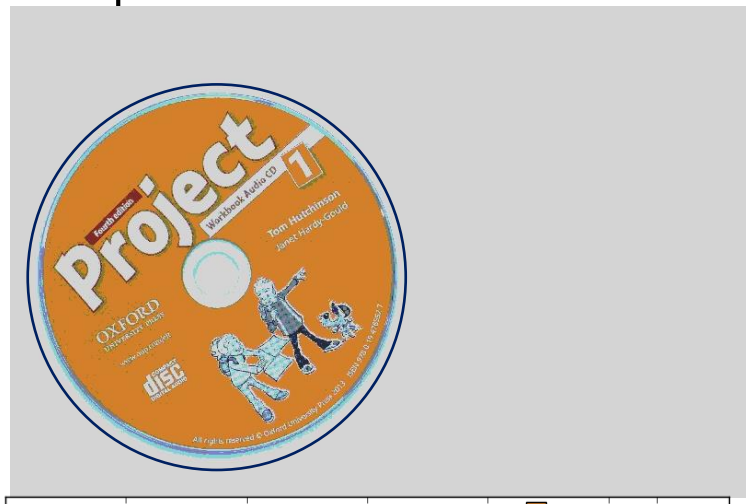


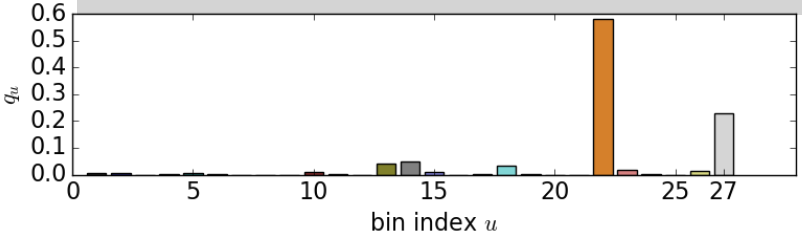


Histogram computed using
Epanechnikov kernel

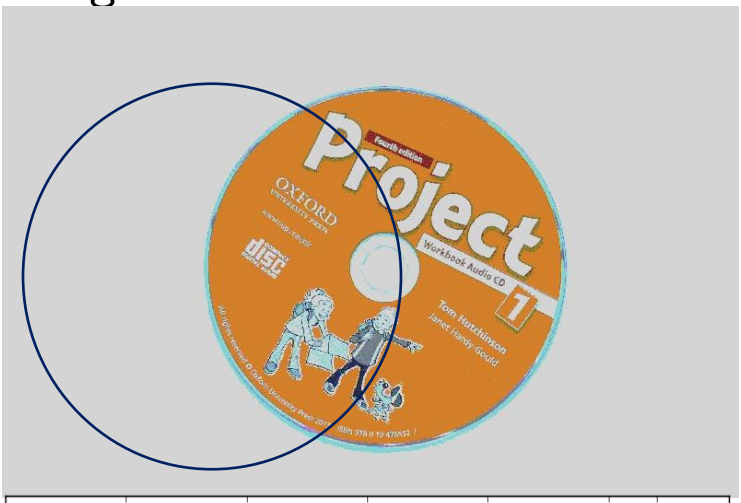
Mean Shift: Example

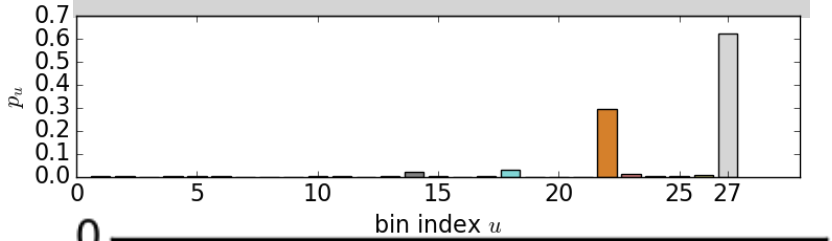
Template



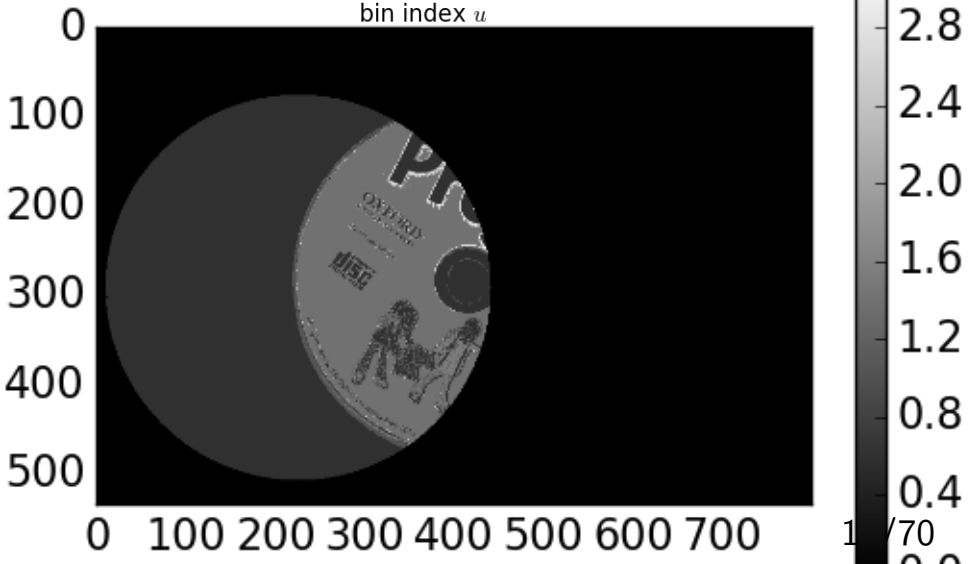


Target candidate





Weights of individual pixels computed based on their histogram bin index



Mean Shift tracking example

Feature space: 16×16×16 quantized RGB

Target: manually selected on 1st frame

Average mean-shift iterations: 4

Mean Shift tracking example

D. Comaniciu, V. Ramesh, P. Meer: *Kernel-Based Object Tracking* TPAMI, 2003

Pros and cons

- Low computational cost (easily real-time)
- Surprisingly robust
  - Invariant to pose and viewpoint
  - Often no need to update reference color model

- Invariance comes at a price
  - Position estimate prone to fluctuation
  - Scale and orientation not well captured
  - Sensitive to color clutter (e.g., teamates in team sports)
- Local search by gradient descent

- Problems:
  - abrupt moves
  - occlusions

slide credit:
Patrick Perez

Mean Shift

This leads to the Mean Shift algorithm:

1. Given intial estimate for candidate position $\mathbf{y}_0$, compute $\rho(\mathbf{p}(\mathbf{y}), \mathbf{q})$.
2. Compute weights w_i according to (2) on previous slide.
3. Update the new location $\mathbf{y}_1$ of target candidate:

$$\mathbf{y}_1 = \frac{\sum_{\mathbf{x}_i \in S_p} \mathbf{x}_i w_i g\left(\left\| \frac{\mathbf{y}_0 - \mathbf{x}_i}{h_p} \right\|\right)}{\sum_{\mathbf{x}_i \in S_p} w_i g\left(\left\| \frac{\mathbf{y}_0 - \mathbf{x}_i}{h_p} \right\|\right)}$$