Mean-Shift Tracking

Ahmet Çakıroğlu

Mean-shift is a kernel based object tracking algorithm. Mean-shift uses weighted histogram of the target region as the kernel. Object is localized by histogram similarities between the kernel and the object region in the new frame. Instead of searching the most similar histogram exhaustively, kernel window moves from the previous location to the new location of the object by a mean-shift procedure. Mean-shift employs Bhattacharyya coefficient to define a distance metric between two histograms. Bhattacharyya coefficient has the following form in terms of densities:

$$\rho(\boldsymbol{y}) \equiv \rho[p(\boldsymbol{y}), q] = \int \sqrt{p_{\boldsymbol{z}}(\boldsymbol{y})q_{\boldsymbol{z}}} d\boldsymbol{z} \tag{1}$$

Since Mean-shift uses weighted histograms as density estimates, Bhattacharyya coefficient takes the following form:

$$\hat{\rho}(\boldsymbol{y}) \equiv \rho[\hat{\mathbf{p}}(\boldsymbol{y}), \hat{\mathbf{q}}] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\boldsymbol{y})\hat{q}_u} \tag{2}$$

where $\hat{q}_u$ is the target histogram, $\hat{p}_u(\boldsymbol{y})$ is the histogram of the region centered at discrete pixel location $\mathbf{y}$ and m is the number of histogram bins. Based on Bhattacharyya coefficient, a distance metric between two histograms is defined:

$$d(\boldsymbol{y}) = \sqrt{1 - \rho[\hat{\mathbf{p}}(\boldsymbol{y}), \hat{\mathbf{q}}]} \tag{3}$$

Weighted histogram computation

Pixel locations of the region centered around $\boldsymbol{y}$ are denoted by $\{\boldsymbol{x}_i\}_{i=1...n}$. Let $b : B^2 \to \{1...m\}$ be a function which maps $\boldsymbol{x}_i$ to the index $b(\boldsymbol{x}_i)$ of the histogram bin corresponding to intensity value at the pixel location $\boldsymbol{x}_i$. The weighted histogram can be written as

$$\hat{q}_u = C \sum_{i=1}^{n} k\left(\left\|\frac{\boldsymbol{y} - \boldsymbol{x}_i}{h}\right\|^2\right) \delta[b(\boldsymbol{x}_i) - u] \tag{4}$$

where C is a normalization constant, h is the scale of the region, δ is the Kronecker delta function, $k : [0, \infty) \to R$ is a function which assigns smaller weights to the pixel locations farther away from the center.

Distance Minimization

Search procedure to find the new location y of the target starts from the location $\boldsymbol{y}_0$ of the target in the previous frame. Minimization of the distance metric is equivalent to maximization of Bhattacharyya coefficient. Taylor expansion of $\rho[p(\boldsymbol{y}), q]$ yields:

$$\rho[p(\boldsymbol{y}), q] \approx \frac{1}{2} \sum_{u=1}^{m} \sqrt{\hat{p}_u(\boldsymbol{y}_0)\hat{q}_u} + \frac{1}{2} \sum_{u=1}^{m} \hat{p}_u(\boldsymbol{y}) \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\boldsymbol{y}_0)}} \tag{5}$$

Since first term is independent of y, second term has to be maximized. Introducing (4) to second term, we get

$$\rho[p(\boldsymbol{y}), q] \approx \frac{1}{2} \sum_{u=1}^{m} \sqrt{\hat{p}_u(\boldsymbol{y}_0)\hat{q}_u} + \frac{C}{2} \sum_{u=1}^{n_h} w_i k\left(\left\|\frac{\boldsymbol{y} - \boldsymbol{x}_i}{h}\right\|^2\right) \tag{6}$$

where

$$w_i = \sum_{i=1}^{m} \delta[b(\boldsymbol{x}_i) - u]\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\boldsymbol{y}_0)}} \tag{7}$$

Maximization is achieved by Mean-shift procedure. Given the distribution $\hat{q}_u$ and the previous location y_0 of the target:

1. Compute the weights $\{w_i\}_{i=1\ldots n_h}$ according to (7).

2. Calculate the new location of the target

$$\boldsymbol{y}_1 = \frac{\sum_{i=1}^{n_h} \boldsymbol{x}_i w_i g\left(\left\|\frac{\boldsymbol{y}_0 - \boldsymbol{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} w_i g\left(\left\|\frac{\boldsymbol{y}_0 - \boldsymbol{x}_i}{h}\right\|^2\right)} \tag{8}$$

3. If $\|\boldsymbol{y}_1 - \boldsymbol{y}_0\| < \epsilon$ Stop.
    Otherwise   Set $\boldsymbol{y}_0 \leftarrow \boldsymbol{y}_1$ and go to Step 1.