SAPIENZA
UNIVERSITÀ DI ROMA

Facoltà di Ingegneria

Laurea Magistrale in **INGEGNERIA INFORMATICA**

Corso:

VISIONE E PERCEZIONE

Docente:

Prof. **FIORA PIRRI**

Tutor:

MATIA PIZZOLI

MEAN SHIFT CLUSTERING

Studente

REDJAN SHABANI

(1013173)

Definizione del mean-shift

Stima di densità multivariata con kernel:

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

con $\boldsymbol{x}, \boldsymbol{x_i} \in \boldsymbol{R^d}$. x_i, i=1…n sono le osservazioni.

K(u) e il kernel da cui dipende la stima e deve soddisfare le seguenti condizioni:

- $\sup_{\mathbf{R^d}}|K(\boldsymbol{u})| < \infty$
- $\int_{\mathbf{R^d}}|K(\boldsymbol{u})|\, d\boldsymbol{u} \ < 0$
- $\lim_{|\boldsymbol{u}|\to\infty}|\boldsymbol{u}| * |K(\boldsymbol{u})| = \boldsymbol{0}$
- $\int_{\mathbf{R^d}} K(\boldsymbol{u})\, d\boldsymbol{u} = \boldsymbol{1}$

Dove K(u) può essere:

- Kernel uniforme: $K_U(\boldsymbol{u}) = \frac{1}{c} * I(|u| \le 1)$
- Kernel di Epanechnikov: $K_E(\boldsymbol{u}) = \frac{1}{c} * (d+2) * (1 - |\boldsymbol{u}|^2) * I(|u| \le 1)$
- Kernel Gaussiano: $K_G(\boldsymbol{u}) = \frac{1}{c} * e^{-\frac{1}{2}*u^2} I(|u| \le 1)$

Stima del gradiente:

$$\hat{\nabla} f(\boldsymbol{x}) = \nabla \hat{f}(\boldsymbol{x}) = \frac{1}{nh^d}\sum_{i=1}^{n} \nabla K\left(\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right)$$

Nel caso si usasse il kernel di Epanechnikov:

$$\hat{\nabla} f(\boldsymbol{x}) = \nabla \hat{f}(\boldsymbol{x}) = \frac{1}{nh^d}\sum_{i=1}^{n} \nabla\left(\frac{1}{2c_d}(d+2)(1-|\boldsymbol{u}|^2)I(|u| \le 1)\right)$$

$$\Rightarrow \quad \hat{\nabla} f(\boldsymbol{x}) = \frac{1}{nh^d c_d} \cdot \frac{d+2}{h^2} \cdot \sum_{\boldsymbol{x}_i \in S_h^d(\boldsymbol{x})} (\boldsymbol{x}_i - \boldsymbol{x})$$

$$\Rightarrow \quad \hat{\nabla} f(\boldsymbol{x}) = \frac{n_x}{nh^d c_d} \cdot \frac{d+2}{h^2} \cdot \left(\frac{1}{n_x} \cdot \sum_{\boldsymbol{x}_i \in S_h^d(\boldsymbol{x})} (\boldsymbol{x}_i - \boldsymbol{x})\right)$$

dove $S_h(\boldsymbol{x})$ è l'ipersfera d-dimensionale di centro $\boldsymbol{x}$, raggio h, volume $h^d c_d$, contenente n_x punti(osservazioni). Il termine:

$$\boldsymbol{m}_{h,U}(\boldsymbol{x}) = \frac{1}{n_x} \cdot \sum_{\boldsymbol{x}_i \in S_h^d(\boldsymbol{x})} (\boldsymbol{x}_i - \boldsymbol{x})$$

è chiamato *vettore di mean-shift.*

La quantità

$$\hat{f}_{h,U}(\boldsymbol{x}) = \frac{n_x}{nh^d c_d}$$

è la stima di densità con kernel uniforme, all'interno dell'ipersfera $S_h^d(\boldsymbol{x})$. Possiamo allora riscrivere:

$$\hat{\nabla} f(\boldsymbol{x}) = \hat{f}_{h,U}(\boldsymbol{x}) \cdot \frac{d+2}{h^2} \cdot \boldsymbol{m}_{h,U}(\boldsymbol{x})$$

Ottenendo cosi l'espressione mean-shift in funzione della $\hat{\nabla} f(\boldsymbol{x})$:

$$\boldsymbol{m}_{h,U}(\boldsymbol{x}) = \hat{\nabla} f(\boldsymbol{x}) \frac{1}{\hat{f}_{h,U}(\boldsymbol{x})} \cdot \frac{h^2}{d+2}$$

Nel caso d=2:

$$\hat{\nabla} f(\boldsymbol{x}) = \frac{n_{\boldsymbol{x}}}{nh^d c_d} \cdot \frac{d+2}{h^2} \cdot \left(\frac{1}{n_{\boldsymbol{x}}} \cdot \sum_{\boldsymbol{x}_i \in S_h^d(\boldsymbol{x})} (\boldsymbol{x}_i - \boldsymbol{x}) \right) = \left(\frac{n_{\boldsymbol{x}}}{nh^2 \pi} \right) \cdot \frac{4}{h^2} \cdot \left(\frac{1}{n_{\boldsymbol{x}}} \cdot \sum_{\boldsymbol{x}_i \in S_h^2(\boldsymbol{x})} (\boldsymbol{x}_i - \boldsymbol{x}) \right)$$

$$\boldsymbol{m}_{h,U}(\boldsymbol{x}) = \frac{1}{n_{\boldsymbol{x}}} \cdot \sum_{\boldsymbol{x}_i \in S_h^2(\boldsymbol{x})} (\boldsymbol{x}_i - \boldsymbol{x})$$

$$\hat{f}_{h,U}(\boldsymbol{x}) = \frac{n_{\boldsymbol{x}}}{nh^2 \pi}$$

Nel caso d=3:

$$\hat{\nabla} f(\boldsymbol{x}) = \frac{n_{\boldsymbol{x}}}{nh^d c_d} \cdot \frac{d+2}{h^2} \cdot \left(\frac{1}{n_{\boldsymbol{x}}} \cdot \sum_{\boldsymbol{x}_i \in S_h^d(\boldsymbol{x})} (\boldsymbol{x}_i - \boldsymbol{x}) \right) = \left(\frac{3}{4\pi} \cdot \frac{n_{\boldsymbol{x}}}{nh^3} \right) \cdot \frac{5}{h^2} \cdot \left(\frac{1}{n_{\boldsymbol{x}}} \cdot \sum_{\boldsymbol{x}_i \in S_h^3(\boldsymbol{x})} (\boldsymbol{x}_i - \boldsymbol{x}) \right)$$

$$\boldsymbol{m}_{h,U}(\boldsymbol{x}) = \frac{1}{n_{\boldsymbol{x}}} \cdot \sum_{\boldsymbol{x}_i \in S_h^3(\boldsymbol{x})} (\boldsymbol{x}_i - \boldsymbol{x})$$

$$\hat{f}_{h,U}(\boldsymbol{x}) = \frac{3}{4\pi} \cdot \frac{n_{\boldsymbol{x}}}{nh^3}$$

Nel caso d=5: $c_d = \frac{\pi^{\frac{d}{2}} \cdot h^d}{\Gamma(\frac{d}{2}+1)} = \frac{8 \cdot \pi^2}{15}$

$$\hat{\nabla} f(\boldsymbol{x}) = \frac{n_{\boldsymbol{x}}}{nh^d c_d} \cdot \frac{d+2}{h^2} \cdot \left(\frac{1}{n_{\boldsymbol{x}}} \cdot \sum_{\boldsymbol{x}_i \in S_h^d(\boldsymbol{x})} (\boldsymbol{x}_i - \boldsymbol{x}) \right) = \frac{15 n_{\boldsymbol{x}}}{8nh^5 \pi^2} \cdot \frac{7}{h^2} \cdot \left(\frac{1}{n_{\boldsymbol{x}}} \cdot \sum_{\boldsymbol{x}_i \in S_h^d(\boldsymbol{x})} (\boldsymbol{x}_i - \boldsymbol{x}) \right)$$

$$\boldsymbol{m}_{h,U}(\boldsymbol{x}) = \frac{1}{n_{\boldsymbol{x}}} \cdot \sum_{\boldsymbol{x}_i \in S_h^5(\boldsymbol{x})} (\boldsymbol{x}_i - \boldsymbol{x})$$

$$\hat{f}_{h,U}(\boldsymbol{x}) = \frac{15 n_{\boldsymbol{x}}}{8nh^5 \pi^2}$$

Il modulo del mean-shift decresce con l'avvicinarsi alla moda(massimo locale).

E' possibile definire la procedura del mean-shift:

1. Calcolare il mean-shift nel punto **x**: $\boldsymbol{m}_{h,U}(\boldsymbol{x})$
2. Spostare l'ipersfera nel nuovo punto x': $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{m}_{h,U}(\boldsymbol{x})$
3. Se $\left\| \boldsymbol{m}_{h,U}(\boldsymbol{x}') \right\|$ diventa "molto piccolo" fermare la procedura restituendo **x'** come una moda, altrimenti ripartire da 1.

Generalizzazione del mean-shift

Introduciamo il profilo di un kernel K tale che:

$$K(\boldsymbol{u}) = k(\|\boldsymbol{u}\|^2)$$

Per esempio il profilo del kernel di Epanechnikov è:

$$k_E(\boldsymbol{u}) = \frac{(d+2)}{c_d} \cdot (1-\boldsymbol{u}) \cdot I(u<1)$$

Utilizzando la nozione di profilo del kernel, riscriviamo la formula della stima di densità multivariata:

$$\hat{f}(\boldsymbol{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} k\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)$$

Assumendo che esiste la derivata di k e ponendo g=-k', la stima del gradiente della densità multivariata, diventa:

$$\hat{\nabla} f_K(\boldsymbol{x}) = \frac{1}{nh^{d+2}} \cdot \sum_{i=1}^{n} (\boldsymbol{x}-\boldsymbol{x}_i) \cdot k'\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right) = \frac{1}{nh^{d+2}} \cdot \sum_{i=1}^{n} (\boldsymbol{x}_i-\boldsymbol{x}) \cdot g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right) =$$

$$\hat{\nabla} f_K(\boldsymbol{x}) = \frac{1}{nh^{d+2}} \cdot \left[\sum_{i=1}^{n} g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)\right] \cdot \left[\frac{\sum_{i=1}^{n} \boldsymbol{x}_i \cdot g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)} - \boldsymbol{x}\right]$$

Sottointeso che si deve avere $\sum_{i=1}^{n} g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right) \neq 0$.

Possiamo estrarre ora il vettore di mean-shift:

$$\boldsymbol{m}_{h,G}(\boldsymbol{x}) = \frac{\sum_{i=1}^{n} \boldsymbol{x}_i \cdot g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)} - \boldsymbol{x} = \frac{\sum_{i=1}^{n} \boldsymbol{x}_i \cdot G\left(\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right)}{\sum_{i=1}^{n} G\left(\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right)} - \boldsymbol{x}$$

calcolato con un kernel $G(\boldsymbol{x}) = c \cdot g(\|\boldsymbol{x}\|^2)$, essendo c una constante di normalizzazione.

La stima della densità in **x** è:

$$\hat{f}_G(\boldsymbol{x}) = \frac{c}{nh^{d+2}} \cdot \left[\sum_{i=1}^{n} g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)\right] = \frac{1}{nh^{d+2}} \cdot \left[\sum_{i=1}^{n} G\left(\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right)\right]$$

calcolato con un kernel $G(\boldsymbol{x})$. Infine si ha:

$$\hat{\nabla} f_K(\boldsymbol{x}) = \hat{f}_G(\boldsymbol{x}) \cdot \frac{2}{ch^2} \cdot \boldsymbol{m}_{h,G}(\boldsymbol{x})$$

ottenendo cosi la struttura generica del vettore di mean-shift:

$$\boldsymbol{m}_{h,G}(\boldsymbol{x}) = \frac{ch^2}{2} \cdot \frac{\hat{\nabla} f_K(\boldsymbol{x})}{\hat{f}_G(\boldsymbol{x})}$$

Linnee guida sull'implementazione

Inizialmente bisognerebbe creare una correlazione fra metrica spaziale e metrica nel range domain, cioè il concetto di distanza deve essere simile sia fra pixel sia fra grigi o colori.

Riferendosi a [2], la procedura di segmentazione è articolata in due fasi:

1° fase: *Mean-Shift Filtering*

Siano $\{x_1 \dots x_n\}$ i punti d-dimensionali dell'immagine originale e $\{z_1 \dots z_n\}$ le future coppie pixel-moda. Sara denotato con s e r rispettivamente "spatial part" e "range part" di un vettore d-dimensionale. Vengono eseguiti i seguenti passi:

Per ogni *j=1 … n*

- Inizializza k=1 e $y_k = x_j$
- Elabora

$$y_{k+1} = \frac{1}{n_k} \sum_{x_i \in S_1(y_k)} x_i \quad , \qquad k \leftarrow k+1$$

fino alla convergenza($y_{k+1} < \varepsilon, y_{conv} \leftarrow y_{k+1}$).

- Assegna $z_j = (\, x_j^s \,,\ y_{conv}^r \,)$

2° fase: *Segmentation*

Siano $\{z_1 \dots z_n\}$ le coppie pixel-moda e $\{L_1 \dots L_n\}$ un insieme di label(scalari). Venegono eseguiti i seguenti passi:

- Identifica i cluster di convergenza, $\{C_1 \dots C_m\}$, raggruppando i z_j che sono vicini tra loro nello spazio congiunto.
- Per ogni *j=1 … n* assegna $L_j = \{p | z_j \in C_p\}$
- Eliminare le regioni contenenti meno di M pixel

Dovendo implementare una versione dell'algoritmo riferendosi a dati presi da delle immagini, si ha un restringimento delle dimensioni dei campioni da trattare. Nel caso d'immagini a scala di grigi, abbiamo a che fare con campioni tridimensionali e cioè punti del tipo:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ ---- \\ \sigma(x_1, x_2) \end{bmatrix} = \begin{bmatrix} x^s \\ x^r \end{bmatrix}$$

Nel caso d'immagini a colori in codifica RGB il generico campione è della forma:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ ---- \\ \rho(x_1, x_2) \\ \gamma(x_1, x_2) \\ \beta(x_1, x_2) \end{bmatrix} = \begin{bmatrix} x^s \\ x^r \end{bmatrix}$$

Nella fase di Mean Shift Filtering, la corrispondente funzione riceve in input l'immagine in forma matriciale ed esegue il calcolo delle mode per i pixel. Trovata la moda corrispondente, si sostituisce il valore del pixel con i valori della moda(grigio o rgb). Infine si restituisce la matrice riscritta che è anche il set delle coppie $z_j = (\, x_j^s \,,\ y_{conv}^r \,)$.

Il Mean Shift Filtering, implementato come indicato in teoria e abbastanza costoso. Perciò servono una serie di ottimizzazioni, a costo di incrementare ilerrore. In [6] sono discusse diverse ottimizzazioni messe in pratica, con particolare riferimento a EDISON, un software per il mean shift processing.

Il classico concetto di località spaziale, tanto diffuso negli algoritmi dei sistemi operativi, gioca ruolo fondamentale nell'ottimizzazione.

Il primo intervento sarebbe di definire una misura di località, e per ogni pixel esaminato assegnare il range della moda a esso ed anche ai pixel che nello spazio congiunto sono vicini nello. Questo metodo(*medium speedup* in EDISON) diminuirebbe il tempo di esecuzione poiché non necessariamente bisogna visitare tutti i pixel ma solo quelli non visitati oppure non assegnati per la località. La misura di località e data come frazione di h, tipicamente ~0.5*h. La località può essere applicata non solo ai pixel simili a quello iniziale, ma anche ai pixel che sono simili ai nodi del percorso che viene fatto per la ricerca della moda(*high speedup* in EDISON). La misura di località spaziale nella catena della ricerca della moda è data come frazione di h. L'high speedup factor è tipicamente minore del medium speedup factor(~0.1), questo perche è più sensibile agli errori.

La seconda fase del, cioè il mean shift segmentation, è relativamente meno costosa della prima. Sempre in [6] sono elencate una serie di metodi efficienti per il raggruppamento dei punti. In questa sede si fa uso delle funzioni morfologiche, fornite in MatLab, per l'assemblaggio dei punti in segmenti. La funzione responsabile alla segmentazione, riceve un'immagine filtrata con mean shift, e restituisce in output una matrice mxnxp, dove p è il numero dei segmenti identificati. Alla funzione viene anche passato un parametro intero che definisce il minimo numero dei pixel che deve includere un segmento per essere accettato come tale.

Si fa notare che la se l'immagine è RGB, il mean shift fltering e fatto nello spazio congiunto di cinque dimensioni, mentre la segmentazione è stata fatta in scala di grigi. A parte il poco tempo a disposizione, si può giustificare tale scelta con la proprietà che l'immagine filtrata a colori ha comunque delle bande ristrette che sono facilmente correlabili a grigi distinguibili tra loro.

L'algoritmo è abbastanza lento. Servirebbero una serie di ottimizzazioni sugli algoritmi interni e/o l'aggiunta di euristiche. Nel caso di EDISON l'algoritmo può usare come supporto il gradiente dell'immagine. Oltre a questo dovrebbero essere collaudate in modo opportuno le costanti parametriche fornite dall'esterno oppure quelle interne.

Bibliografia:

[1]. D.Comanciu P.Meer. MEAN SHIFT: A ROBUST APPROACH TOWARD FEATURE SPACE ANALYSIS. IEEE Transaction on Pattern Analysis and Machine Learning, Vol.24 No.5 2002

[2]. D.Comanciu P.Meer. MEAN SHIFT ANALYSIS AND APPLICATIONS. Department of Electrical and Computer Engineering, Rutgers University.

[3]. Torindo Nesci. MEAN SHIFT TRACKING. Tesi di Laurea "Sapienza" Universita di Roma, 2007.

[4]. Halil Îbrahim Cüce. MEAN SHIFT ANLYSIS FOR IMAGE AND VIDEO APPLICATIONS. Tesi di Laurea "Bilkent" University of Ankara (Turkey), 2005.

[5]. K.G.Derpanis. MEAN SHIFT CLUSTERING. 15.18.2005 http://www.cse.yorku.ca/~kosta/CompVis_Notes/mean_shift.pdf

[6]. J.N.Kaftan, A.A.Bell, T.Aach. **MEAN SHIFT SEGMENTATION EVALUATION OF OPTIMIZATION TECHNIQUES.** Institute of Imaging and Computer Vision RWTH Aachen University, Germany