# 均值漂移算法的收敛性*

李乡儒[1,2+], 吴福朝[1,2], 胡占义[1,2]

[1](中国科学院 自动化研究所 模式识别国家重点实验室,北京 100080)

[2](中国科学院 研究生院,北京 100039)

# Convergence of a Mean Shift Algorithm

LI Xiang-Ru[1,2+], WU Fu-Chao[1,2], HU Zhan-Yi[1,2]

[1](National Laboratory of Pattern Recognition, Institute of Automation, The Chinese Academy of Sciences, Beijing 100080, China)

[2](Graduate School, The Chinese Academy of Sciences, Beijing 100039, China)

+ Corresponding author: Phn: +86-10-62542946, Fax: +86-10-62551993, E-mail: xrli@nlpr.ia.ac.cn, http://www.ia.ac.cn

Received 2004-05-30; Accepted 2004-08-10

**Li XR, Wu FC, Hu ZY. Convergence of a mean shift algorithm. *Journal of Software*, 2005,16(3):365−374.** DOI: 10.1360/jos160365

**Abstract**: Mean shift is an effective iterative algorithm widely used in clustering, tracking, segmentation, discontinuity preserving smoothing, filtering, edge detection, and information fusion etc. However, its convergence, a key property of any iterative method, has not been rigorously proved till now. In this paper, the traditional mean shift algorithm is first extended to account for both the local property at different sampling points and the anisotropic property at different directions, then a rigorous convergence proof is provided under these extended conditions. Finally, some approaches to adaptively selecting the algorithm's parameters are outlined. The results in this paper contribute substantially to the establishment of a sound theoretical foundation for the mean shift algorithm.

**Key words**: mean shift; convergence; clustering analysis; image processing

**摘 要**: 均值漂移是一种有效的统计迭代算法,已广泛应用于聚类分析、跟踪、图像分割、图像平滑、滤波、图像边缘提取和信息融合等方面.但是,其收敛性仍没有得到严格的证明,而收敛性是任何迭代算法的必要前提.推广并严格证明了该算法的收敛性.首先将均值漂移算法做了以下推广:反映不同样本点处局部空间结构的差异及其各向异性.然后,在推广的条件下从数学上严格证明了均值漂移算法的收敛性.最后,探讨了均值漂移算法中参数的自适应选择方法.从而为该算法的应用奠定了理论基础.

**关键词**: 均值漂移;收敛性;聚类分析;图像处理

**中图法分类号**: TP301 **文献标识码**: A

* Supported by the National Natural Science Foundation of China under Grant No.60375006 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2003AA133060 (国家高技术研究发展计划(863))

**作者简介**: **李乡儒**(1972－),男,山东潍坊人,博士生,主要研究领域为天体光谱识别;**吴福朝**(1957－),男,研究员,博士生导师,主要研究领域为计算机视觉中的多视点几何学(摄像机自标定,三维重建,机器人自定位);**胡占义**(1961－),男,博士,研究员,博士生导师,主要研究领域为摄像机标定,三维重建,主动视觉,Hough 变换,视觉机器人导航,基于图像的建模和绘制.

均值漂移是一种有效的统计迭代算法,它使每一个点“漂移”到密度函数的局部极大值点.近年来,均值漂移算法已广泛应用于计算机视觉领域[1−19],例如:跟踪、图像分割、图像平滑、滤波、图像边缘提取、信息融合、运动预测、特征空间分析以及视频数据分析等方面.

假设 $X$ 是 $m$ 维欧氏空间中的总体, $\{\boldsymbol{x}_i, 1\le i\le n\}$ 是来自总体 $X$ 的独立同分布样本集,如果总体的密度估计是

$$\hat{f}(\boldsymbol{x})=\sum_{i=1}^{n} w_i k\left(\beta\left\|\boldsymbol{x}_i-\boldsymbol{x}\right\|^2\right).$$

Cheng[20]提出了如下形式的均值漂移算法:

$$\boldsymbol{y}_{j+1}=\sum_{i=1}^{n} w_i k'\left(\beta\left\|\boldsymbol{x}_i-\boldsymbol{y}_j\right\|^2\right)\boldsymbol{x}_i \Big/ \sum_{i=1}^{n} w_i k'(\beta\left\|\boldsymbol{x}_i-\boldsymbol{y}_j\right\|^2),$$

其中, $w_i>0$ 是权重, $k(x)$ 是文献[20]中定义的轮廓函数,即概率密度估计中的窗口函数, $k'(x)$ 是 $k(x)$ 的导数.Cheng[20]在下述假设条件下证明了漂移序列 $\{\boldsymbol{y}_j, j=1,2,...\}$ 的收敛性:

(1) $k(x)=e^{-x}$;

(2) 总体 $X$ 的密度曲面的理想模式具有如下形式:

$$q(\boldsymbol{x})=e^{-\gamma\|\boldsymbol{x}\|^2}, \gamma<\beta.$$

由于第 2 个条件在实际问题中很难得到满足,所以它限制了均值漂移算法的应用范围.

文献[2,4,7]试图在 $w_i=1/n$ ,并且 $k(x)$ 为单调递减的凸函数条件下证明该迭代算法的收敛性,但是在其证明的关键步骤中都存在错误.文献[2,4]在证明迭代点序列 $\{\boldsymbol{y}_j, j=1,2,...\}$ 的收敛性时,认为由“$\left\|\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right\|$ converges to zero”就能得到“$\{\boldsymbol{y}_j, j=1,2,...\}$ converge”.然而,这是一个谬误,反例如下:

例 1:令 $y_j=\sum_{i=1}^{j}\frac{1}{i}$ ,则

$$\left\|y_{j+1}-y_j\right\|=\frac{1}{j+1}\to 0(j\to\infty).$$

但是,众所周知, $\{\boldsymbol{y}_j, j=1,2,...\}$ 既不是 Cauchy 序列,也不收敛.

文献[7]在证明迭代序列 $\{\boldsymbol{y}_j, j=1,2,...\}$ 的收敛性时,主要依据是下述不等式:

$$\left\|\boldsymbol{y}_{j+m}-\boldsymbol{y}_{j+m-1}\right\|^2+...+\left\|\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right\|^2\ge\left\|\boldsymbol{y}_{j+m}-\boldsymbol{y}_j\right\|^2 \quad (1)$$

但是,这个不等式不成立(反例见例 2、例 3),所以文献[7]的证明也是错误的.

例 2:当 $m$=2 时,

$$\begin{aligned}\left\|\boldsymbol{y}_{j+2}-\boldsymbol{y}_j\right\|^2&=\left\|\boldsymbol{y}_{j+2}-\boldsymbol{y}_{j+1}+\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right\|^2\\&=\left\|\boldsymbol{y}_{j+2}-\boldsymbol{y}_{j+1}\right\|^2+\left\|\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right\|^2+2\left(\boldsymbol{y}_{j+2}-\boldsymbol{y}_{j+1}\right)^T(\boldsymbol{y}_{j+1}-\boldsymbol{y}_j).\end{aligned}$$

由文献[7]的定理 2 可知,

$$(\boldsymbol{y}_{j+2}-\boldsymbol{y}_{j+1})^T(\boldsymbol{y}_{j+1}-\boldsymbol{y}_j)\ge 0.$$

所以,

$$\left\|\boldsymbol{y}_{j+2}-\boldsymbol{y}_j\right\|^2\ge\left\|\boldsymbol{y}_{j+2}-\boldsymbol{y}_{j+1}\right\|^2+\left\|\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right\|^2.$$

但这与式(1)矛盾.

例 3:假设 $m$=2, $y_j=1$ , $y_{j+1}=2$ , $y_{j+2}=3$ ,则 $\left\|y_{j+2}-y_j\right\|^2=\left\|3-1\right\|^2=4$ ,而且,

$$\left\|y_{j+2}-y_{j+1}\right\|^2+\left\|y_{j+1}-y_j\right\|^2=1+1=2,$$

因此,

$$\left\|y_{j+2}-y_j\right\|^2\ge\left\|y_{j+2}-y_{j+1}\right\|^2+\left\|y_{j+1}-y_j\right\|^2.$$

但这与式(1)矛盾.

众所周知,迭代点的收敛性是任何迭代算法都必须具有的基本性质.而且据我们所知,均值漂移算法的基础

性工作是文献[2,4,7,20].因此,有必要进一步研究均值漂移算法的理论基础.另外,这些文献中的均值漂移理论还有下面两个局限性:

(1) 认为各个训练样本在算法中的贡献都相同.但是,位于外围的样本点常常由于受噪声、不同类型之间的混叠和离群点等因素的影响较严重而使可信度较低.所以,算法的鲁棒性不能得到保证.

(2) 没有充分考虑总体 $X$ 在不同样本 $x_i$ 周围的局部尺度信息和局部空间结构的差异性,以及各个样本周围局部空间结构的各向异性.文献[4]仅仅考虑了不同样本点之间局部尺度的差异,而在文献[2,7,20]中,既没有考虑不同样本点之间局部尺度的差异,也没有考虑每个样本点处局部空间结构的各向异性.然而,这只有在样本集 $\{\boldsymbol{x}_i, 1\le i\le n\}$ 独立同分布,并且每个样本 $x_i$ 的不同方向结构也相同时才较为准确.但是,实际问题中很少出现这种理想的情况.所以,上述文献中的结果仅仅是在极其特殊的情况下才能够快速、准确地找到局部极大值点.详细讨论见本文第 4.2 节.

根据计算机视觉方面的应用需要,本文推广了均值漂移算法,使其克服了上述局限性,主要工作是系统地研究了以下几个方面的问题:(1) 在均值漂移算法和密度估计中引入权重参数,刻画了不同样本点的贡献差异,提高了算法的鲁棒性;(2) 在均值漂移算法和密度估计中引入局部空间结构参数,它刻画了不同样本点处局部空间结构的差异和窗口尺度的各向异性;(3) 从数学上严格地证明了更一般意义下的均值漂移算法的收敛性;(4) 探讨了参数的自适应选择方法.从而为均值漂移算法的应用奠定了理论基础.

本文第 1 节讨论基于样本点的非参数密度估计.第 2 节给出均值漂移算法,并对其进行详细的分析和讨论.第 3 节给出该算法的收敛性证明.第 4 节讨论均值漂移算法中一些参数的意义和选择规则.第 5 节是结论.

## 1 非参数密度估计

本节的研究目的是从样本集 $\{\boldsymbol{x}_i, 1\le i\le n\}$ 出发给出密度函数 $f(\boldsymbol{x})$的非参数估计.非参数方法对先验知识要求最少,它完全依靠训练数据进行估计,而且可以用于任意形状密度函数的估计,因此它得到了广泛的应用.其中,基于核函数的密度估计是最常用的非参数估计方法之一.该方法是从核函数 $k(x)$(又称为窗口函数)出发,对每个样本 $x_i$,用正定矩阵 $\Sigma_i$ 表示 $\boldsymbol{x}_i$ 处的窗口形状和相对尺度.也就是说,用正定矩阵来刻画总体 $X$ 在 $\boldsymbol{x}_i$ 周围的局部空间结构.

**定义 1**. 如果函数 $k(x)$在$[0,\infty]$上非负,单调递减,$0<k(0)<+\infty$,并且 $\int_0^{+\infty} k(x) < +\infty$,则称 $k(x)$为核函数.

高斯核函数 $kg(x)=e^{-x}$, $x\ge 0$ 是一种常用的核函数.在实际应用中,为了提高运算速度,也常用各种截断的核函数.例如,截断的高斯核函数:

$$kg_c(x)=\begin{cases} e^{-x}, & \text{if} \quad 0\le x<M \\ (M+1-x)e^{-M}, & \text{if} \quad M\le x<M+1, \\ 0, & \text{if} \quad x\ge M+1 \end{cases}$$

其中 $M>0$.

注 1:由文献[21]中的定理 5.59 可知,核函数几乎处处存在有限导数,也就是说,不可导点集的 Lebesgue 测度为 0.因此,在测度为 0 的不可导点集上补充定义后,可以假设 $k'(x)$在核函数 $k(x)$的定义域上都有定义.由于核函数 $k(x)$非增,所以可以假设 $k'(x)\le 0$.

样本 $\boldsymbol{x}_i$ 对整体贡献 $f(\boldsymbol{x}|\boldsymbol{x}_i)$的核估计是

$$\hat{f}(\boldsymbol{x}|\boldsymbol{x}_i)=K_i(\boldsymbol{x})=c_{k,i,h}k\left((\boldsymbol{x}-\boldsymbol{x}_i)^T\Sigma_i^{-1}(\boldsymbol{x}-\boldsymbol{x}_i)/h^2\right) \tag{2}$$

其中,$h>0$,它从整体上调节窗口大小;常数

$$c_{k,i,h}=\frac{1}{\int k\left((\boldsymbol{x}-\boldsymbol{x}_i)^T\Sigma_i^{-1}(\boldsymbol{x}-\boldsymbol{x}_i)/h^2\right)\mathrm{d}\boldsymbol{x}} \tag{3}$$

主要目的是确保 $K_i(\boldsymbol{x})$为一个密度函数.若用 $w_i$ 表示 $\boldsymbol{x}_i$ 的权重,即 $\boldsymbol{x}_i$ 发生的先验概率,则总体 $X$ 的密度估计是[22,23]:

$$\hat{f}(\boldsymbol{x})=\sum_{i=1}^{n} w_i K_i(\boldsymbol{x}) \tag{4}$$

显然,$w_i$必须满足如下条件:

$$\sum_{i=1}^{n} w_i = 1,\ w_i > 0\ . \tag{5}$$

本文第 4 节将详细讨论参数$\Sigma_i$和 $w_i$的意义及其自适应选取方法.为了后文表述方便,定义如下两个简写符号:

$$H_i = \Sigma_i^{-1}/h^2,\ \|\boldsymbol{x}-\boldsymbol{x}_i\|_{H_i}^2 = (\boldsymbol{x}-\boldsymbol{x}_i)^{\mathrm{T}} H_i (\boldsymbol{x}-\boldsymbol{x}_i) \tag{6}$$

于是,从式(2,4,6)可以得到:

$$\hat{f}(\boldsymbol{x}) = \sum_{i=1}^{n} w_i K_i(\boldsymbol{x}) = \sum_{i=1}^{n} w_i c_{k,i,h} k\left(\|\boldsymbol{x}-\boldsymbol{x}_i\|_{H_i}^2\right) \tag{7}$$

## 2 均值漂移算法

均值漂移是一种统计迭代算法,本节首先从密度函数估计出发通过直观推导得到它的迭代公式,然后给出均值漂移算法.

### 2.1 从密度函数到均值漂移

均值漂移作为一种非常直观的统计迭代算法,它使每个待处理的点“漂移”到分布密度函数$\hat{f}(\boldsymbol{x})$的局部极大值点处.众所周知,梯度方向是函数的增长方向,所以我们将从密度估计函数$\hat{f}(\boldsymbol{x})$的梯度出发进行推导.

当核函数$k(x)$在$\left\{\|\boldsymbol{x}-\boldsymbol{x}_i\|_{H_i}^2, 1\le i\le n\right\}$处可微时,从式(7)可以得到$\hat{f}(\boldsymbol{x})$的梯度为

$$\begin{aligned}\nabla \hat{f}(\boldsymbol{x}) &= \sum_{i=1}^{n} w_i \nabla K_i(\boldsymbol{x}) \\ &= 2\sum_{i=1}^{n} w_i c_{k,i,h} k'\left(\|\boldsymbol{x}-\boldsymbol{x}_i\|_{H_i}^2\right) H_i(\boldsymbol{x}-\boldsymbol{x}_i)\ .\end{aligned}$$

令

$$g(\boldsymbol{x}) = -k'(\boldsymbol{x}) \tag{8}$$

$$L_i(\boldsymbol{x}) = 2 w_i c_{k,i,h} g\left(\|\boldsymbol{x}-\boldsymbol{x}_i\|_{H_i}^2\right) H_i \tag{9}$$

则

$$\begin{aligned}\nabla \hat{f}(\boldsymbol{x}) &= \sum_{i=1}^{n} (\boldsymbol{x}_i-\boldsymbol{x}) L_i(\boldsymbol{x}) \\ &= \sum_{i=1}^{n} L_i(\boldsymbol{x})\boldsymbol{x}_i - \sum_{i=1}^{n} L_i(\boldsymbol{x})\boldsymbol{x}\ .\end{aligned}$$

由定义 1、式(8)和注 1 可知$g\left(\|\boldsymbol{x}-\boldsymbol{x}_i\|_{H_i}^2\right)\ge 0, i\le n$.所以,由式(2,5,8,9)可知,当$\sum_{i=1}^{n} g\left(\|\boldsymbol{x}-\boldsymbol{x}_i\|_{H_i}^2\right)\ne 0$时,$\sum_{i=1}^{n} L_i(\boldsymbol{x})$是正定矩阵.这时,

$$\nabla \hat{f}(\boldsymbol{x}) = \left(\sum_{i=1}^{n} L_i(\boldsymbol{x})\right)\left[\left(\sum_{i=1}^{n} L_i(\boldsymbol{x})\right)^{-1}\sum_{i=1}^{n} L_i(\boldsymbol{x})\boldsymbol{x}_i - \boldsymbol{x}\right] \tag{10}$$

上式的第 2 项就是我们想要的均值漂移向量:

$$ms_k(\boldsymbol{x}) \overset{\Delta}{=} \left(\sum_{i=1}^{n} L_i(\boldsymbol{x})\right)^{-1}\sum_{i=1}^{n} L_i(\boldsymbol{x})\boldsymbol{x}_i - \boldsymbol{x} \tag{11}$$

从式(11)可以得到均值漂移的迭代公式:

$$\boldsymbol{y}_{j+1} = \boldsymbol{y}_j + ms_k(\boldsymbol{y}_j) = \left(\sum_{i=1}^{n} L_i(\boldsymbol{y}_j)\right)^{-1}\sum_{i=1}^{n} L_i(\boldsymbol{y}_j)\boldsymbol{x}_i\ ,$$

其中,$\boldsymbol{y}_1$是迭代的出发点,$j>0$.

### 2.2 均值漂移算法

根据第 2.1 节的讨论,并且考虑到注 1,均值漂移算法可以表示为

$$\text{如果}\sum_{i=1}^{n}L_i(\boldsymbol{y}_j)\text{可逆},\ \boldsymbol{y}_{j+1}=\boldsymbol{y}_j+ms_k(\boldsymbol{y}_j)=\left(\sum_{i=1}^{n}L_i(\boldsymbol{y}_j)\right)^{-1}\sum_{i=1}^{n}L_i(\boldsymbol{y}_j)\boldsymbol{x}_i \tag{12}$$

$$\text{如果}\sum_{i=1}^{n}L_i(\boldsymbol{y}_j)\text{不可逆},\ \boldsymbol{y}_{j+1}=\boldsymbol{y}_j \tag{13}$$

其中 $j>0$ , $j\in N$ .

注 2:由注 1、式(3,5,8,9)可知,只有当 $\nabla\hat{f}(\boldsymbol{y}_j)=0$ 时才会出现 $\boldsymbol{y}_{j+1}=\boldsymbol{y}_j$ 的情况.

注 3:在文献[2,7,20]中,仅仅考虑了 $\Sigma_i=I_{m\times m}$ 的情况,而在文献[4]中只考虑了 $\Sigma_i=h_i^2I_{m\times m}$ 的情况.本文考虑了空间结构的各向异性以及不同样本点周围局部结构的差别,对现有的均值漂移算法结果做了极大的推广.详细讨论见本文第 4.2 节.

值得指出的是,在文献[2,4,7]中,关于迭代点 $\{\boldsymbol{y}_j,j=1,2,...\}$ 的收敛性证明存在错误(详细解释见本文开始部分).收敛性是任何迭代算法的必要前提,但据我们所知,目前还没有相关文献给出收敛性的严格证明.

注 4:从直观上看,式(12)中 $\boldsymbol{y}_{j+1}$ 非常像是 $\boldsymbol{x}_i$ 的加权平均值.特别是当

$$\Sigma_i=\frac{h_i^2}{h^2}I\ ,\ h_i\neq 0 \tag{14}$$

时,式(9)成为

$$L_i(\boldsymbol{x})=2w_ic_{k,i,h}g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)I\ ,$$

从而式(12)简化为

$$\begin{aligned}\boldsymbol{y}_{j+1}&=\left(\sum_{i=1}^{n}L_i(\boldsymbol{y}_j)\right)^{-1}\sum_{i=1}^{n}L_i(\boldsymbol{y}_j)\boldsymbol{x}_i\\&=\sum_{i=1}^{n}w_ic_{k,i,h}g\left(\left\|\frac{\boldsymbol{y}_j-\boldsymbol{x}_i}{h}\right\|^2\right)\boldsymbol{x}_i\Bigg/\sum_{i=1}^{n}w_ic_{k,i,h}g\left(\left\|\frac{\boldsymbol{y}_j-\boldsymbol{x}_i}{h}\right\|^2\right),\end{aligned}$$

其中, $j>0$ , $L_i$ 是正实数, $\boldsymbol{y}_{j+1}$ 恰好是 $\boldsymbol{x}_i$ 的加权平均值,这是均值漂移称谓的来源.

注 5:当 $k(x)$ 在点集 $\left\{\|\boldsymbol{x}-\boldsymbol{x}_i\|_{H_i}^2,1\le i\le n\right\}$ 上可导并且 $\sum_{i=1}^{n}g\left(\|\boldsymbol{x}-\boldsymbol{x}_i\|_{H_i}^2\right)=0$ 时,由式(5,9)和定义 1 可知, $\sum_{i=1}^{n}L_i(\boldsymbol{y}_j)$ 不可逆,并且 $\sum_{i=1}^{n}L_i(\boldsymbol{y}_j)=0$ ;从而 $\hat{f}'(\boldsymbol{y}_j)=0$ ,也就是说,这时 $\boldsymbol{y}_j$ 是一个稳定点.

如果迭代点在漂移时满足 $\boldsymbol{y}_{j+1}\neq\boldsymbol{y}_j$ ,则从式(10,12)可知,均值漂移的方向就是梯度方向.因此,漂移序列向着函数值升高的方向移动.本文第 3.1 节将严格证明 $\left\{\hat{f}(\boldsymbol{y}_j),j=1,2,...\right\}$ 确实是递增的.

## 3 收敛性

### 3.1 漂移点列的密度值严格递增

从前面的讨论可知,均值漂移序列向着函数值升高的方向移动,而且移动的步长是自适应的.但是,在迭代算法中,如果移动的步长太大,即使是沿着函数值升高的方向移动,也可能使函数值减小.下面将证明,均值漂移的自适应步长非常好,恰好使函数值严格递增.

**定义 2**. 对于函数 $k:[0,+\infty)\to R$ ,如果存在有界、连续的 $k'(x)$ ,并且满足:

$$k(x_2)-k(x_1)>k'(x_1)(x_2-x_1)\ ,\ \forall\ x_1\ge 0,x_2\ge 0 \tag{15}$$

则称函数 $k$ 是凸的.

注 6:第 1 节中的核函数 $kg(x)$ 和 $kg_c(x)$ 都是凸的.

**定理 1**. 如果核函数是凸的,则序列 $\left\{\hat{f}\left(\boldsymbol{y}_j\right), j=1,2,\ldots\right\}$收敛并且严格单调递增.

证明:因为从核函数的定义知道 $k(x)$有界,所以由式(2,4,5)可知,密度估计函数 $\hat{f}$ 有界,从而序列 $\left\{\hat{f}\left(\boldsymbol{y}_j\right), j=1,2,\ldots\right\}$也有界.因此,只要证明 $\left\{\hat{f}\left(\boldsymbol{y}_j\right), j=1,2,\ldots\right\}$单调递增,就已说明 $\left\{\hat{f}\left(\boldsymbol{y}_j\right), j=1,2,\ldots\right\}$的收敛性.对于任意 $j \geq 1$,

(1) 如果 $\boldsymbol{y}_j=\boldsymbol{y}_{j+1}$,显然 $\hat{f}(\boldsymbol{y}_j) \leq \hat{f}(\boldsymbol{y}_{j+1})$.

(2) 如果 $\boldsymbol{y}_j \neq \boldsymbol{y}_{j+1}$,则从迭代公式的定义式(12,13)易知 $\sum_{i=1}^{n} L_i\left(\boldsymbol{y}_j\right)$可逆,并且

$$\begin{aligned}
\hat{f}(\boldsymbol{y}_{j+1})-\hat{f}(\boldsymbol{y}_j) &\overset{(7)}{=} \sum_{i=1}^{n} w_i c_{k,i,h}\left[k\left(\left\|\boldsymbol{y}_{j+1}-\boldsymbol{x}_i\right\|_{H_i}^2\right)-k\left(\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}^2\right)\right] \\
&\overset{(15)}{\geq} \sum_{i=1}^{n} w_i c_{k,i,h} k'\left(\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}^2\right)\left[\left\|\boldsymbol{y}_{j+1}-\boldsymbol{x}_i\right\|_{H_i}^2-\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}^2\right] \\
&\overset{(8)}{=} \sum_{i=1}^{n} w_i c_{k,i,h} g\left(\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}^2\right)\left[\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}^2-\left\|\boldsymbol{y}_{j+1}-\boldsymbol{x}_i\right\|_{H_i}^2\right].
\end{aligned}$$

同时,

$$\begin{aligned}
\left\|\boldsymbol{y}_{j+1}-\boldsymbol{x}_i\right\|_{H_i}^2 &= \left\|\boldsymbol{y}_{j+1}-\boldsymbol{y}_j+\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}^2 \\
&= \left\|\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right\|_{H_i}^2+2\left(\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right)^T H_i\left(\boldsymbol{y}_j-\boldsymbol{x}_i\right)+\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}^2.
\end{aligned}$$

因此,

$$\hat{f}(\boldsymbol{y}_{j+1})-\hat{f}(\boldsymbol{y}_j) \geq \sum_{i=1}^{n} w_i c_{k,i,h} g\left(\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}^2\right)\left[-\left\|\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right\|_{H_i}^2-2\left(\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right)^T H_i\left(\boldsymbol{y}_j-\boldsymbol{x}_i\right)\right].$$

又因为从式(9,12)可得

$$\begin{aligned}
&\boldsymbol{y}_{j+1} \overset{(12)}{=} \left(\sum_{i=1}^{n} L_i\left(\boldsymbol{y}_j\right)\right)^{-1} \sum_{i=1}^{n} L_i\left(\boldsymbol{y}_j\right) \boldsymbol{x}_i \\
&\Rightarrow \left(\sum_{i=1}^{n} L_i\left(\boldsymbol{y}_j\right)\right) \boldsymbol{y}_{j+1}=\sum_{i=1}^{n} L_i\left(\boldsymbol{y}_j\right) \boldsymbol{x}_i \\
&\overset{(9)}{\Rightarrow} \left(\sum_{i=1}^{n} 2 w_i c_{k,i,h} H_i g\left(\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}\right)\right) \boldsymbol{y}_{j+1}=\sum_{i=1}^{n} 2 w_i c_{k,i,h} H_i g\left(\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}\right) \boldsymbol{x}_i.
\end{aligned}$$

对上式两边同时左乘 $\left(\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right)^T / 2$ 得到

$$\left(\sum_{i=1}^{n} w_i c_{k,i,h} k'\left(\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}\right)\left(\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right)^T H_i\right) \boldsymbol{y}_{j+1}=\sum_{i=1}^{n} w_i c_{k,i,h} k'\left(\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}\right)\left(\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right)^T H_i \boldsymbol{x}_i.$$

所以

$$\begin{aligned}
\sum_{i=1}^{n} w_i c_{k,i,h} k'\left(\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}^2\right)\left(\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right)^T H_i\left(\boldsymbol{y}_j-\boldsymbol{x}_i\right) &= \sum_{i=1}^{n} w_i c_{k,i,h} k'\left(\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}^2\right)\left(\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right)^T H_i \boldsymbol{y}_j \\
&\quad -\sum_{i=1}^{n} w_i c_{k,i,h} k'\left(\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}^2\right)\left(\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right)^T H_i \boldsymbol{y}_{j+1} \\
&= -\sum_{i=1}^{n} w_i c_{k,i,h} k'\left(\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}^2\right)\left\|\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right\|_{H_i}^2.
\end{aligned}$$

因此,

$$\hat{f}\left(\boldsymbol{y}_{j+1}\right)-\hat{f}\left(\boldsymbol{y}_j\right) \geq \sum_{i=1}^{n} w_i c_{k,i,h} g\left(\left\|\boldsymbol{y}_j-\boldsymbol{x}_i\right\|_{H_i}^2\right)\left[\left\|\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\right\|_{H_i}^2\right] \tag{16}$$

因为 $\{H_i\}$ 都是正定矩阵,所以当 $\boldsymbol{y}_{j+1} \neq \boldsymbol{y}_j$ 时,从式(5,8)以及 $k(x)$和 $g(x)$的定义知道 $\hat{f}\left(\boldsymbol{y}_{j+1}\right)-\hat{f}\left(\boldsymbol{y}_j\right)>0$.即,当 $\boldsymbol{y}_j \neq \boldsymbol{y}_{j+1}$时, $\hat{f}\left(\boldsymbol{y}_{j+1}\right)>\hat{f}\left(\boldsymbol{y}_j\right)$.

综上所述,只要核函数是凸的,则对于任意的 $j\ge 1$ 有 $\hat f(\boldsymbol{y}_j)\le \hat f(\boldsymbol{y}_{j+1})$,即序列 $\{\hat f(\boldsymbol{y}_j), j=1,2,...\}$ 单调递增,所以由前面的分析可知其收敛.

由上面的证明(2)可知,对于任意的 $j\ge 1$,如果 $\boldsymbol{y}_j\ne \boldsymbol{y}_{j+1}$,则 $\hat f(\boldsymbol{y}_{j+1})>\hat f(\boldsymbol{y}_j)$,所以序列 $\{\hat f(\boldsymbol{y}_j), j=1,2,...\}$ 严格单调递增. □

### 3.2 漂移点列的收敛性

对于迭代算法来说,必须要确保迭代点收敛.由定理 1 可知,只需在下面的集合 $S_0$ 上研究迭代序列 $\{\boldsymbol{y}_j, j=1,2,...\}$ 的收敛性即可:

$$S_0=\{\boldsymbol{y}\mid \hat f(\boldsymbol{y})\ge \hat f(\boldsymbol{y}_1)\}.$$

**定理 2**. 如果核函数 $k(x)$是凸的,且 $\hat f(\boldsymbol{x})$在 $S_0$ 上存在有限个稳定点,则序列 $\{\boldsymbol{y}_j, j=1,2,...\}$ 收敛.

注 7:对于 $\hat f(\boldsymbol{x})$,因为稳定点一般代表问题域中的模式或类,由于在实际应用中它们的个数是有限的,所以该定理中的条件“$\hat f(\boldsymbol{x})$在 $S_0$ 上存在有限个稳定点”在应用中易于得到满足.

证明:对于序列 $\{\boldsymbol{y}_j, j=1,2,...\}$,如果存在某个 $j_0>0$ 满足 $\boldsymbol{y}_{j_0}=\boldsymbol{y}_{j_0+1}$,从迭代算法的定义式(12,13)可知,$\boldsymbol{y}_j=\boldsymbol{y}_{j+1}=\boldsymbol{y}_{j+2}=...$,所以此时序列 $\{\boldsymbol{y}_j, j=1,2,...\}$ 收敛;否则,对任意 $j>0$ 都有

$$\boldsymbol{y}_j\ne \boldsymbol{y}_{j+1} \tag{17}$$

下面证明,在这种情况下,$\{\boldsymbol{y}_j, j=1,2,...\}$ 也是收敛的:

由定理 1 可知,序列 $\{\hat f(\boldsymbol{y}_j), j=1,2,...\}$不仅单调递增而且收敛,所以根据式(3,5,16)、定义 1 和定义 2,

$$\|\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\|_{H_i}^2\to 0(j\to\infty),\ \forall\ i=1,2,...,n.$$

因为矩阵 $H_i$ 正定,所以从式(6)得到,

$$ms_k(\boldsymbol{y}_j)=\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\to 0(j\to\infty) \tag{18}$$

由定义 2 可知 $k'(x)$有界,所以从式(8,9)可知 $\sum_{i=1}^{n}L_i(\boldsymbol{y}_j)$有界.于是,从式(10,11,12,18)得到,

$$\nabla \hat f(\boldsymbol{y}_j)=\left(\sum_{i=1}^{n}L_i(\boldsymbol{y}_j)\right)(\boldsymbol{y}_{j+1}-\boldsymbol{y}_j)\to 0(j\to\infty) \tag{19}$$

因为 $\hat f(\boldsymbol{x})$在 $S_0$ 上存在有限个稳定点,不妨假定有 $m_0$ 个:$\{\boldsymbol{x}'_k, 1\le k\le m_0\}$,则

$$\nabla \hat f(\boldsymbol{x}'_k)=0,\ 1\le k\le m_0,$$

并且,$\nabla \hat f(\boldsymbol{x})\ne 0$,$\boldsymbol{x}\in S_0$ 且 $\boldsymbol{x}\notin\{\boldsymbol{x}_k, 1\le k\le m_0\}$.令

$$d_0\overset{\Delta}{=}\min\left\{\|\boldsymbol{x}'_j-\boldsymbol{x}'_k\|, 1\le j\ne k\le m_0\right\} \tag{20}$$

假设$\varepsilon$是满足 $0<\varepsilon<d_0/3$ 的任意实数.由定义 2 和式(8,9,10)可知,$\nabla \hat f(\boldsymbol{x})$ 在有界闭集 $V_0=S_0-\bigcup_{i=1}^{m_0}\{\boldsymbol{x}\,\|\boldsymbol{x}-\boldsymbol{x}'_i\|<\varepsilon\}$上连续且 $\nabla \hat f(\boldsymbol{x})\ne 0$,所以 $\min_{x\in V_0}\|\nabla \hat f(\boldsymbol{x})\|\ne 0$,于是存在 $c_\varepsilon>0$ 满足:

$$\|\nabla \hat f(\boldsymbol{x})\|>c_\varepsilon,\text{对任意 } \boldsymbol{x}\in V_0 \tag{21}$$

由式(18,19)可知,存在 $N_\varepsilon>0$ 同时满足下面两个不等式:

$$\|ms_k(\boldsymbol{y}_j)\|=\|\boldsymbol{y}_{j+1}-\boldsymbol{y}_j\|<\varepsilon,\ j\ge N_\varepsilon \tag{22}$$

$$\|\nabla \hat f(\boldsymbol{y}_j)\|<c_\varepsilon,\ j\ge N_\varepsilon \tag{23}$$

令 $S_{\varepsilon,i}=\left\{\boldsymbol{x}\,\|\boldsymbol{x}-\boldsymbol{x}'_i\|<\varepsilon, \boldsymbol{x}\in S_0\right\}, 1\le i\le m_0$.式(21,23)意味着,$\{\boldsymbol{y}_j, j\ge N_\varepsilon\}\subset\bigcup_{i=1}^{m_0}S_{\varepsilon,i}$.假设 $\boldsymbol{x}_1^*$ 和 $\boldsymbol{x}_2^*$ 是任意两个不同超球内的点,例如 $\boldsymbol{x}_1^*\in S_{\varepsilon,i_1}$, $\boldsymbol{x}_2^*\in S_{\varepsilon,i_2}$,则

$$\left\|x_1^*-x_2^*\right\|=\left\|x_1^*-x'_{i_1}+x'_{i_1}-x'_{i_2}+x'_{i_2}-x_2^*\right\|\ge\left\|x'_{i_1}-x'_{i_2}\right\|-\left\|x_1^*-x'_{i_1}\right\|-\left\|x'_{i_2}-x_2^*\right\|\ge d_0-\varepsilon-\varepsilon\ge\varepsilon,$$

即 $\left\|x_1^*-x_2^*\right\|\ge\varepsilon$.所以,式(22)说明 $\{y_j,j\ge N_\varepsilon\}$ 只可能在其中的一个超球 $S_{\varepsilon,i_0}$ 中,所以有 $\left\|y_j-x'_{i_0}\right\|<\varepsilon$.故序列 $\{y_j,j=1,2,...\}$ 收敛. □

## 4 权重与局部空间结构

本节我们讨论均值漂移算法中一些参数的自适应选择方法及其意义.

### 4.1 参数选择

我们觉得,参数 $w_i$ 和 $\Sigma_i$ 可以采用以下几种方法选取.

4.1.1 基于近邻的方法

假设 $k$ 是一个整数,$x_{i_1},...,x_{i_k}$ 是样本 $x_i$ 的 $k$ 个近邻,则有以下 4 种方式能够定义 $\Sigma_i$:

$$\Sigma_i=\frac{1}{k}\sum_{j=1}^{k}\left(x_{i_j}-x_i\right)\left(x_{i_j}-x_i\right)^T \tag{24}$$

或定义为

$$\Sigma_i=\frac{1}{k}\sum_{j=1}^{k}\left(x_{i_j}-\overline{x_i}\right)\left(x_{i_j}-\overline{x_i}\right)^T \tag{25}$$

其中 $\overline{x_i}=\sum_{j=1}^{k}x_{i_j}\Big/k$ , $k\ge m$ , $m$ 是样本空间的维数,即 $x_i\in R^m$ . $k\ge m$ 是为了保证 $\Sigma_i$ 的正定性.

为了简单起见,$\Sigma_i$ 也可以定义为

$$\Sigma_i=tr\left(\frac{1}{k}\sum_{j=1}^{k}\left(x_{i_j}-x_i\right)\left(x_{i_j}-x_i\right)^T\right)I_{m\times m} \tag{26}$$

或者定义为

$$\Sigma_i=tr\left(\frac{1}{k}\sum_{j=1}^{k}\left(x_{i_j}-\overline{x_i}\right)\left(x_{i_j}-\overline{x_i}\right)^T\right)I_{m\times m} \tag{27}$$

这时,$\Sigma_i$ 仅仅反映了不同样本点之间的差异,但没有体现各个样本点处局部空间结构的各向异性.令

$$d_{1,i}=\max_j\left\|x_i-x_{i_j}\right\| \tag{28}$$

$$d_{2,i}=\max_j\left\|x_i-\overline{x_{i_j}}\right\| \tag{29}$$

$$d_{3,i}=\frac{1}{k}\sum_{j=1}^{k}\left\|x_i-x_{i_j}\right\| \tag{30}$$

$$d_{4,i}=\frac{1}{k}\sum_{j=1}^{k}\left\|x_i-\overline{x_{i_j}}\right\| \tag{31}$$

则,$w_i$ 可以定义为

$$w_i=d_{l,i}\Big/\sum_{p=1}^{n}d_{l,p}\ ,\ l=1,2,3,4 \tag{32}$$

$$w_i=\left|\Sigma_i\right|\Big/\sum_{p=1}^{n}\left|\Sigma_p\right| \tag{33}$$

$$w_i=tr\left(\Sigma_i\right)\Big/\sum_{p=1}^{n}tr\left(\Sigma_p\right) \tag{34}$$

4.1.2 基于超球的方法

假设 $r>0$ , $x_{i_1},...,x_{i_{k_{r,i}}}$ 是所有满足 $\left\|x-x_i\right\|\le r$ 的样本,也就是说,$x_{i_1},...,x_{i_{k_{r,i}}}$ 是所有落在以 $x_i$ 为中心,$r$ 为半径的超球中的样本.于是,$\Sigma_i$ 可以按以下 4 种方式定义:

$$\Sigma_i = \frac{1}{k_{r,i}} \sum_{j=1}^{k_{r,i}} \left(\boldsymbol{x}_{i_j} - \boldsymbol{x}_i\right)\left(\boldsymbol{x}_{i_j} - \boldsymbol{x}_i\right)^T \tag{35}$$

$$\Sigma_i = \frac{1}{k_{r,i}} \sum_{j=1}^{k_{r,i}} \left(\boldsymbol{x}_{i_j} - \overline{\boldsymbol{x}_i}\right)\left(\boldsymbol{x}_{i_j} - \overline{\boldsymbol{x}_i}\right)^T \tag{36}$$

其中,$\overline{\boldsymbol{x}_i} = \frac{1}{k_{r,i}} \sum_{j=1}^{k_{r,i}} \boldsymbol{x}_{i_j}$;

$$\Sigma_i = tr\left(\frac{1}{k_{r,i}} \sum_{j=1}^{k_{r,i}} \left(\boldsymbol{x}_{i_j} - \boldsymbol{x}_i\right)\left(\boldsymbol{x}_{i_j} - \boldsymbol{x}_i\right)^T\right) I_{m\times m} = \left[\frac{1}{k_{r,i}} \sum_{j=1}^{k_{r,i}} \left(\boldsymbol{x}_{i_j} - \boldsymbol{x}_i\right)^T \left(\boldsymbol{x}_{i_j} - \boldsymbol{x}_i\right)\right] I_{m\times m} \tag{37}$$

$$\Sigma_i = tr\left(\frac{1}{k_{r,i}} \sum_{j=1}^{k_{r,i}} \left(\boldsymbol{x}_{i_j} - \overline{\boldsymbol{x}_i}\right)\left(\boldsymbol{x}_{i_j} - \overline{\boldsymbol{x}_i}\right)^T\right) I_{m\times m} = \left[\frac{1}{k_{r,i}} \sum_{j=1}^{k_{r,i}} \left(\boldsymbol{x}_{i_j} - \overline{\boldsymbol{x}_i}\right)^T \left(\boldsymbol{x}_{i_j} - \overline{\boldsymbol{x}_i}\right)\right] I_{m\times m} \tag{38}$$

$w_i$ 可以定义为

$$w_i = k_{r,i} \Big/ \sum_{j=1}^{n} k_{r,j} \tag{39}$$

从上面对参数 $w_i$ 和 $\Sigma_i$ 的定义可知,它们都决定于样本集合,所以本文算法中窗口的结构和相对尺度以及不同样本点贡献的相对权重都是自适应的.

### 4.2 权重$\boldsymbol{w_i}$

因为位于外围的样本点常常由于受噪声、不同类型之间的混叠和离群点等因素的影响较严重而使可信度较低,所以适当设置权重 $w_i$ 能够提高密度估计和均值漂移算法的鲁棒性.并且,可以通过适当设置权重 $w_i$ 而使我们关心的模式处更为显著.

虽然 Cheng[20]证明均值漂移算法时仅仅要求 $w_i > 0$,但据我们所知,在涉及参数 $w_i$ 选取时,文献中都假定 $w_1 = \ldots = w_n = 1/n$,即认为各个训练样本在算法中的贡献都相同.因为位于外围的样本点一般都比较稀疏,所以本文推荐的选择方法能够克服上述缺点,从而提高算法的鲁棒性.

### 4.3 局部空间结构$\Sigma_i$

通过选取不同的参数 $\Sigma_i$ 可以表达总体 $X$ 在各个样本 $\boldsymbol{x}_i$ 周围不同的局部尺度信息和局部空间结构.它既刻画了窗口在不同样本点之间的相对尺度差异,又表示了窗口在样本 $\boldsymbol{x}_i$ 处的各向异性.文献[4]中假设 $\Sigma_i = h_i I_{m\times m}$,而文献[2,7,20]假设 $\Sigma_i = I_{m\times m}$.然而,这个估计只有在样本集$\{\boldsymbol{x}_i\}$不相关,同分布,并且每个样本 $x_i$ 的不同方向结构也相同时才较为准确.但是,实际问题中很少出现这种理想的情况.所以,上述文献中的结果仅仅在极其特殊的情况下才能够快速而准确地找到局部极大值点.局部空间结构 $\Sigma_i$ 依赖于样本点在空间中的分布关系,它有很强的方向性和相关性.而本文推荐的选择方法(式(24),式(25),式(35),式(36))能够更好地刻画各个样本周围不同的局部尺度信息和局部空间结构.方法(式(26),式(27),式(37),式(38))虽然与文献[4]一样,仅仅刻画了各个样本 $\boldsymbol{x}_i$ 周围不同的局部尺度信息,但是文献[4]使用迭代算法确定这些参数,算法较复杂,本文方法则直接从样本点计算得到,更为简单.

## 5 结 论

本文将均值漂移算法推广到了下面的情况:反映了不同样本点局部空间结构的差异及其各向异性,并且算法中窗口的结构和相对尺度都是自适应的,从而拓宽了该算法的应用范围.并且,从数学上严格证明了均值漂移算法的收敛性,为该算法的应用提供了理论保证.此外,本文所给出的权重与局部空间结构参数的选择方法对实际应用也有一定的指导意义.

当然,该算法还有许多需要进一步研究的问题,例如:迭代过程的收敛速度,什么样的核函数 $k(x)$ 能使 $\hat{f}(\boldsymbol{x})$ 在 $S_0$ 上存在有限个局部极值点,以及参数的不同选择方式对算法有何影响等.

**致谢** 在此,我们非常感谢中国科学院数学院应用数学研究所的关庆阳同志,在工作开始阶段对收敛性的讨论.

**References**:

[1] NummiaroK, Koller-Meier E, Van Gool L. An adaptive color-based particle filter. Image and Vision Computing, 2002,21(1): 99−110.

[2] Comaniciu D, Ramesh V, Meer P. Real-Time tracking of non-rigid objects using mean shift. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2000. 142−149.

[3] Comaniciu D, Ramesh V. Mean shift and optimal prediction for efficient object tracking. In: Mojsilovic A, Hu J, eds. Proc. of the IEEE Int'l Conf. on Image Processing (ICIP). 2000. 70−73.

[4] Comaniciu D, Ramesh V, Meer P. The variable bandwidth mean shift and data-driven scale selection. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV). 2001. 438−445. http://citeseer.csail.mit.edu/comaniciu01variable.html

[5] Comaniciu D, Meer P. Mean shift analysis and applications. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV). 1999. 1197−1203. http://citeseer.ist.psu.edu/comaniciu00realtime.html

[6] Bradski GR. Computer vision face tracking for use in a perceptual user interface. Intel Technology Journal, 1998. http://developer.intel.com/technology/itj/q21998/articles/art_2.htm

[7] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002,24(5):603−619.

[8] Comaniciu D. An algorithm for data-driven bandwidth selection. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003, 25(2):281−288.

[9] Comaniciu D. Nonparametric information fusion for motion estimation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2003. 59−66. http://csdl.computer.org/comp/proceedings/cvpr/2003/1900/01/190010059abs.htm

[10] Comaniciu D, Ramesh V, Meer P. Kernel-Based object tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003, 25(5):564−575.

[11] Collins RT. Mean-Shift blob tracking through scale space. In: Proc. of the Conf. on Computer Vision and Pattern Recognition (CVPR). 2003. 18−20. http://csdl.computer.org/comp/proceedings/cvpr/2003/1900/02/190020234abs.htm

[12] Bradski GR. Real time face and object tracking as a component of a perceptual user interface. In: Proc. of the 4th IEEE Workshop on Applications of Computer Vision (WACV). 1998. 19−21. http://csdl.computer.org/comp/proceedings/wacv/1998/8606/00/86060214abs.htm

[13] Comaniciu D, Meer P. Robust analysis of feature spaces: Color image segmentation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 1997. 750−755. http://csdl.computer.org/comp/proceedings/cvpr/1997/7822/00/78220750abs.htm

[14] Zhou XS, Comaniciu D, Krishnan S. An information fusion framework for robust shape tracking. In: Proc. of the 3rd Int'l Workshop on Statistical and Computational Theories of Vision. 2003. http://csdl.computer.org/comp/trans/tp/2005/01/i0115abs.htm

[15] Barash D, Comaniciu D. A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift. Image and Video Computing, 2003,22(1):73−81.

[16] Comaniciu D, Meer P. Distribution free decomposition of multivariate data. Pattern Analysis and Applications, 1999,2(1):22−30.

[17] Comaniciu D, Ramesh V. Robust detection and tracking of human faces with an active camera. In: Proc. of the IEEE Int'l Workshop on Visual Surveillance. 2000. 11−18. http://csdl.computer.org/dl/proceedings/vs/2000/0698/00/06980011.pdf

[18] Comaniciu D. Image segmentation using clustering with saddle point detection. In: Proc. of the IEEE Int'l Conf. on Image Processing (ICIP). 2002. 297−300. http://ieeexplore.ieee.org/xpl/abs_free.jsp?arNumber=1038964

[19] Comaniciu D, Ramesh V, Bue AD. Multivariate saddle point detection for statistical clustering. In: Proc. of the European Conf. Computer Vision (ECCV). 2002. 561−576. http://link.springer.de/link/service/series/0558/bibs/2352/23520561.htm

[20] Cheng YZ. Mean shift, mode seeking, and clustering. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1995,17(8): 790−799.

[21] Xue CX. Real Function Theory and Functional Analysis. Beijing: Higher Education Press, 1993. 187−188 (in Chinese).

[22] Peters CA, Valafar F. Comparison of three nonparametric density estimation techniques using Bayes' classifier applied to microarray data analysis. In: Proc. of the Int'l Conf. on Mathematics and Engineering Techniques in Medicine and Biological Sciences. 2003. 119−125. http://www-rohan.sdsu.edu/~faramarz/papers/ME281-Peters-Valafar.pdf

[23] Bian ZQ, Zhang XG. Pattern Recognition. 2nd ed., Beijing: Tsinghua University Press, 2000. 65−72 (in Chinese).

**附中文参考文献**:

[21] 薛昌兴.实变函数与泛函分析.北京:高等教育出版社,1993.187−188.

[23] 边肇祺,张学工.模式识别.北京:清华大学出版社,2000.65−72.