均值漂移在背景像素模态检测中的应用*)

梁英宏　王知衍　曹晓叶　许晓伟
(华南理工大学计算机科学与工程学院　广州 510640)

摘　要　自适应背景更新是视频序列运动分割中的重要步骤，而背景像素分布的不规律性是对背景进行更新的困难所在。本文首先对背景像素值分布的模态性特点进行描述，然后提出采用均值漂移(Mean Shift)方法检测背景像素的模态数量，从而为背景建模提供依据，可以针对不同模态数量的背景像素采用不同的建模方法。这种基于背景像素模态分类的方法能够实现背景更新在精度和速度上的折中。

关键词　均值漂移，背景更新，模态检测，运动分割

Application of Mean Shift Algorithm in Mode Seeking of Background Pixel Values

LIANG Ying-Hong　WANG Zhi-Yan　CAO Xiao-Ye　XU Xiao-Wei
(School of Computer Science & Engineering, South China University of Technology, Guangzhou 510640)

Abstract　The background updating step is crucial to motion segmentation in video sequences. However, the irregular distributions of background pixel values make the background modeling complicated. First, the multi-modality problem of background pixel values is described. Then a novel method for background pixel classification by using mean shift based mode-seeking algorithm is presented, which can classify the background pixels as single mode or multiple mode pixels so that different updating methods can be applied. The presented method can help improve the speed of background reconstruction without reducing its precision.

Keywords　Mean shift, Background updating, Modal detection, Motion segmentation

1　引言

视频序列运动分割(运动前景分割、运动检测)是计算机视觉领域的一个研究热点，应用范围包括智能监控、人机交互、三维建模等方面。对视频序列进行运动分割，难点在于对背景进行建模和更新，以适应背景的动态变化。场景中的背景不仅会受到视频采集设备本身电子噪声的影响[1]，而且会受到光线以及噪声运动的影响，导致背景像素不会全部服从单模态分布[2]。如果只采用单模态分布进行背景建模，显然会存在一定的误差，所以一些学者提出一些方法[2~6]来解决背景像素所呈现的多模态性问题，而这些方法无一例外都对所有像素采用同一种模型进行描述。

对于背景自适应更新问题，我们考虑的解决方法是首先对某个场景进行背景像素模态检测，即判断背景中哪些像素服从单模态分布(单高斯分布)，哪些像素服从多模态分布。针对不同模态的像素采用不同的模型进行描述，例如对于服从单模态分布的像素，采用单高斯分布和 IIR 滤波器[14]进行更新，运算量低；对于多模态像素采用混合高斯模型[2]或者非参数模型[3]进行建模，精度高的同时运算时间也相应增加。采用这种方法依据在于：对于绝大部分场景，背景中的噪声主要是视频采集设备的电子噪音，而噪声运动区域只占一个较小的部分，所以没有必要对所有像素采用同一种复杂的建模方法。因此我们设计背景更新方法分成两个步骤，初始阶段进行背景像素模态检测，然后对像素进行建模更新，从而实现背景更新在精度和速度上的折中。图 1 显示了基于背景像素模态分类的背景更新方法流程图。

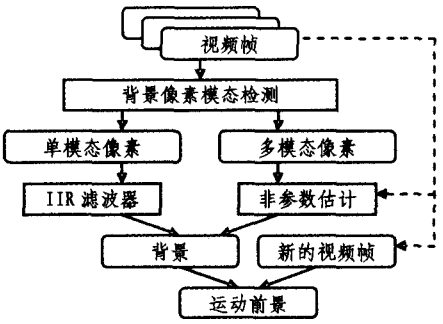


图 1　基于背景像素分类的背景更新方法流程图

均值漂移[7~11]由非参数核函数估计理论发展而来，通过不断迭代，可以寻找到概率密度函数的极值点。文[8,9]提出利用均值漂移可以检测概率密度函数中存在的模态，因此我们考虑采用均值漂移方法对背景像素值的概率分布进行模态检测。

本文第 2 节对背景像素的多模态性进行描述；第 3 节介绍本文提出的背景像素模态检测算法；第 4 节对算法进行实验；最后总结全文。

2　运动分割中的背景像素模态性问题

当场景中不存在任何微小运动区域的时候，获得的背景

*)科技部科技型中小型企业技术创新基金无偿资助项目(立项代码：02C26214400224)；广东省科技计划资助项目(项目编号：2002A1020104)。
梁英宏　博士生，研究方向：模式识别、图像处理；**王知衍**　教授，博士生导师，研究方向：计算机辅助设计、图像处理、模式识别；**曹晓叶**　博士生，研究方向：计算机图形图像；**许晓伟**　博士生，研究方向：计算机图形图像。

是相对静止的，背景中的像素仅仅受到电子噪声的影响，在分布上服从单高斯分布 $N(\mu,\sigma^2)$[14]。随着时间的变化，由于场景中的光线会发生缓慢变化，参数也会逐渐变化。当在背景相对静止的情形下，利用单高斯模型对背景进行更新，误差小且速度快，但背景中往往会存在一些噪声运动区域。例如：室外场景中随风摆动的树木、河面上的水纹，以及室内场景中闪烁的显示器等，这些噪声运动区域中像素的分布呈现多模态性，难以用单高斯模型描述。

对于解决背景中存在的噪声运动区域，文[2]提出采用混合高斯方法进行建模，将每个像素的分布用多个高斯分布进行描述。这种参数统计的方式最大的问题在于将像素分布的模态数量限定在事先设定的阈值范围内，不可能完全描述噪声运动区域的变化规律。文[3]提出背景某些区域的像素变化是不规律的，难以事先设定其参数模型，提出采用非参数统计模型进行描述，直接利用样本值以及选定的核函数估计概率，实际上是基于 Parzen 窗[12]的概率估计方法。如果提供足够的样本，选择合适的窗宽，非参数统计方法能够描述复杂背景的变化规律。其他的一些方法[4~6]能够很好地解决静态背景（单高斯分布背景），但是对动态背景（多模态背景、存在运动噪声区域的背景）的估计精确度较差。总结所有的背景更新方法，混合高斯方法以及非参数核函数估计方法是目前描述动态背景最精确的方法。然而前者存在的问题包括事先设定的混合高斯数量的阈值（通常为 3～5），不能满足对所有像素的变化规律进行描述，而增加参与混合的高斯数量将带来更多的计算量。后者不仅计算复杂，而且需要较大的内存空间。

实际上，背景中大部分像素都是服从单模态分布的，没有必要对所有的像素建立复杂的统计模型。由于视频运动分割对速度的要求较为苛刻，同时需要为后面的目标识别与场景理解操作保留一定的计算时间，复杂的背景建模方法显然难以满足实时应用的需求。我们提出对单模态像素和多模态像素采用不同的建模方法，以实现精度和速度上的折中。

对背景像素进行模态检测，也就是对背景中的每个像素在一段时间内的像素值的分布曲线进行峰值计数，以确定该像素是单模分布还是多模分布。文[13]将背景中的像素划分为静态背景、噪声背景和冲激背景，并采用多层神经网络训练方法对背景进行分类。该方法的问题在于训练时间长，且所构造的神经网络参数难以确定。我们的目标是找到一种能够在较短时间内完成的背景像素分类方法，对于概率分布的模态检测，均值漂移方法能够在较短的时间内收敛且精度高，下面将对该方法进行介绍。

3 背景像素模态检测算法

3.1 检测方法

令 $(x_1,x_2,\cdots,x_N)$ 为一段时间内的某个像素的灰度值，则该像素在 x 点的均值漂移的基本形式（不采用核函数）为

$$M(x)=\frac{1}{N}\sum_{i=1}^{N}(x_i-x) \tag{1}$$

$M(x)$ 指向 x 点出发的概率密度的梯度方向。但样本无论离 x 远近都对 $M(x)$ 的计算做同样的贡献，这显然是不合理的。所以引入核函数来计算 $M(x)$，采用核函数的均值漂移的形式为

$$M(x)=\frac{\sum_{i=1}^{N}K\left(\frac{x_i-x}{h}\right)w(x_i)(x_i-x)}{\sum_{i=1}^{N}K\left(\frac{x_i-x}{h}\right)w(x_i)} \tag{2}$$

式中，K 为核函数，用于根据样本点与 x 的距离对计算施加控制；$w(x_i)$ 为权值且 $\sum_{i=1}^{N}w(x)=1$，用于控制样本点参与计算的重要性；h 为带宽，用于控制概率密度估计的平滑度。对式(2)进行变形：

$$M(x)=\frac{\sum_{i=1}^{N}K\left(\frac{x_i-x}{h}\right)w(x_i)x_i}{\sum_{i=1}^{N}K\left(\frac{x_i-x}{h}\right)w(x_i)}-x=m(x)-x$$

$$m(x)=\frac{\sum_{i=1}^{N}K\left(\frac{x_i-x}{h}\right)w(x_i)x_i}{\sum_{i=1}^{N}K\left(\frac{x_i-x}{h}\right)w(x_i)} \tag{3}$$

式中，$m(x)$ 可以看作对 x 沿着概率密度梯度的方向作估计。如果设定每个样本的权值都相同，则 $m(x)$ 简化成

$$m(x)=\frac{\sum_{i=1}^{N}K\left(\frac{x_i-x}{h}\right)x_i}{\sum_{i=1}^{N}K\left(\frac{x_i-x}{h}\right)} \tag{4}$$

通过该像素一段时间的灰度值序列 $(x_1,x_2,\cdots,x_N)$，从每一个 x 出发，都可以计算 $m(x)$，并使 $m(x)$ 不断沿着概率密度梯度方向移动，达到峰值，即该序列的一个概率密度极值点。该过程是一个迭代过程，被称为均值漂移算法（mean shift algorithm）[8]。实际上很容易理解均值漂移算法是数据集的概率密度梯度的估计过程。如果计算从每一个点出发到达的峰值，不同峰值点的数量即为该序列概率密度的模态数。

如果采用颜色值（RGB 三个通道值）来检测背景像素模态，即在三维颜色空间中寻找概率密度的峰值数量，式(4)变形为

$$m(X)=\frac{\sum_{i=1}^{N}K\left(\frac{\|X_i-X\|}{h}\right)X_i}{\sum_{i=1}^{N}K\left(\frac{\|X_i-X\|}{h}\right)} \tag{5}$$

式中，X 是颜色向量 (x^R,x^G,x^B)。

我们将高斯核函数 $K(x)=(\sqrt{2\pi}h)^{-1}e^{-1/2(x/h)^2}$ 代入，则式(4)和式(5)改写成

$$m(x)=\frac{\sum_{i=1}^{N}e^{-\frac{1}{2}\left(\frac{x_i-x}{h}\right)^2}x_i}{\sum_{i=1}^{N}e^{-\frac{1}{2}\left(\frac{x_i-x}{h}\right)^2}} \tag{6}$$

$$m(X)=\frac{\sum_{i=1}^{N}e^{-\frac{(X_i-X)(X_i-X)'}{2h^2}}X_i}{\sum_{i=1}^{N}e^{-\frac{(X_i-X)(X_i-X)'}{2h^2}}} \tag{7}$$

上面提到模态的检测是一个迭代过程。为了使该过程迅速收敛，设定一个容许误差 ε，每次计算出 $m(x)$，都与 x 相比较。如果 $|m(x)-x|<\varepsilon$（对于向量 X，是 $\|m(X)-X\|<\varepsilon$），则停止检测；否则令 $x=m(x)$，继续计算 $m(x)$。如此反复，直到满足要求的 $m(x)$ 出现，此时 $m(x)$ 近似于概率密度的极值。文[9]中证明了该迭代过程的收敛性和收敛条件，因此通过一定次数的计算一定可以找到该点附近的极值。

图 1 显示了某个室外背景中的两个像素的值在一段时间内的分布情况（1000 帧，为了便于显示，只采用 R 和 G 两个颜色通道），以及这两个像素的模态检测结果。

图 1(b)左显示了像素（100，100）的分布，该像素属于静态背景像素，没有发生显著变化，属于单高斯分布；图 1(b)右显示像素（230，180）的分布，该像素所处的位置是一棵树木，所以该像素的分布有跳跃性。图 1(c)显示了这两个像素采用均值漂移算法的模态检测结果。显然，前一个是单模态像

素，最终可以收敛到一个峰值，而后一个属于多模态像素，无法用一个峰值来表示。

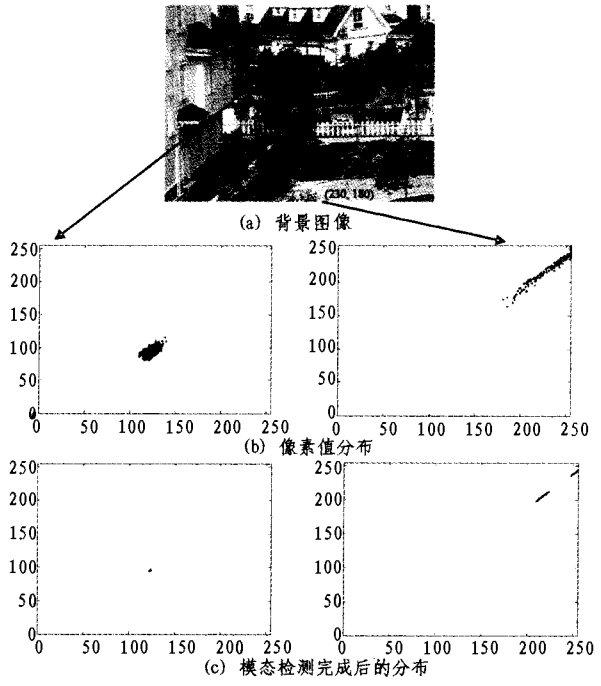


图 1　某个场景中的像素值分布及其模态检测结果

3.2　样本数和带宽的选择

均值漂移来源于非参数统计理论中的 Parzen 窗概率密度估计。样本的大小和窗宽都会影响到 Parzen 窗概率密度曲线的精确性[12]。一般来说，当样本数一定时，窗宽大，估计结果分辨率低，有可能会将靠近的峰值合并；窗宽小，统计的稳定性不足，会产生很多尖峰。窗宽的选择显然与样本数量相关，当样本数增加时，概率密度估计的精确度也会提高。

样本数 N(这里是指用于模态检测的视频序列的帧数)当然是越多越好，但考虑到内存占用和计算时间的限制，N 通常取一个合理值，例如 300，即取 300 个连续的图像帧。如果每秒取 15 帧，那么就是 20s 的场景视频，已经可以覆盖噪声运动的周期性(如树木随风摆动的周期)，无需进一步增大样本数。

文[15,16]提出的非参数估计的带宽选择方法，可以通过以下公式简单计算获得优化带宽：

$$h_j^{opt}=\left(\frac{4}{2d+1}\right)^{1/(d+4)}\hat{\sigma}_j\, n^{-1/(d+4)} \tag{8}$$

式中，d 表示维度，n 是样本数量，$\hat{\sigma}_j^2$ 表示第 j 维分布的方差。如果采用灰度图进行计算，带宽 h 的计算公式为

$$h^{opt}\approx 1.06\hat{\sigma}n^{-1/5} \tag{9}$$

3.3　算法描述

均值漂移模态检测算法可以总结为：采集一段时间的背景视频序列，对于背景中的每一个像素，首先计算优化带宽，然后从每一个值开始运行均值漂移迭代过程，一直到误差在设定的容许误差以内时才停止。如果所有均值都漂移到同一个值，就表示该像素是一个单模态像素，反之就属于多模态像素。如果对多模态像素采用混合高斯分布模型进行更新，可以对多模态像素进一步划分，以便确定参与混合的高斯分布数量。

本文提出的背景模态检测算法描述如下：

$I(x, y, i)$为像素(x, y)在第 i 帧的像素值，$i = 0,1,2,\cdots,N$。

Step 1 设定容许误差 ε。

Step 2 对背景中每一个像素(x, y)，执行 Step 3-8。

Step 3 采用公式(9)计算优化带宽 $h^{opt}(x, y)$。

Step 4 For $i = 0, 1, \cdots, N$。

Step 5 采用公式(6)计算均值 m(对于颜色向量，采用公式(7))。

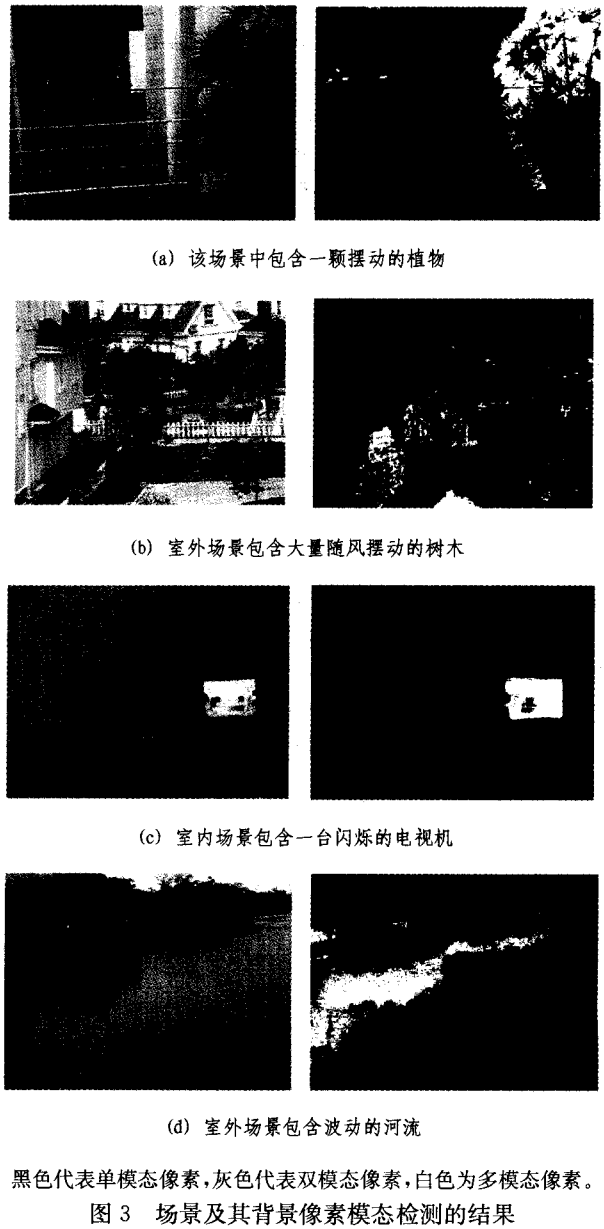
Step 6 如果$|m-I(x,y,i)|>\varepsilon$，令 $I(x,y,i)=m$，重复 Step 5。当$|m-I(x,y,i)|<=\varepsilon$时停止对均值 m 进行计算。(注意，当 $I(x, y, i)$ 是颜色向量时，$|m-I(x,y,i)|>\varepsilon$ 应该是 $\|m-I(x,y,i)\|>\varepsilon$)。

Step 7 将 m 的计算结果存放于数组 Results 中。

Step 8 数组 Results 中如果两个值的差别小于容许误差 ε，则合并这两个值。计算 Results 中不同值的数目。如果数目为 1，则代表该像素为单模态像素，反之为多模态像素。

Step 9 一旦背景中所有像素的模态检测完成，就可以通过不同的背景更新方法进行背景更新。

4　试验



黑色代表单模态像素，灰色代表双模态像素，白色为多模态像素。

图 3　场景及其背景像素模态检测的结果

(下转第 237 页)

Background Segmentation by Tracking Spatial-Color Gaussian Mixture Models. IEEE Workshop on Motion and Video Computing 2007, Austin, TA, Feb. 2007
6 Wang Y, Loe K, Tan T, et al. Spatiotemporal Video Segmentation Based on Graphical Models. IEEE Trans. On Image Processing, 2005, 14(7): 937～947
7 Chien S, Huang Y, Hsieh B, et al. Fast Video Segmentation Algorithm with Shadow Cacellation, Global Motion Compensation, and Adaptive Threshold Techniques. IEEE Trans. on Multimedia, 2004, 6(5): 732～748
8 Liu Y, Zheng Y F. Video Object Segmentation and Tracking Using Learning Classification. IEEE Trans. on Circuits System and Video Technology, 2005, 15(7): 885～899
9 Chen P C, Su J J, Tsai Y P. Coarse-To-Fine Video Object Segmentation By MAP Labeling of Watershed Regions, Bulletin of College of Engineering, National Taiwan University, 2004(9): 25～34
10 Malik J, Shi J. Normalized cuts and image segmentation. In: Proceedings of IEEE Conf. Computer Vision and Pattern Recognition, 1997
11 Feleznszwalb P, Huttenlocher D. Image segmentation using local variation. In: Proceedings of IEEE Conf. Computer Vision and Pattern Recognition, 1998. 98～104
12 Cormen T H, Leiserson C E, Rivest R L, Stein C. Introduction to Algorithms, Second Edition. MIT Press and McGraw-Hill, ISBN 0-262-03293-7. Chapter 26: Maximum Flow, 2001. 643～700
13 孙仲康，沈振康. 数字图像处理及其应用[M]. 北京：国防工业出版社，1985
14 Zitova B, Flusser J. Image RegistrationMethods: A Survey[J]. Image and Vision Computing, 2003, 21 (11): 977～1000
15 Lisa G B. A Survey of Image Registration Techniques[J]. ACM Computing Surveys, 1992, 24 (4): 325～376
16 曹菲，杨小冈，缪栋，等. 景象匹配制导基准图选定准则研究[J]. 计算机应用研究，2005，22 (7)：137～139
17 杨小冈，曹菲，缪栋，等. 基于相似度比较的图像灰度匹配算法研究[J]. 系统工程与电子技术，2005，27(5)：918～921
18 Cohen J, Cohen P, West S G, Aiken L S. Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, 2003

(上接第 230 页)

我们对多个视频序列(均为 320×240 的分辨率)采用本文提出的算法进行测试，包括室外场景和室内场景，每个场景都包含噪声运动区域。为了提高检测速度，采用灰度图进行计算。图 3 显示了其中的 4 个测试视频及其检测结果。图 3(a)显示某个场景中存在一个摆动的植物，经过背景模态检测后，植物的枝叶部分被划分成多模态；图 3(b)显示一个包含大量树木的室外场景，这是一种比较极端的情况，经过背景模态检测，运动的植物区域被分类出来；图 3(c)显示一个室内场景，只有电视机部分以及一些反光的边缘被划分为多模态背景区域；图 3(d)显示存在一条河流的场景，河流波动较大的区域被划分成多模态背景区域。

总结 与文[13]提出的背景分类方法相比，本文提出的方法在速度上有着明显的优势，平均在数分钟内就能完成计算，而文[13]提出的基于神经网络的方法需要数小时乃至数天的运算时间。由于在大部分监控应用中，所监控的场景都不是频繁变化的，所以本文的方法能够满足实际的需求。

对于背景更新问题，所提出的方法都是采用单一模型对所有像素进行更新，没有考虑到背景中像素的模态性差异。很多复杂的方法尽管能够描述多模态背景像素值的分布，但是计算复杂，难以应用于实际。我们认为：由于一般情况下，背景中的动态区域所占背景的比例较小，所以没有必要为所有像素采用同一个复杂的更新方法，所以事先对背景像素进行模态检测就可以为后续的背景更新操作提供方法选择上的依据，从而在不降低精确度的同时大幅提高背景更新速度。下一步的工作是优化该算法，使其运算速度进一步提高。

参 考 文 献

1 Toyama K, Krumm J, Brumitt B, et al. Wallflower: principles and practice of background maintenance. In: Proceedings of the IEEE International Conference on Computer Vision. Kerkyra Greece, 1999. 255～261
2 Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol 2, 1999. 23～25
3 Elgammal A, Harwood D, Davis L. Non-parametric model for background subtraction. In: Proceedings of the European Conference on Computer Vision. Dublin Ireland, 2000. 751～767
4 侯志强，韩崇昭. 基于像素灰度归类的背景重构算法. 软件学报，2005，16(9)：1569～1576
5 Haritaoglu I, Harwood D, Davis L S. W4: real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 809～830
6 Ridder C, Munkelt O, Kirchner H. Adaptive background estimation and foreground detection using Kalman-filter. In: Proceedings of the International Conference on Recent Advances in Mechatronics. UNESCO Chair on Mechatronics, 1995. 193～199
7 Fukanaga K, Hostetler L D. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans on Information Theory, 1975, 21(1): 32～40
8 Cheng Y. Mean shift, mode seeking and clustering. IEEE Trans on Pattern Analysis and Machine Intelligence, 1995, 17(8): 790～799
9 Comanicu D, Meer P. Mean shift: A robust approach toward feature space analysis. IEEE Trans Pattern Analysis and Machine Intelligence, 2002, 24(5): 603～619
10 Meer P, Georgescu B. Edge detection with embedded confidence. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(12): 1351～1365
11 Christoudias C, Georgescu B, Meer P. Synergism in low level vision. In: Proceedings of the 16th International Conference of Pattern Recognition. Quebec City, Canada, 2001. 150～155
12 Parzen E. On the estimation of a probability density function and mode. Ann Math Stat, 1962, 33: 1065～1076
13 Gil-Jim'enez P, Maldonado-Basc'on S, Gil-Pita R, et al. Background Pixel Classification for Motion Detection in Video Image Sequences. Lecture Notes in Computer Science, 2003, 2686: 718～725
14 Wren C, Azarbayejani A, Darrell T, et al. Pfinder: Real-time Tracking of the Human Body. IEEE Trans on Pattern Analysis and Machine Intelligence, 1997, 19(7): 780～785
15 Scott D W. Mulivariate Density Estimation. Wiley-Interscience, 1992
16 Turlach B A. Bandwidth selection in kernel density estimation: A review. Discussion Paper. 9317, Institut de Statistique, Voie du Roman Pays 34, B-1348 Louvain-la-Neuve, 1993