

# 基于像素层背景模型的复杂场景运动目标检测

韩建平<sup>1),2)</sup> 张明敏<sup>1)</sup> 潘志庚<sup>1)</sup>

<sup>1)</sup> (浙江大学 CAD&CG 国家重点实验室, 杭州 310027)

<sup>2)</sup> (杭州电子科技大学图形图像研究所, 杭州 310018)

**摘要** 为了从复杂变化背景中鲁棒地检测、提取运动目标,提出一种基于像素层背景模型的运动目标检测算法。该算法采用快速均值漂移方法将背景帧上具有相同统计特性的像素划分为一个像素层,背景模型从而被表示为一组像素层,通过与邻域像素对应的层匹配来检测运动前景像素。实验结果表明,该方法可以实时、准确地检测运动目标,特别是在摄像机颤动等原因造成的背景时域不规则变化情况下,比经典的基于混合高斯背景模型的方法具有更好的检测效果。

**关键词** 视频分析 运动目标检测 背景模型 视觉监控

中图法分类号: TP309.7 文献标志码: A 文章编号: 1006-8961(2010)08-1201-06

## A Pixel Layer Based Background Model for Moving Objects Detection in a Dynamic Scene

HAN Jianping<sup>1),2)</sup>, ZHANG Mingmin, PAN Zhigeng<sup>1)</sup>

<sup>1)</sup> (State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310027)

<sup>2)</sup> (Institute of Graphics and Image, Hangzhou Dianzi University, Hangzhou 310018)

**Abstract** A novel background model based on pixel layer for moving objects detection in a dynamic scene is presented in this paper. Fast mean shift approach is used to cluster into layers where those pixels share similar statistics. The background is then modeled as a group of pixel layers. An incoming pixel is detected as foreground if it does not adhere to these layer-models of the background. The experiments show that the proposed method performs better than the traditional MoG method under dynamic background and especially in the presence of nominal camera motion.

**Keywords** video analysis, moving objects detection, background model, visual surveillance

## 0 引言

实时检测运动目标是智能视频监控、行为识别等视觉应用的基础。背景减除是在摄像机固定条件下常用的一种运动目标检测方法,即通过当前帧和背景模型匹配来检测和提取运动目标。构建背景模型是背景减除方法中的关键技术,决定运动目标检测的效率、准确性和鲁棒性。其困难在于自然场景

中背景是变化的,例如背景中树枝随风飘动、水面波动以及摄像机颤动,可以呈现一定的伪周期性,也可以是突发的,这些复杂的背景变化给运动目标检测提出了挑战。

国内外研究者针对实际问题,提出了许多背景模型及其运动目标检测方法。文献[1]建立参数化的混合高斯模型(MoG)来表示每个像素的概率分布,通过K-均值近似算法对模型进行自适应更新,一定程度上能适应多峰分布的背景变化。此后,研

基金项目:国家高技术研究发展计划(863)项目(2009AA062704);上海世博科技专项(09-524005-001)

收稿日期:2009-01-08;改回日期:2009-04-22

第一作者简介:韩建平(1969—),男,副教授。浙江大学CAD&CG国家重点实验室博士研究生。主要研究方向为计算机视觉、数字图像处理。E-mail: hanjp@hdu.edu.cn

通讯作者:张明敏。E-mail: zmm@cad.zju.edu.cn

究者又对混合高斯背景模型进行了改进。文献[2]利用最大似然估计在混合高斯模型中引入了高斯分布个数的自动选择机制;文献[3]提出自适应的学习速率,从而改善模型参数更新的收敛速度。文献[4]通过应用颜色和纹理信息的线性组合,以提高分割的准确性。高斯混合模型依赖高斯分布以及参数结构的先验假设,难以准确描述复杂的背景变化。为此,文献[5]提出了一种基于核密度估计的非参数背景建模算法(KDE),该方法不用对背景的概率模型做任何假设,直接利用先前获得的像素值估计当前像素为背景的概率,不仅能够适应复杂的像素分布,而且能够适应像素在短时间内发生的频繁变化。文献[6]在非参数核密度估计中引入快速高斯变换以提高非参数背景模型的计算速度。文献[7]提出自适应前景、背景阈值选择,以解决误报率和漏报率间的矛盾。相对于 MoG 方法、KDE 方法可以得到更好的检测结果,但计算复杂度高,一般难以满足实时要求。

当背景变化在时间窗内呈现类周期性时,上述基于时域分布的方法一般可以相对准确地检测运动目标。然而如果背景上出现非平稳变化,如阵风引起树木晃动或摄像机的颤动,导致背景物局部运动而被误检为运动像素。为此,文献[8]在混合高斯模型的框架内,将匹配及参数更新拓展到以像素  $7 \times 7$  的邻域像素,检测鲁棒性显著提高,但计算复杂度大大增加。文献[9-11]则在运动对象检测或分割中采用了空域全局背景模型。但由于全局模型丢失了位置信息,准确性显然会受到影响,文献[9]用局部和全局的混合加权背景模型,单高斯模型来表示每个像素的时域分布,而用混合高斯模型来表示空域全局分布特征,但两者权重系数的确定困难且缺乏足够的理论依据。文献[10]则假定背景是不变的,因而各背景像素对应的空域全局高斯分布也是确定的,算法的空间和时间复杂度都大大下降,但显然无法处理动态变化背景。

针对上述问题,提出一种基于像素层背景模型的运动目标检测方法。算法着眼于利用背景运动的局部性和时空变化相关性特征,在复杂变化背景下鲁棒、准确地检测运动目标。背景上具有相同统计特性的像素被划分为一个像素层,背景模型从而表示为一组像素层。运动前景像素的检测通过与空域相邻的层匹配来实现。在背景上出现时域非平稳变化时,仍然可以快速鲁棒地检测运动目标。

## 1 算法思想

在无运动目标遮挡的情况下,除光照变化、噪声影响外,背景像素颜色时域的变化一般是空域变化引起的,如树枝等背景物体的晃动、摄像机的颤动等。假定背景运动具有局部性特征,即在一个相对小的范围内运动。当局部性运动频繁具有一定类周期性时,会使得相应像素值成为一定时间内分布的峰值之一,从而可以被 MoG、KDE 等基于时域统计方法正确判定为背景。然而在许多应用中,特别是户外条件下,突发的背景物体局部运动经常发生,如阵风导致树枝晃动或摄像机颤动,此时基于像素的时域统计方法难免产生错误判定。为此,不在时域统计像素颜色分布信息,而在背景运动局部性假设的基础上,着眼于利用像素空间相关性,将背景模型表示为一组像素层,其中像素层为背景帧上具有相同统计特性的像素集,用高斯分布描述层的全局统计特性。

图 1 描述了本文算法的基本框架。在初始化阶段,在对背景帧初步划分的基础上,利用快速均值漂移(mean shift)算法来对背景帧进行颜色特征聚类,将背景帧分解为一组像素层,并估计每个层高斯分布参数。前景运动像素的检测通过与空域相邻的层匹配来实现,以处理背景物体的局部运动。设  $p$  为当前帧上任意像素,运动检测就是判定  $p$  属于前景运动物体还是背景层。经过初始化阶段的像素层划分,得到一组背景像素层,以及表明每个像素所属背景层的标记矩阵。

在户外条件下,波动的水面、随风飘动的树叶以及摄像头晃动等干扰因素,都可能使得像素周围的背景层运动而出现在  $p$  处。背景像素具有不确定性,呈现的不是单一不变的像素层;但同时背景运动具有局部性,出现在  $p$  处的背景层一般就来自  $p$  附近。为此,在以  $p$  为中心的  $w \times w$  窗口内,匹配  $p$  对应的背景层,从而判定像素  $p$  为前景或背景。

## 2 算法过程描述

### 2.1 背景层划分与背景模型构建

设  $B = \{y_i \mid i = 1, \dots, N\}$  为背景帧全体像素集合, $y_i$  为像素  $i$  的 RGB 向量。采用均值漂移算法在 RGB 颜色空间对背景帧进行聚类,将背景帧划分为

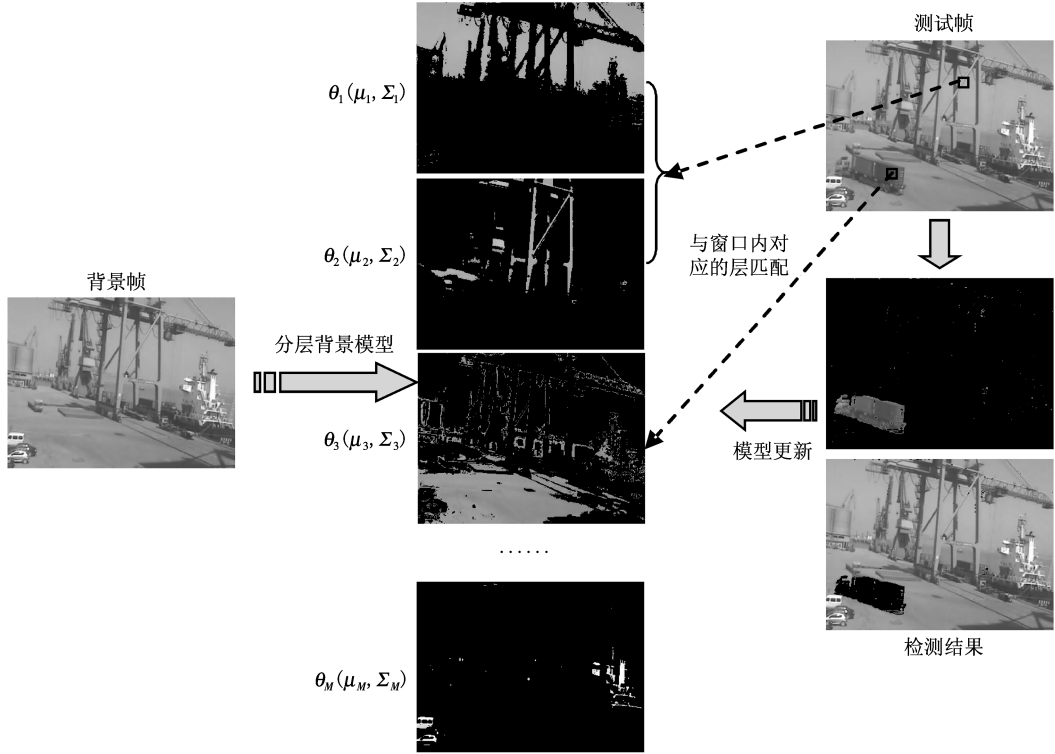


图 1 算法框架

Fig.1 Algorithm framework

一组层  $L_k(\mu_k, \Sigma_k)_{k=1, \dots, M}$ , 并估计各层的参数。  $\mu_k$ ,  $\Sigma_k$  分别为第  $k$  层的均值向量和协方差矩阵,  $M$  为层数。

均值漂移算法的显著优势在于鲁棒性以及不依赖于聚类数目先验知识, 层数  $M$  不需要人为设定或复杂计算。

本文的目标是在复杂变化的背景下, 快速鲁棒地检测运动物体, 而不是细致分割前景目标。为了提高模型构建的效率, 采用快速简化的均值漂移算法<sup>[12]</sup>。首先将背景帧全体像素集合  $B$  初步划分为若干个子集  $S = \{s_j(n_j, c_j) | j = 1, \dots, m\}$ ,  $n_j, c_j$  分别为子集  $s_j$  包含的元素个数和均值向量。步骤如下:

- 1) 将  $S$  初始化为  $\{s_1(y_1, 1)\}$ 。
- 2) 对  $k = 2, 3, \dots, N$ , 如果存在  $j$ , 使得范数  $\|y_k - c_j\|_\infty$  小于阈值  $r$ , 那么  $y_k$  划分到  $s_j$ ; 否则生成新的子集  $s_{new}(y_k, 1)$ 。

- 3) 更新各子集的均值向量  $c_j = (\sum y_i \in s_j) / n_j$ 。在  $\{s_j(n_j, c_j) | j = 1, \dots, m\}$  上, 从  $c_j^{(0)} = c_j$  开始进行如式(1)的均值漂移过程。

$$c_j^{(k+1)} = \frac{\sum_{i=1}^m n_i c_i g\left(\left\|\frac{c_j^{(k)} - c_i}{h}\right\|^2\right)}{\sum_{i=1}^m n_i g\left(\left\|\frac{c_j^{(k)} - c_i}{h}\right\|^2\right)} \quad (1)$$

式中,  $g(\|x\|^2)$  为一个中心在零点, 定义域积分为 1 的非负函数, 本文采用高斯核函数,  $h$  为平滑带宽, 实验中选择  $h = 16$ 。

通过均值漂移迭代过程, 可以得到一组聚类中心  $\mu_i (i = 1, \dots, Q)$ 。对任意两个聚类中心  $\mu_a$  和  $\mu_b$ , 如果  $\|\mu_a - \mu_b\| < \varepsilon$ , 则将它们合并。设  $\{c_{k1}, c_{k2}, \dots, c_{kt}\}$  为所有收敛于  $\mu_k$  的子集中心,  $c_{k1}, c_{k2}, \dots, c_{kt}$  对应的全部像素构成一个像素层,  $\mu_k$  作为该层的均值向量。为了简化计算协方差矩阵  $\Sigma_k$ , 假定 RGB 通道变化是相互独立的,  $\Sigma_k$  为对角矩阵, 利用属于同一层的像素分别计算得到每个通道的方差。

### 2.2 运动检测

给定像素  $p$ , 设  $L_k(\mu_k, \Sigma_k)$  是以  $p$  为中心的  $w \times w$  窗口内像素所属的一个背景层, 如果像素  $p$  的颜色  $y_p$  与层  $L_k$  如式(2)的 Mahalanobis 距离的平方小于设定阈值  $T_{dist}$ , 则  $y_p$  与层  $L_k$  高斯分布匹配,  $p$  属于背景层  $L_k$ , 是背景像素。相反, 如果窗口内所有

像素对应的背景层都不与  $y_p$  匹配,则像素  $p$  被判定为前景。

$$dist_k^2 = (y_p - \mu_k)^T \Sigma_k^{-1} (y_p - \mu_k) \quad (2)$$

尽管匹配过程涉及  $w \times w$  个像素,但前景判定的计算复杂度并不高。一方面,虽然背景模型可能包含许多个层,但空间相关性决定了  $w \times w$  窗口内包含的背景层一般较少。另一方面,由于多个像素对应同一层,可以通过预计算来避免距离的冗余计算,提高匹配计算效率。

窗口大小  $w$  的取值与背景局部运动的程度相关。如果摄像机是完全静止的,且背景没有局部运动, $w$  可以为 1,此时本文方法与文献[10]是相似的。在户外动态复杂变化背景以及摄像机颤动等条件下, $w$  取值可以为 7~15,实验中  $w$  取值为 11。

### 2.3 背景更新

背景更新一方面是对背景层的参数进行更新,以适应光线变化对背景模型的影响,另一方面使得运动物体静止后可以融为背景。

对匹配同一背景层的像素点,我们不逐一计算更新背景层参数,而是在统计直方图的基础上实现批量更新,从而显著提高更新计算效率。设当前帧中所有与背景层  $j$  匹配的像素点的直方图为  $H^c[k]$  ( $c=R,G,B;k=0,1,\dots,255$ ),则对背景层  $j$  的高斯分布参数  $(\mu_j, \Sigma_j)$  作如下更新:

$$\mu_j^c = (1 - \alpha)\mu_j^c + \alpha \left( \frac{\sum_{k=k_{\min}^c}^{k_{\max}^c} H^c[k]k}{\sum_{k=k_{\min}^c}^{k_{\max}^c} H^c[k]} \right) \quad (3)$$

$$\Sigma_j^c = (1 - \alpha)\Sigma_j^c + \alpha \left( \frac{\sum_{k=k_{\min}^c}^{k_{\max}^c} (k - \mu_j^c)^2 H^c[k]}{\sum_{k=k_{\min}^c}^{k_{\max}^c} H^c[k]} \right) \quad (4)$$

式中, $\alpha$  为更新速率, $k_{\min}^c, k_{\max}^c$  分别为通道  $c$  的最小值和最大值,与一个背景层匹配的像素,其各通道取值显然分布在一个很小的范围内,当  $k < k_{\min}^c$  或  $k > k_{\max}^c$  时, $H^c[k]$  值为 0。

背景的局部运动会带来像素所属层的变化,但本文的运动目标检测方法对像素所属背景层的局部变化并不敏感,因此可以每隔  $T$  帧对标记矩阵  $X$  更新一次。对判定为背景的像素,用其所属的背景层序号替换原标记矩阵  $X$  中对应的序号;而对判定为前景的像素,原标记矩阵  $X$  中对应的序号保持不变。

如果运动物体持续静止,按照第 2.1 节介绍的

背景层划分方法,将物体分解为一组新的背景层,合并到背景模型中。当一个背景层如果持续未被匹配超出一个较长的时间,即从背景模型中删除。

## 3 实验结果与分析

在多个视频上测试了本文算法,并与经典的 MoG 算法进行了比较。本文所用的测试视频如图 2 所示,序列 1 为 waving trees 视频<sup>[13]</sup>,背景中包含树枝显著晃动;序列 2 包含了树叶飘动、喷泉等不断变化的杂乱背景;序列 3 和序列 4 分别是交通和港口监控的视频,在背景动态变化的同时,摄像机在测试帧发生显著抖动。图 2 中所有的检测结果均未作腐蚀、膨胀、滤波等处理。MoG 对时域多峰背景具有一定的适应性,但由于在视频序列 1 和序列 2 中测试帧背景变化在时域不规则,依然产生了较多的虚警,本文算法能够克服这种不规则动态背景变化,虚警大大低于 MoG 算法。由于摄像机发生颤动,MoG 算法在序列 3 和序列 4 的测试帧上产生了大量的虚警,而本文算法的检测结果明显优于 MoG 方法。

图 3 给出了 MoG 算法及本文算法对上述 4 个视频检测结果的定量评价。评价指标采用计算真报警率 (TPR) 和误报警率 (FPR)。

$$TPR = TP/TFG \quad (5)$$

$$FPR = FP/TBG \quad (6)$$

式中, $FP$  指被错误检测为前景的像素数, $TP$  是被正确检测为前景的像素数, $TFG$  和  $TBG$  为手工分割结果的前景和背景像素数。图 4 给出了 waving trees 视频中包含运动前景的第 243~259 帧分割结果的比较分析。可以看出:对真报警率指标,本文算法与 MoG 算法基本相当,而对误报警率指标,本文算法显著优于 MoG 算法,特别是在摄像机发生颤动时。

图 5 比较了 MoG 算法及本文算法的处理速度,实验测试硬件环境为 CPU P4/3.0 GHz、内存 1GB 的台式计算机,算法实现的编程环境为 Visual C++ 7.0。本文算法与 MoG 算法相比速度接近,稍优于 MoG 算法。

本文算法在“基于网络流媒体的海事场景智能监控系统”中得到应用(如图 6)。该系统的功能之一是用户定义监控区域,系统检测受控区域内的运动物体并实时告警。海事场景中自然景物扰动频繁,监控摄像机也常在阵风中颤动,本文算法在这样的条件下可以实时监控 4~8 路网络视频流。



图 2 运动目标检测结果比较

Fig. 2 Comparison of moving object detection results

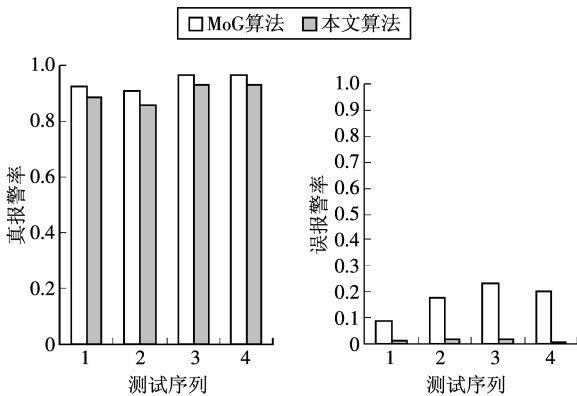


图 3 真报警率和误报警率的定量比较

Fig. 3 Quantitative evaluations by TPR and FPR

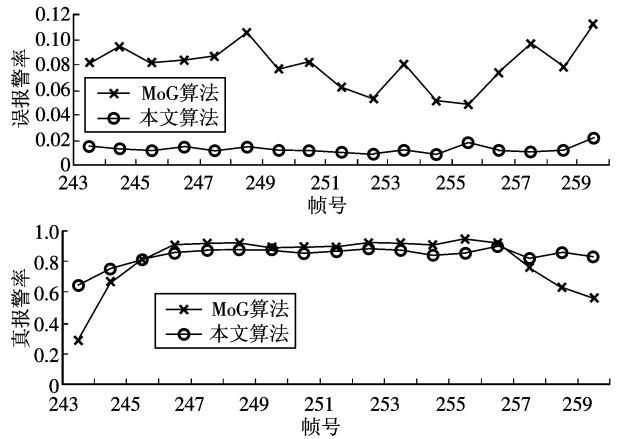


图 4 Waving trees 检测结果的比较

Fig. 4 Quantitative comparison on waving trees by TPR and FPR

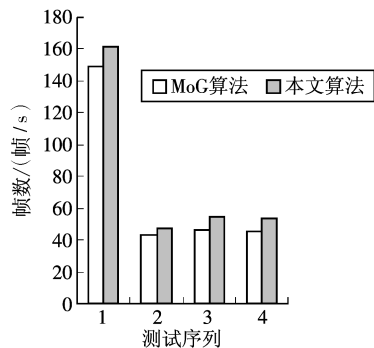


图 5 检测速度的比较

Fig. 5 Comparison of processing speed



图 6 海事场景智能监控系统

Fig. 6 Real-time marine video surveillance system

## 4 结 论

提出了一种复杂场景下的运动目标检测新方法。有效利用了背景运动的局部性和时空变化相关性特征,建立基于像素层的背景模型,在背景物突发运动和摄像机抖动等原因造成的像素时域变化分布不规则的情况下,仍然可以鲁棒地检测运动目标。此外,本文算法的算法效率也稍优于经典的 MoG 算法。

需要指出的是:本文算法基于这样的假设,除噪声和光照变化的影响,背景像素的变化是背景物空域局部运动引起的,这反映了实际应用中的一般情况。但仍然存在一些特殊的情形,如背景中有一个红绿灯交替闪烁,此时本文算法将会产生错误的检测结果,这是进一步工作要解决的问题。

## 参考文献 (References)

- [ 1 ] Stauffer C, Grimson W. Adaptive background mixture models for real-time tracking [ C ] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC, USA: IEEE Computer Society, 1999: 246-252.
- [ 2 ] Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction [ C ] // Proceedings of the International Conference on Pattern Recognition. Washington DC, USA: IEEE Computer Society, 2004: 28-31.
- [ 3 ] Lee D. Effective Gaussian mixture learning for video background subtraction [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5): 827-832.
- [ 4 ] Xiang Shiming, Chen Rui, Deng Yu, et al. Motion segmentation via on-line Gaussian mixture model and texture [ J ]. Journal of Computer Aided Design & Computer Graphics, 2005, 17(7): 1504-1509. [ 向世明, 陈睿, 邓宇, 等. 在线高斯混合模型和纹理支持的运动分割 [ J ]. 计算机辅助设计与图形学学报, 2005, 17(7): 1504-1509. ]
- [ 5 ] Elgammal A, Duraiswami R, Harwood D, et al. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance [ J ]. Proceeding of the IEEE, 2002, 90(7): 1151-1163.
- [ 6 ] Elgammal A, Duraiswami R, Davis L S. Efficient kernel density estimation using the fast gauss transform with applications to color modeling and tracking [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(11): 1499-1504.
- [ 7 ] Xu Dongbin, Liu Changping, Huang Lei. Adaptive background model for motion detection based on statistic of probability [ J ]. Journal of Image and Graphics, 2008, 13(2): 351-358. [ 徐东彬, 刘昌平, 黄磊. 基于概率统计自适应背景模型的运动目标检测方法 [ J ]. 中国图象图形学报, 2008, 13(2): 351-358. ]
- [ 8 ] Gerald D, Joshua M, Grimson W E L. Background subtraction for temporally irregular dynamic textures [ C ] // Proceedings of IEEE Workshop on Applications of Computer Vision. Washington DC, USA: IEEE Computer Society, 2008: 1-7.
- [ 9 ] Sun J, Zhang W, Tang X, et al. Background cut [ C ] // Proceedings of IEEE Europ Conference on Computer Vision. Washington DC, USA: IEEE Computer Society, 2006: 628-641.
- [ 10 ] Jodoin P M, Mignotte M, Konrad J. Statistical background subtraction using spatial cues [ J ]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(12): 1758-1763.
- [ 11 ] Han J, Zhang M, Pan Z. Accurate foreground segmentation unaffected by cast shadows based on graph cut [ J ]. Journal of Computational Information Systems, 2008, 4(4): 1607-1612.
- [ 12 ] Zhang K, Tang M, James T K. Applying neighborhood consistency for fast clustering and kernel density estimation [ C ] // Proceedings of Computer Vision and Pattern Recognition. Washington DC, USA: IEEE Computer Society, 2007: 1001-1007.
- [ 13 ] Toyama K, Krumm J, Brumitt B, et al. Wallflower: principles and practice of background maintenance [ C ] // Proceedings of IEEE International Conference on Compute Vision. Washington DC, USA: IEEE Computer Society, 1999: 255-261.