

# 聚类差分图像核密度估计前景目标检测

胡 闽 刘纯平 崔志明 王朝晖 张书奎

(苏州大学计算机科学与技术学院, 苏州 215006)

**摘 要** 针对非参数核密度估计学习阶段信息冗余与重复计算,估计阶段的估计错误噪声和计算量大的问题,提出了一种基于聚类分析的差分图像核密度估计前景目标检测算法。该方法在非参数核密度估计的学习阶段基于最大最小聚类原理从原采样全样本中提取那些具有较高频率和多样性的小样本来包含尽可能多的关键样本信息,在估计阶段采用基于自适应阈值的图像差分滤去非典型的运动像素,再利用高斯核密度估计进行运动像素分类。实验结果表明该方法限制了非典型运动像素估计错误产生的噪声,并减少了核密度估计计算量,提高了算法的实时性。

**关键词** 核密度估计 聚类 差分图像 前景目标检测

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2009)10-2126-06

## Clustering Difference Image Kernel Density Estimation for Foreground Object Detection

HU Min, LIU Chun-ping, CUI Zhi-ming, WANG Chao-hui, ZHANG Shu-kui

(Department of Computer Science and Technology, Soochow University, Suzhou 215006)

**Abstract** For non-parametric kernel density estimation information redundancy and repetition computation in the training stage estimate error and large amount of calculation in the estimated phase, this paper proposed a method of clustering difference image kernel density estimation for foreground object detection. We first choose those samples that have higher frequency and diversity to contain important information based on max-min distance clustering in training sequence. A Gaussian KDE is built to estimate a motion object after adaptive threshold image difference calculation. Experimental results were given to demonstrate that the proposed algorithms are elimination of the typical non-movement noise point for estimated error and improving real-time capability.

**Keywords** kernel density estimation (KDE), clustering, difference image, foreground object detection

## 1 引 言

从视频序列中分离出前景目标,是计算机视觉中的重要研究内容,在视频监控、人机交互、视频会议等领域有着广泛的应用。背景减法是在摄像机静止条件下广泛应用的一种前景目标检测技术,是当

前计算机视觉领域的研究热点之一。

为此研究人员提出了不同的方法。Velastin 等人利用最近  $n$  帧的中值作为背景模型<sup>[1]</sup>,该方法在场景中存在多个运动目标或目标运动缓慢时,前景会污染背景;Ridder 等人对每个像素构建一个模型,每个像素有一个 Kalman 滤波,使得系统能适应场景中的光照变化,但是其背景自适应变化很缓慢,不能

基金项目:国家自然科学基金项目(60873116);江苏省自然科学基金项目(BK2009116, BK2009593);江苏省科技支撑计划项目(BE2009048)

收稿日期:2009-06-17;改回日期:2009-07-16

第一作者简介:胡 闽(1984 ~ ),男。现为苏州大学计算机科学与技术学院硕士研究生。主要研究方向为图像处理、模式识别。

E-mail:123234334@qq.com

通讯作者:刘纯平, E-mail:cpliu@suda.edu.cn

很好地处理多层背景的情况<sup>[2]</sup>;Colombari 等人把像素分布看作高斯分布,对背景建模<sup>[3]</sup>;Stauffer 等人在此基础上用  $k$  个高斯混合分布(GMM)来表示像素分布,解决了像素的多峰分布的问题,但是其需要事先假定背景分布模型及进行模型参数估计和优化<sup>[4-5]</sup>;Elgammal 等人提出了基于核密度估计的非参数背景模型<sup>[6-8]</sup>,该方法能够适应不同的场景,不同于 GMM,它充分利用最近的历史帧信息来表示背景模型,能够适应复杂的像素分布密度,克服像素值在短时间内频繁变化,因此能得到较准确的估计。这种方法属于惰性学习法,它在学习阶段几乎不进行计算只进行样本值统计,但是在估计阶段需要在统计的样本集基础上进行核密度估计,由于采样样本集中包含噪声且存在重复样本,导致在核密度估计时计算量大且容易导致估计错误。该方法对图像非典型运动像素也进行了核密度估计,由此不可避免地产生估计错误噪声。

本文提出了一种基于聚类采样的差分图像核密度估计前景目标检测方法,该方法在非参数估计的学习阶段基于最大最小聚类原理从原采样全样本中提取那些具有较高频度和多样性的小样本来包含尽可能多的关键样本信息,在估计阶段先利用自适应全局阈值图像差分粗略检测出场景的所有典型运动,然后对这些典型运动区域利用核密度估计进行运动背景抑制,从而实现前景目标的检测。该方法保留了 Elgammal 非参数背景模型<sup>[6]</sup>能处理复杂背景以及敏感度强的优点,且充分利用聚类思想克服原样本信息冗余和重复计算问题,并通过自适应阈值图像差分过滤非典型运动像素,减少了估计阶段非典型运动像素核密度估计错误造成的噪声影响,并使核密度估计计算简单有效。

## 2 非参数核密度估计运动检测

视频序列中每一像素取值可视为一个随机过程。由于场景的复杂性,背景密度分布不能简单地用一个单模态模型来完全描述,而且背景分布形式通常无法事先假设。采用完全由视频数据来得到背景模型的核估计方法<sup>[6]</sup>非常适用于密度分布未知的计算机视觉分析问题。

假设  $x_1, x_2, \dots, x_N$  为一图像序列中某对应像素点最近  $N$  个采样值,那么在  $t$  时刻观测到对应像素点的概率  $P(x_t)$  可以通过下式进行估计:

$$P(x_t) = \frac{1}{N} \sum_{i=1}^N K(x_t - x_i) \quad (1)$$

式中,  $K$  为核估计子,假设  $K$  服从正态分布。对于 RGB 图像可以取 R, G, B 分量作为特征值,如果它们相互独立,则  $N$  个样本的概率之和为

$$P(x_t) = \frac{1}{N} \sum_{i=1}^N \prod_{m=1}^d \frac{1}{\sqrt{2\pi}\sigma_{i,m}} e^{-\frac{(x_t^m - x_i^m)^2}{2\sigma_{i,m}^2}} \quad (2)$$

式中,  $d$  为像素的特征维数,  $x_t^m$  为像素点的第  $m$  个特征值,  $\sigma_{i,m}$  为第  $m$  个特征的核宽,可由像素相邻帧样本的绝对差中位数来计算<sup>[6]</sup>。式(2)得到的概率如果满足

$$P(x_t) < T_f \quad (3)$$

则像素点  $x_t$  为前景,否则被认为是背景点。其中  $T_f$  为整副图像的全局阈值。

## 3 聚类差分图像核密度估计运动检测

聚类采样差分图像核密度估计运动检测算法主要由 5 部分组成:训练样本集聚类采样、差分图像自适应阈值运动检测、核密度估计运动分类、前景对象分割、样本集与参考图像更新。整个算法的框图如图 1 所示。

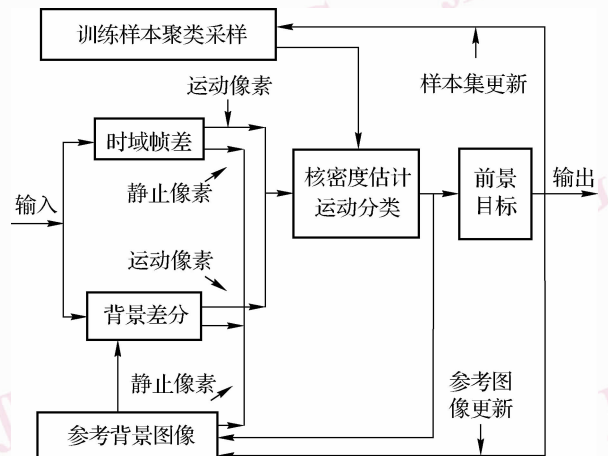


图 1 算法的框图

Fig. 1 Block diagram of algorithm

### 3.1 训练样本集聚类采样

Elgammal 提出的非参数核密度估计算法<sup>[6]</sup>在学习阶段对  $N$  帧图像建立全样本集。  $x_1, x_2, \dots, x_N$  为一图像序列中某对应像素点  $N$  个采样值全样本

集。这种方法属于惰性学习法,在学习阶段几乎不进行计算,只统计样本值。由于采样样本中存在噪声和重复样本导致在核密度估计时计算量大且容易导致估计错误。针对全样本信息冗余与重复计算问题,提出了采样样本集聚类机制将样本集空域多样性推广到时域多样性,从训练样本集中提取关键信息。

首先从全样本集中选取具有最大出现频率的值  $y_1$ ,然后在剩下的样本中选取与之在欧氏距离相差最远的值  $y_2$ ,再在剩余样本集中选择最大出现频率的值  $y_3$ ,然后在剩余样本中选取二者欧氏距离相差最远的值样本值  $y_4$ ,如此反复直到新样本数达到预先设定的值  $M$ ,最后整个样本集可由  $M$  个新样本  $y_1, y_2, y_3, \dots, y_M$  来表示,从而建立能代表全样本关键信息的小样本集。其中  $M$  的大小主要由全样本集频率的极值数决定。 $M$  越大用于模型建立的样本越多,模型越准确,但计算量越大。新样本中每个样本对密度估计的作用不同,由下式(4)计算新的权值

$$\alpha_i = \frac{N_i}{N} \quad i = 1, 2, \dots, M \quad (4)$$

式中,  $N_i$  是通过对聚类中心计算最小距离得到的原样本数。则对应像素点的概率可以通过下式进行估计:

$$P(x_i) = \sum_{i=1}^N \alpha_i \prod_{m=1}^d \frac{1}{\sqrt{2\pi}\sigma_{i,m}} e^{-\frac{(x_i^m - s_i^m)^2}{2\sigma_{i,m}^2}} \quad (5)$$

## 3.2 差分图像核密度估计

### 3.2.1 差分图像自适应阈值运动检测

时域帧差法能够适应环境的动态变化,实现实时运动分割,但是容易导致内部空洞,并对速度过快和过慢的运动也容易导致检测错误。而在参考背景图像获得较好的情况下,背景差分能得到比较完整的运动目标,但是不易获取准确的参考背景图像且难进行背景自适应更新。

根据时域帧差法和背景差分方法的特点,在这一阶段,不重要的非典型运动像素被通过简单的帧间差分 and 背景差分过滤<sup>[9]</sup>。  $I(s, t) = \{I_c(s, t)\}$  表示  $t$  时刻输入图像,  $B(s, t) = \{B_c(s, t)\}$  表示在  $t$  时刻维护的背景图像,  $c \in \{r, g, b\}$  表示颜色成分。帧间差分图像  $F_{fd}(s, t)$  和背景差分图像  $F_{bd}(s, t)$  通过自适应的基于噪声模型空间分布的全局阈值法<sup>[10]</sup>进行运动像素和静止像素分类。使用自适应阈值的目的是

是为了得到相对精确的检测结果。

### 3.2.2 核密度估计运动分类

在上一阶段差分图像自适应阈值运动检测得到的典型运动像素存在两种情况:运动像素属于运动前景,运动像素属于运动背景。这两种情况将被通过核密度估计进行区分。在这一阶段如果  $F_{fd}(s, t) = 1$  或  $F_{bd}(s, t) = 1$ , 像素  $I_i(s, t)$  将通过式(5)在聚类样本集的基础上计算对应的概率  $P(x_{i,i})$ , 如果其满足式(3)则  $I_i(s, t)$  判定为前景, 否则被认为是背景点。

### 3.2.3 前景对象分割

经过差分图像运动检测和核密度估计分类的识别,只有很小比例的背景点被错误地识别成前景点,而且更多的是被识别成孤立的前景点。形态学开闭操作被用来连接前景点以及消除这些孤立错误点,并使用消除小区域的方法来保留关键的兴趣区域。

针对差分图像使得前景目标的一些运动像素点在第一阶段就被判定为背景像素点导致前景目标不完整的问题,考虑前景目标噪声较少的情况,先利用形态学闭操作进行前景目标膨胀,再利用开操作腐蚀噪声点。分割的前景图像用  $O(s, t)$  来表示。

### 3.2.4 样本集与参考图像更新

采样样本集和参考背景图像都需要进行实时更新。采样样本集更新采用 short-term 模型策略<sup>[7]</sup>, 它采用了选择更新机制,只有分类为背景的样本才更新到样本集中。对当前被识别为背景的样本点,计算其特征值  $x_i$  与前一时间刻样本集  $\{y_1, y_2, \dots, y_m\}_{t-1}$  中每一样本的距离,如果  $x_i$  值在某一样本  $y_i$  的  $2\sigma_i$  内,那么当前  $x_i$  不用来更新样本模型,否则将  $x_i$  作为一个新的样本点加入到样本集模型中。

参考背景图像的更新考虑前景分割的结果,如果  $O(s, t) = 0$  并且  $F_{fd}(s, t) = 1$  或者  $F_{bd}(s, t) = 1$ , 这说明背景发生了较强改变,需要被新的背景特征更新:

$$B_c(s, t+1) = I_c(s, t) \quad c \in \{r, g, b\} \quad (6)$$

否则说明背景没有发生较强改变,可以通过下式进行背景更新:

$$B_c(s, t+1) = (1 - \alpha_1)B_c(s, t) + \alpha_1 I_c(s, t) \quad (7)$$

通过以上的更新,采样样本集和参考背景图像可以适应背景模型改变。

## 4 实验及结果分析

本文算法对多种复杂场景下的视频进行了实验,使用双核 Intel(R) Pentium(R) Dual CPU E2160 @ 1.80GHz,1GB 内存的 PC 机器。所有实验训练阶段采样 100 帧,参考图像样本更新率  $\alpha_1$  取 0.1。

### 4.1 全样本聚类采样 $M$ 样本值取值分析

对 Li 给的 Campus 视频序列<sup>[9]</sup>,以及 PETS (performance evaluation of tracking and surveillance) 2001 的 camera 1 和 camera 2 视频序列复杂场景中像素点分布进行了实验分析。图 2 是像素点分布分析实验中具有代表性的 Campus 视频序列<sup>[9]</sup>,图 2(b) 是该视频序列 1—200 帧图像的像素点 (60,50) 进行采样分析的灰度统计直方图。因为 Campus 视频序列中该像素点是摇动的树叶,且存在光照变化情况,因此该像素变化很剧烈,是复杂场景中像素点分布的典型代表。从图 2(b) 可以看出,该像素点灰度值变化非常剧烈,直方图中约有 40~60 个不同取值,且有 3 个明显的较大峰值,其大概峰值点在 10~15 个左右。

实验表明在一般复杂场景下,像素点样本分布一般在 30~40 左右,特别剧烈的如图 2 在 40~60 左右,其主要峰值一般约在 8~15。为了更多地表示全样本的多样性,所以在后面实验中进行全样本聚类时一般设  $M$  值为 20。

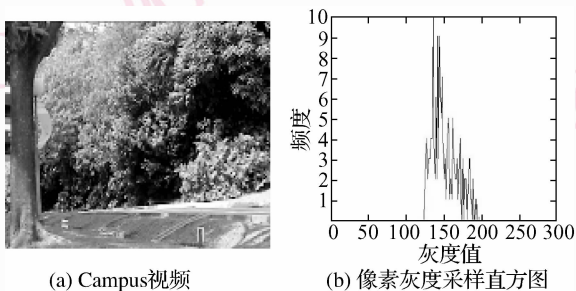


图 2 Campus 视频像素灰度采样直方图

Fig. 2 Histogram of pixel gray sampling for Campus video

### 4.2 实验结果比较

在 PETS2001 的两个视频序列上,将本文算法与 Elgammal 等人提出的非参数核密度估计算法<sup>[6]</sup>进行了实验对比。

长为 2 342 帧的图 3 camera 1 视频序列是校园

停车场的一段复杂场景,视频序列中背景存在摇动的树木,以及背景物体变成前景物体和前景物体变背景物体的情况。图 3 中的前景目标图像都没有进行形态学操作。

图 4 为长为 2 452 帧的 camera 2 视频序列。相对于 camera 1 视频序列,在具有相同背景变化情况的基础上 camera 2 中存在光照变化。

从图 3 中可以看出,对复杂场景 camera 1 来说,非参数核密度估计<sup>[6]</sup>虽然可以解决运动背景以及背景物体变前景和前景物体变背景等问题,但是其在学习阶段直接将视频序列像素值作为样本,而在实际应用中,学习时视频序列包含运动物体,这种做法势必将前景物体的取值也作为背景点计算进去,使得前景点包含在背景中导致误检,同时其对每个像素都进行了核密度估计计算,导致有些非典型运动像素由于上述原因导致了估计错误,产生了大量的噪声,如图 3、图 4 的原算法实验结果所示。并且由于是对所有像素进行估计,使得估计阶段计算量大,实时性不好。而本文算法通过差分图像滤去非典型运动像素,只对典型的运动像素进行核密度估计,且在训练阶段采用全样本集聚类采样得到具有代表性的小样本集,减少了全样本集中的噪声,使得在得到前景目标准确位置和大致轮廓的基础上,前景图像只有很少的估计错误噪声,并使得估计阶段计算量减少,提高了算法的实时性。但是本文算法因为差分图像的问题使得前景目标的一些运动像素点在差分阶段就被判定为背景像素点,导致前景目标不完整,如第 1 551 帧的后一个行人轮廓不清晰。需要在后期先进行形态学的 2 次闭操作再进行 1 次开操作来连接前景点。

从图 4 中可以看出,对复杂场景 camera 2 来说,非参数核密度估计较好的地方同 camera 1 实验分析。但是在光照变化的一段时间,因为非参数背景模型的训练样本集需要一段时间进行更新才能包含这种变化的像素点样本,这就导致在光照发生变化到样本集包含该改变化样本的一段时间内背景被误判定为前景,图 4 第 2 371 帧显示了该情况下的误检。而本文算法在差分阶段因为相邻帧间变化较少,通过差分处理减少了光照变化的影响,并及时更新了参考背景图像,使得得到的前景图像较准确。

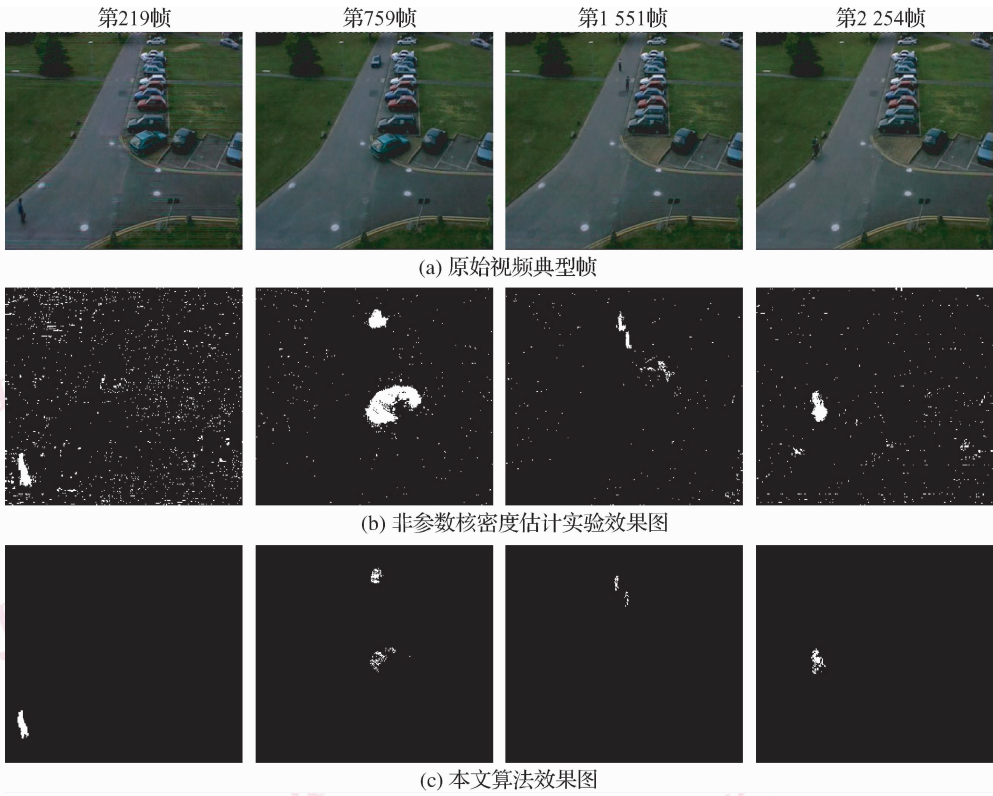


图 3 阴天室外复杂场景 camera 1 视频实验结果比较

Fig. 3 Result of on cloudy shy outdoor scenes video camera 1 viedo

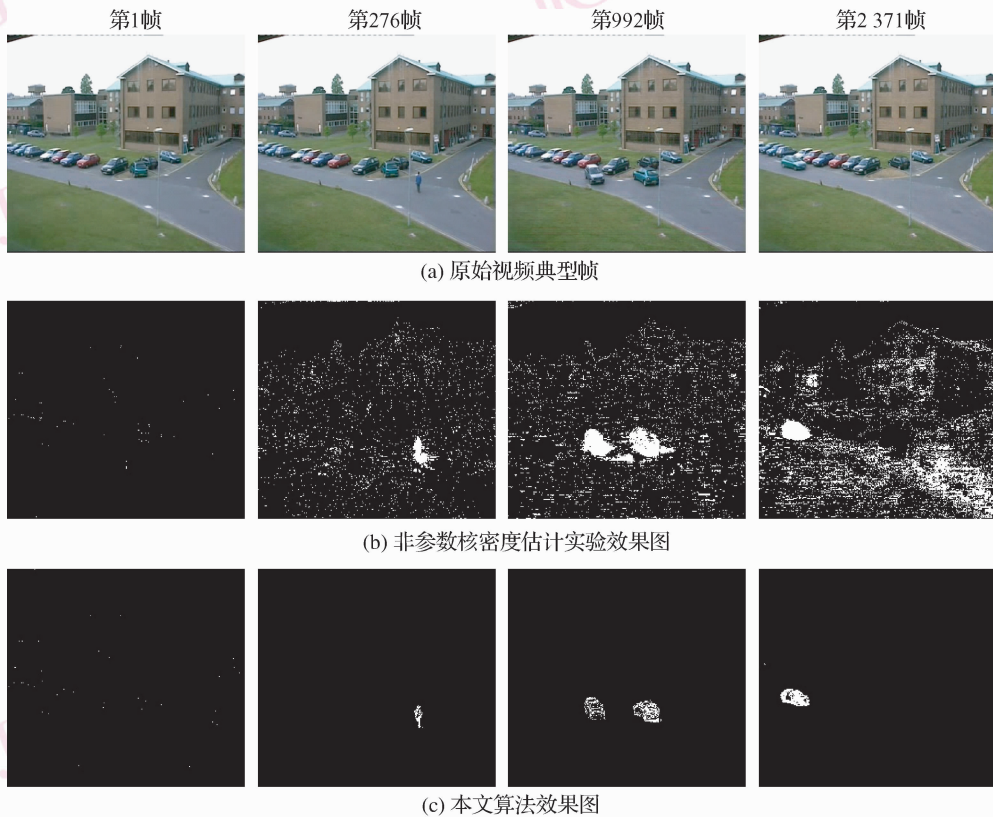


图 4 晴天光照变化的室外复杂场景 camera 2 视频实验结果比较

Fig. 4 Result of on sunny light changes outdoor scenes video camera 2 viedo

### 4.3 全样本核密度估计与聚类小样本核密度估计时间对比

通过对视频 camera 1 和 camera 2 彩色图像的核密度估计时间统计(表 1),全样本核密度估计的时间比较平均,一般都在 140 ms 左右,每秒大概能处理 7 帧左右。聚类差分图像核密度估计因为所要估的样本点不均匀,耗费时间不平均,一般在 70 ~ 80ms 之间,每秒大概能处理 12 帧左右。实验结果表明,经过自适应阈值差分图像过滤非典型运动像素和全样本聚类分析后,简化了估计阶段核密度的计算,使处理速度更加迅速,提高了算法的实时性。

表 1 核密度估计时间对比

Tab. 1 Kernel density estimation time compared

视频序列	当前帧	全样本(ms)	聚类样本(ms)
Camera1 (320 × 240)	第 219 帧	142	76
	第 759 帧	140	82
	第 1 551 帧	140	73
	第 2 254 帧	141	78
Camera2 (320 × 240)	第 276 帧	142	72
	第 2 371 帧	140	74

## 5 结 论

本文提出了一种基于聚类采样的差分图像核密度估计运动检测,在利用差分图像进行过滤非典型运动像素点的基础上,进行核密度估计区分运动前景和背景,并在非参数估计训练阶段基于最大最小聚类原理从原始全样本中提取那些具有较高频度和多样性的一个较小样本,使得估计阶段计算量和内存空间减少,并消除了非典型运动像素核密度估计错误造成的噪声。

同时,差分图像滤去非典型运动像素时阈值化结果的好坏将影响到运动检测的结果,而当前使用的阈值化方法是基于噪声模型空间分布的单通道全局阈值方法,没有考虑 RGB 3 个颜色通道之间的相

关性,由此导致前景目标像素点也有部分被滤去。如何使用更好的全局阈值法来进行差分图像检测运动是进一步深入研究的问题。

### 参考文献 (References)

- Lo B P L, Velastin S A. Automatic congestion detection system for underground platforms [A]. In: Proceedings of International Symposium on Intelligent Multimedia, Video, and Speech Processing [C], Hong Kong, China, 2001: 158-161.
- Ridder C, Munkelt O, Kirchner H. Adaptive background estimation and foreground detection using Kalman-filtering [A]. In: Proceedings of the Int'l Conference on Recent Advances Sinmechatronics [C], Istanbul, Turkey, 1995: 193-199.
- Colombari A, Fusiello A, Murino V. Segmentation and tracking of multiple video objects [J]. Pattern Recognition, 2007, 40(4): 1307-1317.
- Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking[A]. In: Proceedings of the Computer Society on Computer Vision and Pattern Recognition [C], FortCollins, USA, 1999:246-252.
- Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction[A]. In: Proceedings of the 17th International Conference on Pattern Recognition [C], Cambridge, United Kingdom, 2004: 28-31.
- Elgammal A M, Hanwood D, Davis L S. Non-parametric model for background subtraction [A]. In: Proceedings of the 6th European Conference on Computer Vision [C], Dublin, Ireland, 2000: 751-767.
- Mittal A, Paragios N. Motion-based background subtraction using adaptive kernel density estimation [A]. In: Proceedings of the Computer Society on Conference on Computer Vision and Pattern Recognition [C], Washington D C, USA, 2004:302-309.
- Xu Dong-bin, Huang Lei, Liu Chang-ping. Adaptive Kernel density estimation for motion detection[J]. Acta Automatica Sinica. 2009, 35(4):379-385. [徐东彬, 黄磊, 刘昌平. 自适应核密度估计运动检测方法[J]. 自动化学报. 2009, 35(4): 379-385.]
- Li L, Huang W, Gu I Y H, et al. Foreground object detection from videos containing complex background[A]. In: Proceedings of 11th ACM Multimedia Conference [C], Berkeley, USA, 2003:2-10.
- Rosin P. Thresholding for change detection [A]. In: Proceedings of IEEE Int'l Conference on Computer Vision [C], Bombay, India, 1998:274-279.