

自适应多模快速背景差算法

梁 华 刘云辉

(国防科学技术大学电子科学与工程学院, 长沙 410073)

摘 要 在多高斯模型的基础上, 从场景中模型分布不均匀性出发, 提出了一种新的快速背景差算法。该算法针对混合高斯模型中固定模型数量不足的问题, 建立了模型产生和退出的机制, 使模型数量能够自动适应场景特点, 实现了高斯模型的实时自适应分布, 即提高了准确性又有效地减少了模型的总量; 同时, 针对混合高斯模型中计算量大的问题, 对模型参数的计算进行了优化, 将耗时的浮点运算转化为整型运算, 减少了计算量; 算法中引入了生存时间和模型重现频率的概念, 通过对模型重现频率的限制有效抑制高频噪声。与混合高斯模型的实验结果对比说明, 该快速算法保持了原算法的优点, 执行速度提高 1 倍以上, 检测结果准确, 算法内存消耗小, 前景轮廓清晰, 抑制高频噪声的能力强, 整体效果优于混合高斯模型背景差算法。

关键词 视频监控 背景差算法 混合高斯模型 快速算法

中图分类号: TP391 文献标识码: A 文章编号: 1006-8961(2008)02-0345-06

Adaptive and Efficient Background Subtraction Using Multiple Models

LIANG Hua LU Yun-hui

(College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073)

Abstract This paper presents an efficient background subtraction algorithm using multiple scene models to cope with variations of noises in a background. A mechanism has been developed to add and delete scene models so that the distribution of the models is adaptive to the background characteristics. The calculation for the model parameters has been optimized so as to avoid time-consuming floating point calculation. We introduced the living time and recurrent frequency to the models so that the algorithm can suppress high frequency background noises effectively by controlling the model recurrent frequency. Experiments using video data have been conducted to compare the performance of our algorithm with that of the mixture Gaussian model algorithm. The experimental results demonstrated that our algorithm can extract the foreground contour more precisely, efficiently and with less memory, while maintaining the advantages of the mixture Gaussian model algorithm. It was also found that high frequency noises that cannot be rejected by the mixture Gaussian model can be suppressed.

Keywords VSAM (video surveillance and monitoring), background subtraction, mixture gaussian model, fast algorithm

1 引言

随着摄像设备价格下降、计算机计算能力增强、网络速度的提高, 大量视频监控设备已经在日常生活得到普及。视频监控相关技术的研究也逐渐成为计算机视觉研究领域的热点。

视频监控中遇到的首要问题往往是提取图像序

列中感兴趣的部分, 即前景, 例如移动的人、车辆等等, 也可以称为运动检测。从 20 世纪末至今, 这方面技术的研究相当广泛, 同时产生了大量的文献。通过对这些文献的分析, 可以看出, 针对固定场景的运动检测问题, 根据对前景区域特征定义的不同, 可以将这些方法大致分为两类: 一类是基于背景差的方法, 出发点是维持监控场景的背景模型, 判定当前图像中与背景差别较大的部分为前景; 另一类是基

基金项目: 国家自然科学基金项目 (6033401Q, 60475029)

收稿日期: 2007-06-20 改回日期: 2007-12-01

第一作者简介: 梁华 (1977~), 男, 国防科学技术大学信息与通信工程专业博士研究生。主要研究方向为视频监控相关技术。E-mail

lh2078@163.com

于光流场的方法,出发点是图像序列中具有明显运动的较大面积区域作为前景。近年来也有一些学者将形态学滤波应用于图像序列的时间、空间结构上^[1,2]。

基于背景差的方法,概念清晰,不需要费时的光流场计算,速度比较快,所以此类方法的研究比较多。但随着研究不断的深入,对算法的要求也在不断的提高。特别是对较复杂场景情况下的前景检测,例如,光照变化引起的背景改变(室外的环境光、室内的灯光等),高噪声的背景区域(树木、水面、风中的旗的反复运动等),背景的改变(车辆停止、背景中物体搬动等),前景物体阴影的去除问题等,都成为研究中必须考虑的问题。

较早期的背景差方法有均值、中值、运动平均、运动高斯模型等方法。从 20 世纪末至今,出现了一批比较成熟的背景差方法。1999 年,Stauffer 等人提出了基于混合高斯模型的背景差方法^[3,4],针对高噪声的复杂背景,采用多个高斯分布描述像素过程;2000 年,Elgammal 等人提出了基于无参数模型的背景差方法^[5],该方法利用核密度估计,产生比多高斯混合分布更准确的无参分布描述像素过程,Olivier 等人提出背景特征化建模的方法^[6],对较长时间内的背景图像数据进行主分量分析,利用主特征矢量张成的子空间描述背景,用当前图像与子空间的距离作为判决前景的依据;2004 年,Han 等人在无参数模型的基础上,提出利用序列核密度进行近似的方法^[7],在效果与无参估计基本相近的基础上提高了算法的速度。

在实际监控系统的研究中,背景差往往作为像素层的方法,和区域层的许多方法相结合。1999 年, W^4 系统^[8]中采用的是最大最小值双门限背景差方法;2000~2001 年佛罗里达中央大学的 KNIGHT 系统^[9~11]中,将混合高斯模型检测出的结果与梯度检测结果结合,对剔除背景中由于物体移动和光照造成的错误前景以及前景中的阴影有良好的效果;2003 年,意大利摩德纳大学的 SAKBOT 系统^[12,13]中,提出了基于知识的方法,在中值背景差方法的基础上,通过对区域的面积、显著性和区域运动特征的分析,将前景中的阴影、鬼影以及鬼阴影区分出来。2004 年,IBM 的 Smart 系统中提出了显著性运动检测的方法^[14],对时间差分图像中较明显的区域计算光流场,将运动较明显的区域作为前景。

从背景差方法的发展可以看出,有两个主要方

向:一是对像素过程的建模不断完善,并不断发展快速算法;另一个是在基本背景差算法的基础上结合其他信息,提高检测的鲁棒性。由于算法复杂度不断增加,而实际视频监控系统对实时性又不断提高,所以一个快速的背景差算法往往是决定监控系统性能的关键。

经过分析发现,视频监控场景中多数区域的背景模型用单高斯模型近似是准确的,只有少数区域需要用多个高斯模型近似,而且模型的数量往往需要根据场景的变化而改变。所以本文提出了自适应多模快速背景差算法。

2 自适应多模快速背景差算法

对于视频监控中固定的场景,如果场景中的像素来自于固定光照条件下的一个物体表面时(即静态的背景),用一个高斯模型描述该像素过程以及图像噪声是比较合理的。而在实际监控的场景中,往往某些区域会处于多个物体表面、边缘交替变化的情况,针对这些区域,需要采用多个高斯模型近似描述。当场景的光线缓慢变化时,用参数能够自适应的高斯模型描述是合理的,但是当光线快速变化或者场景快速改变时,参数来不及适应快速变化,这时就需要删除旧模型,产生新的模型来适应场景变化。

图 1 是在不同图像序列中固定某一位置像素观测其在一段时间内蓝绿分量分布的状况,其中横轴为蓝色分量,纵轴为绿色分量。由图 1(a)可见,坐标(200, 20)处的地面和(150, 220)处的天空的像素颜色变化较小,分布用单模近似比较准确,而(150, 120)处水面的像素颜色变化较大,需要用多模近似。图 1(b)中草地(150, 30)和路面(130, 80)中像素的分布用单模近似准确,而树叶区域的像素需要用多模近似。图 1(c)中室内灯光不变时墙壁(280, 180)上像素的分布用单模近似准确。图 1(d)中关灯前后墙壁(280, 180)中像素的分布变化剧烈,必须用不同的模型进行近似。

可见,像素颜色分布在同一场景中存在不同的情况,有些区域只需要用一个高斯模型就可以准确的近似,而有些区域如图中的水面、树叶等区域需要用多个高斯模型描述。在室内场景中由于灯光变化,像素的分布发生了快速的变化,变化前后需要用不同的模型描述。针对这种情况,可对每个像素建立多个数量不定的高斯模型,并对高斯模型建立产

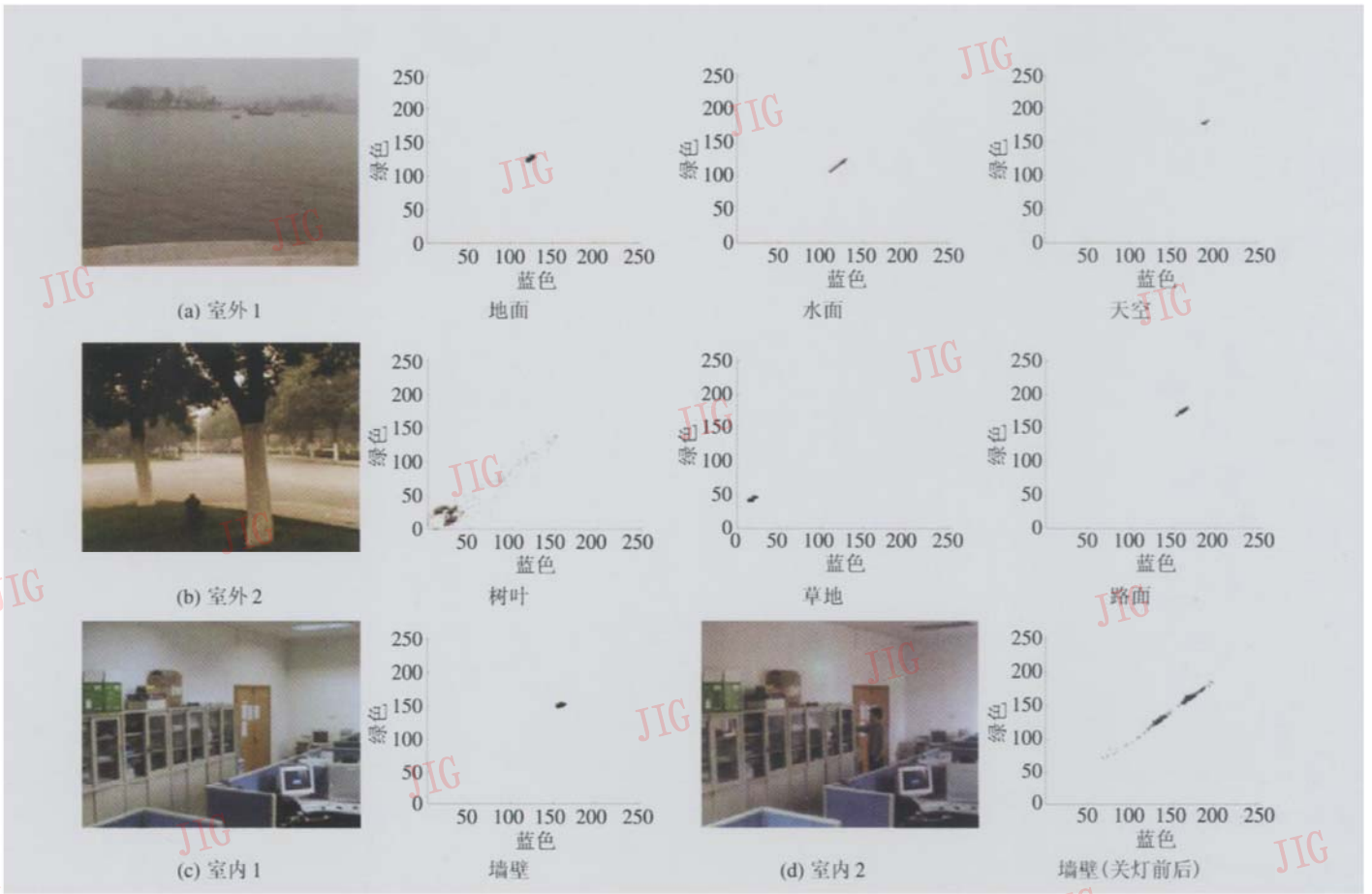


图 1 图像序列中像素颜色分布

Fig 1 Distribution of pixel process in image sequences

生和退出的机制, 让模型的数量和参数能够实时变化, 自动适应场景的不同特点。这样即能够准确描述复杂的像素过程, 又可以减少模型的总量。

设图像为 I 像素为 x , 时间为 t 每个像素 x 的过程用 N 个高斯模型 M 近似:

$$\{M_1, M_2, \dots, M_n, \dots, M_N\} \quad 1 \leq n \leq N \quad (1)$$

设每个高斯模型的均值为 μ 权值为 w , 生存时间为 l 重现频率为 f 。

首先用前若干帧图像 (一般取 255 单位) 对模型进行初始化。初始化时用第 1 帧图像为每个像素建立一个高斯模型, 模型参数 μ 设置为像素的值 (灰度或色彩), w 为 1, l 为 255 单位, f 为 0。后续图像的像素与这个模型进行匹配。像素 x 在 t 时刻与第 n 个模型的匹配定义为

$$|I_t(x) - \mu_n(x)| \leq Th_t \quad (2)$$

其中, $I_t(x)$ 为图像 I 在 t 时刻像素 x 处的像素值, $\mu_n(x)$ 表示 t 时刻像素 x 处第 n 个模型的均值, Th_t 为 t 时刻模型匹配的门限, 由下式决定:

$$Th_t = (\max(I_t) - \min(I_t)) / DIV \quad (3)$$

模型匹配的门限与当前图像中色彩变化的幅度

成正比, 其中 $\max(I_t)$ 和 $\min(I_t)$ 分别是 t 时刻图像 I 中像素值的最大值和最小值, DIV 是比例系数, DIV 值越大, 门限越低, 模型的分辨率越高, 当然也会产生较多的模型。由于监控场景的图像受光照的变化, 摄像机曝光强度、自动增益控制、白平衡等因素的影响, 整体的亮度和色彩往往会发生变化, 如果采用相同的门限, 在图像较暗时检测的分辨率会下降。采用随时间变化的门限值, 每间隔一段时间进行更新, 可以保证在不同情况下都有较好的分辨率, 所以, 这样设置门限值是合理的。

如果 t 时刻像素 x 与第 n 个模型匹配, 则模型参数更新如下:

$$w_{t+1n}(x) = w_{tn}(x) + 1 \quad (4)$$

$$\mu_{t+1n}(x) = \mu_{tn}(x) + \text{sgn}(I_t(x) - \mu_t) \quad (5)$$

$$\begin{cases} f_{t+1n}(x) = f_{tn}(x) + 1 & l_{tn}(x) \neq 255 \\ f_{t+1n}(x) = f_{tn}(x) & l_{tn}(x) = 255 \end{cases} \quad (6)$$

$$l_{t+1n}(x) = 255 \quad (7)$$

对于不匹配的模型, 参数更新如下:

$$w_{t+1n}(x) = w_{tn}(x) \quad (8)$$

$$\mu_{t+1n}(x) = \mu_{tn}(x) \quad (9)$$

$$f_{t+1n}(x) = f_{tn}(x) \quad (10)$$

$$l_{t+1n}(x) = l_{tn}(x) - 1 \quad (11)$$

其中, $w_{tn}(x)$, $\mu_{tn}(x)$, $f_{tn}(x)$, $l_{tn}(x)$ 分别为 t 时刻像素 x 处第 n 个模型的权值、均值、重现频率和生存时间, $\text{sgn}(x)$ 是符号函数。如果像素没有找到匹配的模型, 那么就产生新的模型, 新模型的初始化设置同第 1 帧。

在初始化过程完成后, 算法加入模型退出的机制。当模型的权值 w 与生存时间 l 都为 0 时将该模型删除。模型的匹配和建立以及参数的更新与式 (1)~式 (11) 基本相同, 只是对权值的更新有所变化, 如果匹配, 则

$$w_{t+1n}(x) = w_{tn}(x) + N_t(x) - 1 \quad (12)$$

如果不匹配, 则

$$w_{t+1n}(x) = w_{tn}(x) - 1 \quad (13)$$

式中, $N_t(x)$ 是 t 时刻像素 x 处模型的数量。

模型的产生、匹配、更新、退出的过程中, 像素 x 的模型权值的总和不变, 反复匹配的模型的权值不断增大, 而没有匹配的模型的权值不断减小, 权值的分配不断向匹配次数较多也就是更像背景的背景模型集中。当模型匹配时, 模型的均值随前景像素值发生改变, 能够使模型适应背景的缓慢变化。模型没有匹配时生存时间倒数计时, 匹配后重新开始计时。同时, 重现频率在模型生存时间内累计模型重复匹配的次数。只有当模型权值和生存时间均为 0 时才退出。因为, 有些背景模型存在时间短, 但是会反复出现 (例如树叶摆动反复覆盖的位置), 导致权值很小但是重现频率比较高, 所以, 不能仅根据权值判断模型是否退出。这样设计在一定程度上使背景模型能够容纳某些混合高斯模型无法容纳的高频噪声, 从而能够对这些噪声产生有效的抑制。

选择权值比较大或者重现频率比较高的模型的均值表示背景。如果像素没有找到匹配的模型, 或者虽然匹配但是匹配的模型权值较小且重现频率比较低时, 判决为前景, 模型权值和重现频率门限设置如下:

$$\begin{aligned} w_{tn}(x) &< 255N_t(x) \\ f_{tn}(x) &< F \end{aligned} \quad (14)$$

式中, 重现频率门限 F 需要根据系统的帧速率、模型生存时间长度和噪声的频率确定。

3 实验结果与分析

从 NUDT 视频和 PETS2001 视频中选取了 4 段

视频数据做实验对比。实验计算机 CPU 为 Celeron M 1.5GHz Windows XP 操作系统, 图像、视频处理平台为 Direct9.0 和 Intel OpenCV。实验中, 混合高斯模型的参数^[3]设置如下: 像素模型个数为 3 模型权值的学习速率 α 取 0.005 单位, 均值的学习因子 ρ 为 0.01, 初始化门限 σ 为 7 (等效的匹配判决门限为 $2.5\sigma = 17.5$), 背景模型权值比例设置为 0.6。自适应多模算法中判决门限系数 DIV 初始化为 13 (根据实验中场景的颜色幅度范围, 等效的匹配判决门限 Th_i 约为 17.5), 重现频率门限 F 设置为 12。这样的参数设置使二者的分辨能力和模型均值的变化速度基本相当, 实验结果对比有可信度。由于自适应多模算法中涉及模型的添加和删除, 所以, 算法实现中采用了结构链表的数据结构方式。对检测出的前景像素进行形态学开运算, 消除噪声。

实验结果如图 2 所示, 分别为混合高斯模型算法和自适应多模算法的实验结果、耗时以及后者模型数量分布图和模型总量统计。图 2 中, (a)、(b) 为 NUDT 校园视频数据的实验结果, (c)、(d) 为 PETS2001 DATASET1 视频数据的实验结果。图 2(a) 中两种算法对湖面背景的噪声抑制都比较好, 但后者的前景区域更准确, 误报率低。图 2(b) 中前者结果中对树叶摇摆的区域出现误检, 后者能够有效的抑制。图 2(c) 中后者模型分布主要集中在行人、车辆密集的区域。图 2(d) 中场景光线较暗, 汽车灯光扫过的区域出现模型数量的显著变化。

从场景中行人和车辆的检测效果来看, 自适应多模算法检测出的前景区域外形更清晰, 背景噪声相对较少。这说明该算法识别率高, 误检率低, 对场景中的噪声抑制能力强。原因来自两个方面, 一方面是由于混合高斯模型的模型数量人工设定, 模型冗余时会造成漏检, 而模型过少时会造成误报。而自适应多模算法中模型数量能适应场景特点变化, 对场景的描述更准确; 另一方面由于自适应多模算法中引入了模型重现频率的判决, 这样算法对高斯模型无法准确描述的高频噪声 (如树叶摆动) 也有很好的抑制作用。

从算法的执行速度来看, 在一般的场景复杂度情况下, 自适应多模算法的执行速度远远高于混合高斯模型算法, 原因也来自两个方面: 一方面是由于自适应多模算法的模型数量比较少; 两种算法的速度很大程度上取决于背景模型中模型总量, 混合高

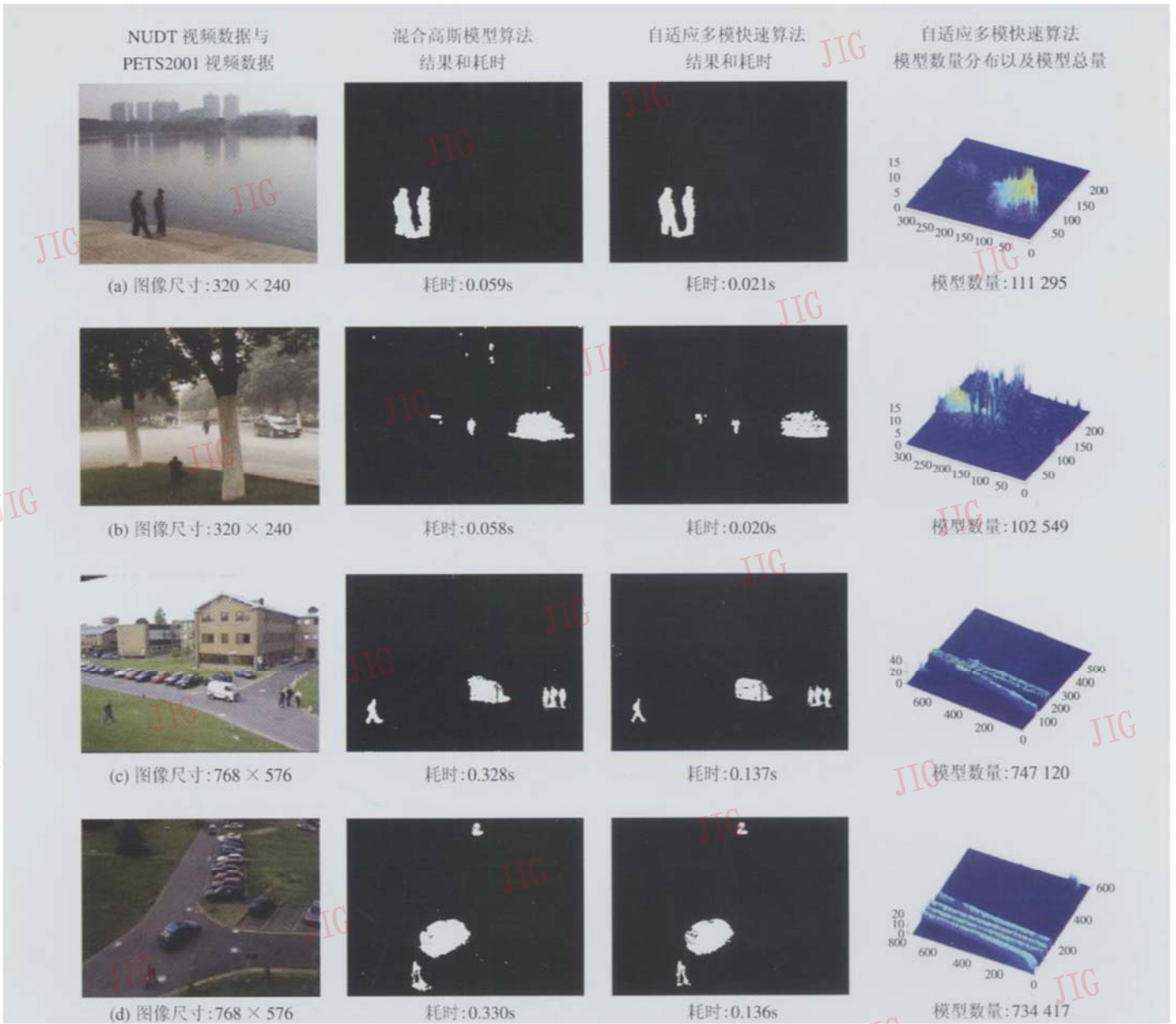


图 2 实验结果

Fig 2 Experiment results

斯模型算法的模型总量设定后执行过程中不改变 (如果像素的模型个数设为 3 那么 320×240 图像的模型总量为 230 400, 768×576 图像的为 1 327 104), 而自适应多模算法的模型数量随场景的变化而变化, 从对图像帧 (一般复杂度的场景, 背景较复杂, 前景面积占一定比例的图像帧) 的模型统计数据可以看出, 模型数量比前者减少一半左右 (4 幅实验图像中模型总量分别为前者的 48%、46%、56%、55%), 由于模型数量小, 所以, 算法内存消耗也小于前者; 另一方面是由于自适应多模算法中对数值运算做了优化, 将混合高斯模型中耗时的浮点数乘除运算和排序运算转化成了整型的加减运算和比较运算, 即使在模型数量相当的情况下, 自适应多模算法的速度也能高出混合高斯模型 20% 以上。

4 结 论

背景差算法是视频监控系统中的关键技术, 而算法的速度往往决定算法的实用价值。通过对相关文献和实验数据的分析可以看出, 当前的主流算法在普通 PC 机计算能力的条件下, 速度仍然较慢, 距离在实际系统中的应用还具有一定的差距。

本文在像素过程多高斯模型描述的基础上, 从场景中模型分布不均匀性出发, 提出了一种自适应多模快速背景差算法。该算法针对混合高斯模型中固定模型数量的不足, 建立了模型产生和退出的机制, 使模型数量能够自动适应场景特点, 实现了模型的实时自适应分布, 有效地减少了模型的数量; 针对

混合高斯模型中计算量大的问题,对模型参数计算进行优化,将耗时的浮点数乘除运算和排序运算转化成了整型数的加减运算和比较运算,并运用结构链表的数据结构方式,减少了计算量;针对多高斯模型对高频噪声描述的局限性,模型中引入了生存时间和模型重现频率,通过对模型重复出现频率的控制,有效抑制了混合高斯模型无法准确描述的高频噪声。通过与混合高斯模型算法实验比较说明,该快速算法内存消耗小,计算速度快,总体耗时在原算法的一半以下,检测结果准确,抑制高频噪声的能力强,整体效果优于混合高斯模型算法。由于该算法中没有浮点运算,所以该算法对某些浮点运算能力较弱的小型系统环境适应能力更强。

混合高斯模型算法在当前背景差算法中就其准确性而言并不是最好的,但其概念清晰、实现简单,在综合内存消耗、计算耗时等性能后,仍不失为一种优秀的算法,并且在许多系统中得到广泛的应用。所以与该算法的实验结果对比是有意义的。

参考文献 (References)

- 1 Manzanera A, Richefeu J. A robust and computationally efficient motion detection algorithm based on Σ - Δ background estimation[A]. In Proceedings of Indian Conference on Computer Vision, Graphics and Image Processing[C], Kolkata, India, 2004: 46~51.
- 2 Richefeu J, Manzanera A. A new hybrid differential Filter for motion detection[A]. In Proceedings of International Conference on Computer Vision and Graphics[C], Warsaw, Poland, 2004: 22~24.
- 3 Stauffer C, Grimson E. Learning patterns of activity using real-time tracking[J]. IEEE Transactions on Pattern Recognition and Machine Intelligence, 2000, 22(8): 747~757.
- 4 Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking[J]. Computer Vision and Pattern Recognition, 1999, 2: 246~252.
- 5 Elgammal A, Hawwood D, Davis L. Non-parametric model for background subtraction[A]. In Proceedings of the 7th IEEE International Conference on Computer Vision Frame-Rate Workshop[C], Kerkyra, Greece, 1999: 246~252.
- 6 Oliver N M, Rosario B, Pentland A P. A bayesian computer vision system for modeling human interactions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22: 831~843.
- 7 Haritaoglu I, Hawwood D, Davis L S W. Real-time surveillance of people and their activities[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 809~830.
- 8 Javed O, Shah M. KNIGHT^M: A Real Time Surveillance System[OL]. <http://www.cs.ucf.edu/~vision/papers>
- 9 Javed O, Shafique K, Shah M. A hierarchical approach to robust background subtraction using color and gradient information[A]. In Proceedings of IEEE Workshop on Motion and Video Computing[C], Washington, DC, USA, 2002: 22~27.
- 10 Javed O, Shah M. Tracking and object classification for automated surveillance[A]. In Proceedings of the Seventh European Conference on Computer Vision[C], Copenhagen, Denmark, 2002: 343~357.
- 11 Javed O, Rasheed Z, Alatas O, et al. KNIGHT^M: A real time surveillance system for multiple overlapping and non-overlapping cameras[A]. In Proceedings of IEEE International Conference on Multimedia and Expo[C], Baltimore, Maryland, USA, 2003.
- 12 Cucchiara R, Grana C, Piccardi M, et al. Detecting moving objects, ghosts and shadows in video streams[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(10): 1337~1342.
- 13 Cucchiara R, Grana C, Piccardi M, et al. Detecting objects, shadows and ghosts in video streams by exploiting color and motion information[A]. In Proceedings of IEEE International Conference on Image Analysis and Processing[C], Palermo, Italy, 2001: 360~365.
- 14 Tian Y L, Lu M, Hampapur A. Robust and efficient foreground analysis for real-time video surveillance[A]. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C], San Diego, CA, USA, 2005: 1182~1187.