

基于改进高斯混合建模和短时稳定度的运动目标检测算法

张超^① 吴小培^{*①} 周建英^① 戚培庆^① 王莹冠^② 吕钊^①

^①(安徽大学计算智能与信号处理教育部重点实验室 合肥 230039)

^②(中国科学院上海微系统与信息技术研究所 上海 200050)

摘要: 传统高斯混合建模方法中前景模型转换为背景模型的风险随着模型权值在一定更新率下的迭代累加而增大,使得传统方法难以有效检测低速运动目标。该文对高斯混合建模中背景匹配失败时生成的前景模型加以利用并引入短时稳定度指标进行综合判断,在深入挖掘前景模型中包含的运动目标信息和像素点级稳定性的基础上实时区分各像素点的状态。多场景下的实验结果表明,该方法对低速目标达到了更高的检出率。

关键词: 运动目标检测; 高斯混合建模; 前景模型; 短时稳定度

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2012)10-2402-07

DOI: 10.3724/SP.J.1146.2012.00158

A Moving Object Detection Algorithm Based on Improved GMM and Short-term Stability Measure

Zhang Chao^① Wu Xiao-pei^① Zhou Jian-ying^① Qi Pei-qing^① Wang Ying-guan^② Lü Zhao^①

^①(Key Lab of Intelligent Computing and Signal Processing Anhui University, Hefei 230039, China)

^②(Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China)

Abstract: In traditional Gaussian Mixture Modeling (GMM) algorithm, the risk that foreground model changes into background model rises with the cumulating of model weight under certain learning rate. That makes the algorithm unable to deal with slow moving object. This paper proposes an algorithm which takes advantage of the foreground models and employs an index of short-term stability measure to make a compound judgment. Each pixel status is decided real-timely considering the information of moving objects contained in foreground models and the pixel-level stability status. The results from different experiments verify that the proposed algorithm achieves a higher detection rate in detecting slow moving objects.

Key words: Moving object detection; Gaussian Mixture Modeling (GMM); Foreground model; Short-term stability measure

1 引言

当今信息化社会的各领域对智能视频系统均有着广泛而紧迫的应用需求,运动目标检测正是智能视频系统中各种智能分析处理的基础。目前,常用的运动目标检测算法主要有背景差分法^[1]、帧间差分法^[2]和光流法^[3]。其中,光流法计算较复杂,一般用于背景变化的场景。帧间差分法实时性强,对于动态环境具有较好的适应性,但其提取的目标前景内部易出现空洞。背景差分法可以得到相对完整的目标,且算法复杂度较低,但现实场景往往存在光照或背景的变化,这种情况下背景差分法的效果同样不佳。

针对现实世界中极为常见的多模态背景特性,文献[4]提出了高斯混合建模(Gaussian Mixture Modeling, GMM)方法。该方法在每个像素位置使用多个高斯模型进行建模,并且利用时间序列上的像素值不断进行模型参数更新,以期克服背景和光照变化等因素给运动目标检测带来的不利影响。文献[5]证明了高斯混合建模较其他方法具有的性能优势和应用可行性。

尽管比其他方法更能有效应对多模态场景,高斯混合建模同样有其局限^[6],如初始阶段算法收敛性能不佳、对低速目标以及复杂背景条件下的目标检测效果变差等。国内外学者一直在寻求对高斯混合建模进行有针对性地改进,文献[7]提出考察空间依存关系并对标准差的更新进行限制,文献[8]中使用色彩和空间一致性准则辅助传统GMM进行判断,文献[9]提出了解决建模时瞬间光线变化的方法,文献

2012-02-22 收到, 2012-07-02 改回

安徽省科技攻关强警专项(1101b0403030)和中国科学院上海微系统与信息技术横向研发基金课题资助

*通信作者: 吴小培 IIPHCI_AHU@163.com

[10]中专门研究了复杂背景下的建模问题，文献[11]在高斯混合建模的思路进一步提出泛化的混合对称稳定模型。

本文针对传统高斯混合建模方法不适于检测低速目标的局限，在对传统GMM算法进行一定程度的分析理解的基础上提出一种改进的运动目标检测方法，该方法对一直以来处于辅助地位的背景模型匹配失败时新生成的前景模型加以利用并引入短时稳定度指标进行综合判断，通过考察前景模型中包含的运动目标信息和像素点级稳定性来克服传统高斯混合建模方法检测低速运动目标时易产生的前景破碎问题。

2 高斯混合建模

高斯混合建模中的建模对象是图像序列，对应空域每个像素位置可视为一随机变量 $X=\{x_1, x_2, \dots, x_t\}$ ，该随机变量由具有 K 个单高斯分布的高斯混合模型表示，每个单高斯模型可以表征当前像素点的不同状态， K 视情况一般取为3到7。 $w_k=P(k)$ 为当前像素点和第 k 个状态匹配的先验概率。 $\theta_k=(\mu_k, \sigma_k)$ 为第 k 个高斯分布的参数，总的参数表示为 $\phi = \{\mu_1, \mu_2, \dots, \mu_K; \sigma_1, \sigma_2, \dots, \sigma_K\}$ 。于是，在随机变量 X 当前取值 x 下，第 k 状态对应的

$$f_X(x | \phi) = \sum_{k=1}^K P(k) f_{X|k}(x | k, \theta_k) = \sum_{k=1}^K w_k \frac{1}{(2\pi\sigma_k^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma_k^2}(x - \mu_k)^2\right\} \quad (1)$$

算法的目的是寻求最大的状态 \hat{k} ，使得

$$\hat{k} = \arg \max(w_k f_{X|k}(x | k, \theta_k)) \quad (2)$$

图1形象地给出了高斯混合建模的建模思路，图1(a)所示为连续视频帧，图1(b)代表每帧图像对应的数据矩阵。建模过程中视频流的每帧图像依时序不断更新空域各个位置上的混合高斯模型，混合高斯建模的结果是在每个像素点位置上建立一个由多个单高斯 $\eta(\theta_{ijk})$ 的加权和表示的背景模型，如图1(c)所示，此背景模型的作用是衡量当前新进数据和已有背景模型的匹配程度来确定当前像素值是否属于背景并以此检测出运动目标。图1(d)则给出了由3个单高斯混合得到的高斯混合模型示意图。

建模过程中， t 时刻新的像素值使用式(3)与 K 个模型逐一匹配。

$$|x_t - \mu_{k,t-1}| < \lambda\sigma_{k,t-1}^2 \quad (3)$$

若与当前模型的均值的差异在某一范围之内，则认为匹配，否则为不匹配。匹配的情况下使用下列式(4)~式(6)更新模型参数^[4]：

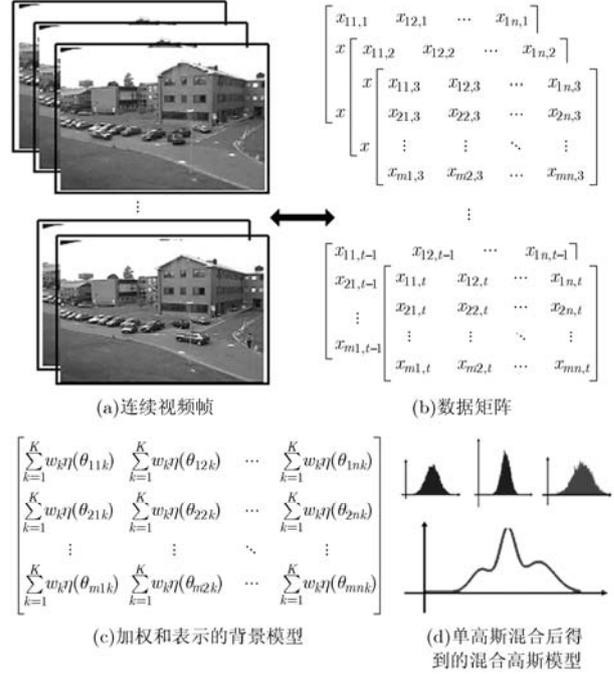


图1 高斯混合建模的形象表示

$$w_{k,t} = (1 - \alpha)w_{k,t-1} + \alpha(M_{k,t}) \quad (4)$$

$$\mu_{k,t} = (1 - \rho)\mu_{k,t-1} + \rho(x_t) \quad (5)$$

$$\sigma_{k,t}^2 = (1 - \rho)\sigma_{k,t-1}^2 + \rho(x_t - \mu_{k,t})^2 \quad (6)$$

其中 α 为学习率，且有

$$M_{k,t} = \begin{cases} 1, & \text{匹配成功} \\ 0, & \text{匹配失败} \end{cases} \quad (7)$$

$$\rho = \alpha\eta(x_t | \mu_{k,t-1}, \sigma_{k,t-1}) \quad (8)$$

不匹配时，应减小权值，且不对均值方差进行更新。若每个模型都不匹配，则要把权值最小的模型使用新的模型取代，其均值为新的样本像素值，方差为一大值，权值为一小值。每次权值更新后，都要对权值进行归一化处理。作为运动目标检测的最后步骤，建模完成后，要实时区分分别代表前景和背景的背景模型。文献[4]中的方法是计算每个模型对应的 $w_{k,t}/\sigma_{k,t}$ ，然后按照降序排列，其依据是当前最为匹配的模型必然具有最大的权值并具有较小的方差，最后，选定满足式(9)

$$B = \arg \min\left(\sum_1^b w_k > T\right) \quad (9)$$

的前 B 个模型作为背景模型，其中 T 为一权值累加门限。

3 一种改进的高斯混合建模方法

前景模型在传统方法中一直居于辅助地位，但事实上其包含了重要的目标信息。为了利用GMM良

好的环境适应性并解决上述传统GMM方法的固有局限, 本文对传统高斯混合建模方法进行改进, 对背景匹配失败时生成的前景模型加以利用并引入短时稳定度指标对前景进行综合判断。算法先按照传统算法的流程, 在当前像素值和对应 K 个背景模型都匹配失败后, 使用当前像素值和较大值替代权值最小模型的均值和方差, 这实际上是生成前景模型 $\eta(\mu_f, \sigma_f)$, 若后来点和前景模型均值相差小于门限 T_f , 则使用式(10)计算短时稳定度并更新前景模型:

$$S = \frac{M \sum_{j=0}^{M-1} x_{t+j}^2 - \left(\sum_{j=0}^{M-1} x_{t+j} \right)^2}{M(M-1)} \quad (10)$$

$$\mu_{f,t+1} = (1 - a_f)\mu_{f,t} + a_f x_{t+1} \quad (11)$$

其中, M 为滑动窗帧数范围, $a_f \in [0, 1]$ 为前景模型的学习率。算法优先使用前景模型对当前像素点进行匹配, 直接降低了前景点和背景模型误匹配的决策风险。

视目标外表面状况, 对外表面颜色较统一的目标, 其短时稳定度计算窗长 M 值可选为 $2 \sim 5$ 之间, 而对于外表色彩丰富多样的运动目标, 为了防止短暂的像素值变化导致目标向背景的转换, 其 M 值可选为 $5 \sim 20$ 之间, 较大的 M 能带来较好的检测效果但指标响应速度较为迟缓。求得稳定度后, 判断门限 S_{th} 可依图像序列实际情况设为经验值也可由 M 帧内稳定度的最大最小值动态求得

$$S_{th} = S_{MIN} + \left(\frac{S_{MAX} - S_{MIN}}{C} \right) \quad (12)$$

其中 C 为一常数。 S_{MIN} 和 S_{MAX} 分别为当前 C 帧内的稳定度最大值和最小值。在连续 C 帧超越短时稳定度门限的情况下, 当前像素点判为前景点。

稳定度的提法也见于文献[12]中使用当前时刻以前的像素值生成像素短时图(transience map)。本文中取当前时刻后的 M 点计算短时稳定度 S , 若 S 在一定范围内连续超越某一门限则说明此时该像素位置仍处于前景过程中。此方法的根本出发点是, 前景初次建模后, 若后来点未产生类似回到背景值的剧烈波动, 则认为当前像素位置一直保持某种稳定的前景状态, 遂在该点位置跳过传统高斯建模方法的后续流程直接判断为前景点。若当前点未能和前景模型匹配或短时稳定度指标不满足条件, 仍然进入传统方法的一般后处理流程。

图2为某视频图像序列某像素位置上的灰度值和该点稳定度的计算结果, 图中箭头表示上部的视频帧和下部像素值以及稳定度指标的对应关系, 空域中选取点的位置见视频帧黑圈所示, 图的底部为选取的视频帧的序号。从图2可见, 稳定度指标(柱

状图)从接近于零升高最后回落到零附近, 其变化超前于像素值的变化(折线图), 稳定度的上扬和回落均与像素值的变化相对应, 因而可以表征前景的出现和持续的状态。稳定度的使用能有效避免运动目标区域像素值短时变化导致前景误检为背景, 前景模型和短时稳定度指标相配合使本文的方法既具有高斯混合建模的环境适应性又能克服传统高斯建模无法检测慢速目标的局限。

4 实验及结果分析

为了验证本文方法的有效性, 我们使用3段视频在matlab平台上开展仿真实验。实验中将本文所提方法在相同条件下与传统高斯混合建模方法进行对比, 为了说明基于GMM类方法较其他方法的性能优势, 我们同时和基于自适应背景减的运动目标检测算法进行对比。实验使用的参数设置如表1所示。实验中高斯混合建模时将使用两种学习率以解决算法初始化时收敛慢的问题, 初始阶段学习率设置为1.5, 稳定阶段设置为0.05。

表1 仿真实验参数设置

高斯混合数 K	3	模型匹配权值 λ	2.5
权值和门限 T	0.7	前景模型学习率	0.01
初始化权值 w_i	0.05	延时帧数 M	2
自适应背景减门限	25	自适应背景减更新率	0.001

第1个对比试验采用一段含有低速运动车辆的公路交通视频, 视频分辨率为 320×240 (下同)。由于部分目标运动较慢, 出于说明问题的需要我们每隔4帧抽取一帧给出处理结果。对比实验的结果见图3。图3(a)为选取的视频帧。图3(b)为使用自适应背景减方法的运动目标检测结果, 从图3(b)可见, 由于算法对环境适应性较差, 检测前景中含有较多的非目标噪声, 包括车辆的阴影和光照条件变化下的路面等均被检测为前景, 但在右上的慢速目标得到了较完整的检测说明其对慢速目标的检测效果更好。图3(c)为传统GMM算法的前景检测结果, 可以明显看出, 传统混合高斯建模方法在较大的更新率下, 易出现运动缓慢的目标前景被背景吸收的现象, 图3(c)中白色运动车辆的后部已经部分融入背景, 而右上方反向车道的低速运行车辆则已全部被背景吸收。但图3(c)所示的传统GMM方法的前景检测结果并未出现类似图3(b)中因光照条件变化而检测到的噪声, 说明即便是传统GMM方法其对环境也具有相当的适应性。图3(d)为本文所提方法的检测效果, 可见前景中非目标噪声明显少于自适应

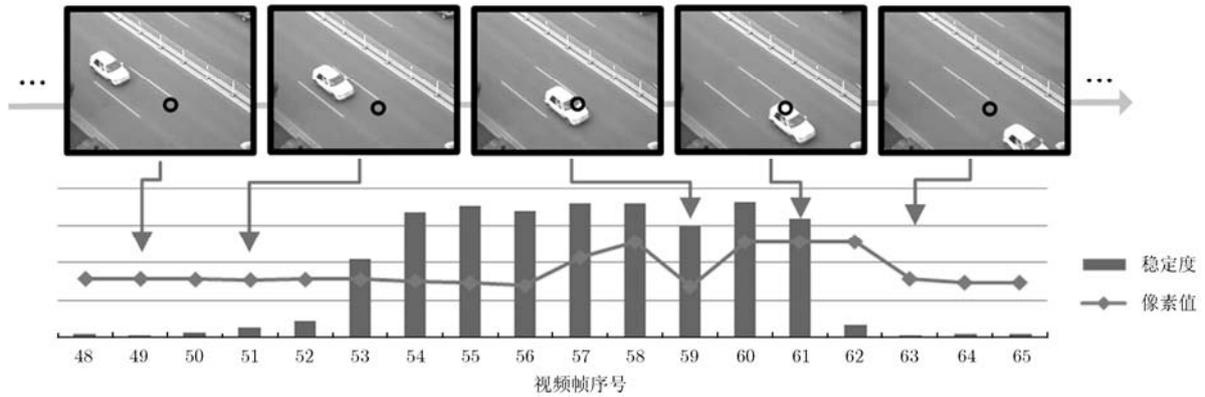


图 2 某像素点位置(黑圈处)上的灰度值和稳定度指标的变化

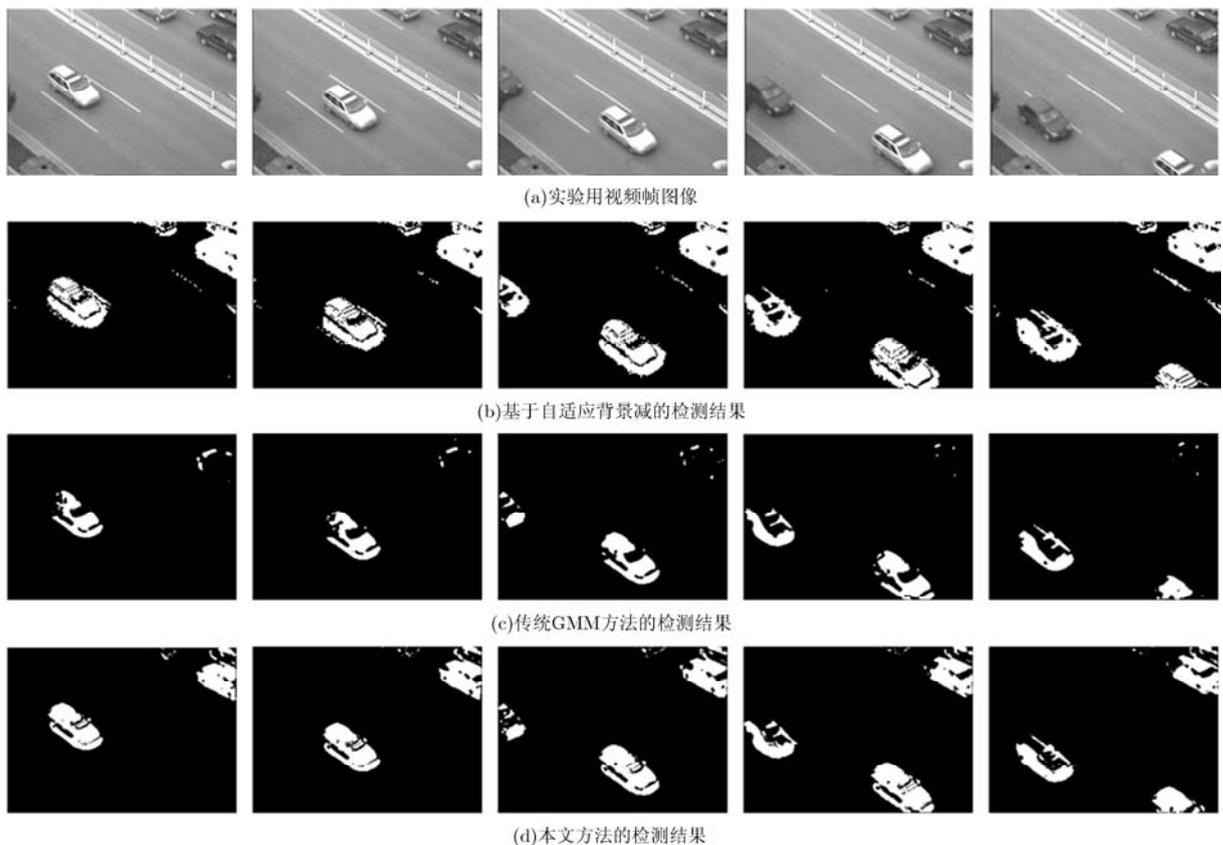


图 3 实验 1 中运动目标检测试验结果对比

背景减方法，检测出的运动目标明显较传统 GMM 方法完整，且算法对车窗等不连续部位也可做到较好的检测。

第 2 段视频为 VS-PETS 2001 dataset 的人字路口监控视频，我们使用视频中两车相遇缓慢行驶的部分进行实验，对比试验的结果见图 4。图 4(a)为选取的视频帧。图 4(b)为使用自适应背景减方法的运动目标检测结果，和上一实验中的情况相同，由于建筑物的外表面等受光照变化的影响产生改变，使得算法检测结果中含有较多的非目标噪声，说明

基于自适应背景减的方法对环境的适应性有限。图 4(c)为传统 GMM 方法的检测结果，两车相遇时运动较为缓慢，此时传统方法检测效果剧烈恶化，从图 4(c)可见相遇的两车大部分已融入背景，单纯从检测到的少量破碎前景已无法辨别目标属性。图 4(d)为本文方法的检测结果，在目标低速运动的情况下，本文方法依旧较完整地检测到了目标，且前景中的噪声明显少于自适应背景减方法的结果。

为了进一步验证本文所提方法的目标检测性能，我们使用更复杂的视频开展实验，视频背景为

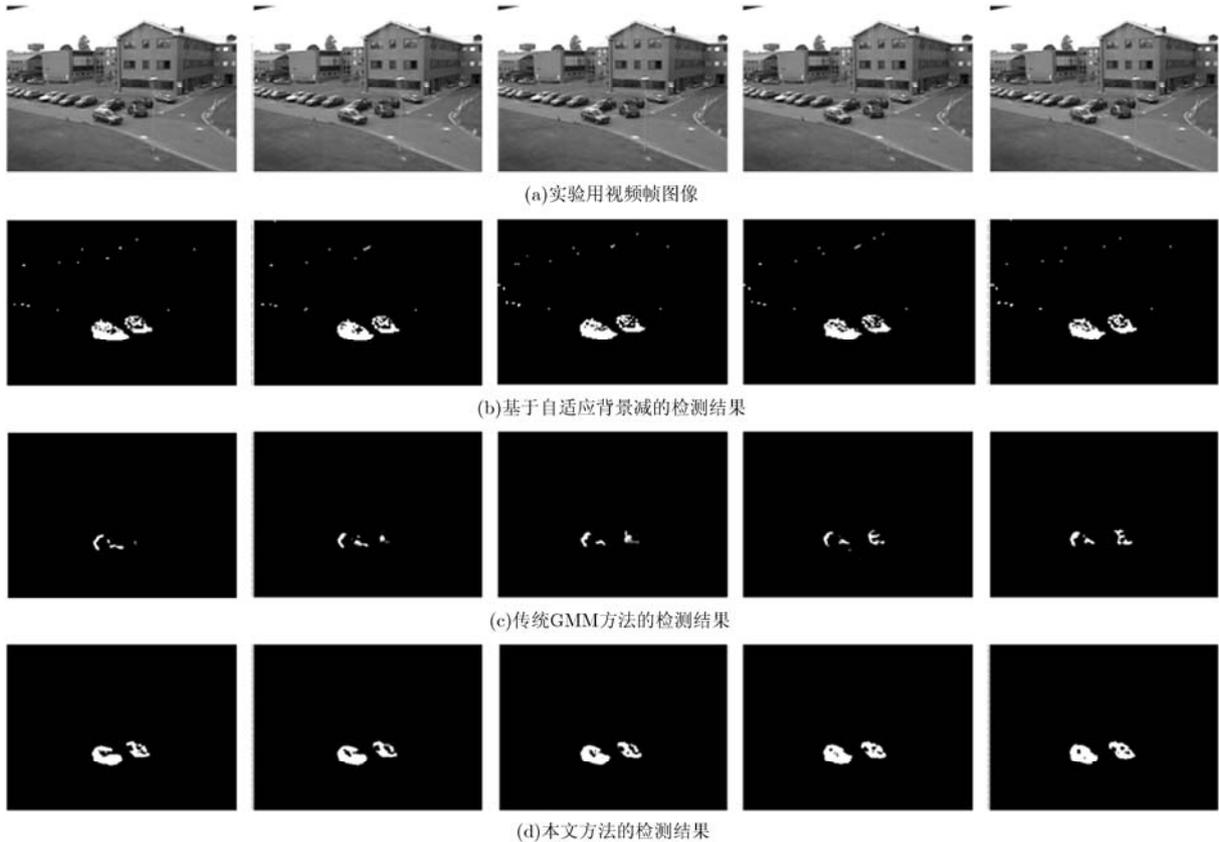


图4 实验2中运动目标检测试验结果对比

开放的有行人走动的路口，整段视频始终有人的走动并包含人的站立交谈和物品遗留等行为，实验的结果见图5。图5(a)为原始视频帧图像。图5(b)为使用自适应背景减方法的运动目标检测结果，由于场景中始终存在较多的运动目标且伴随有光照条件的变化，这影响了自适应背景减算法对背景的估计，使得检测结果中噪声较多且算法对类似人脚下阴影的光线变化过于敏感，存在较多误检。图5(c)为传统高斯混合建模前景检测结果，从图5(c)第3帧图像可见，站立交谈的行人已经大部分被吸收入背景，而从图5(c)第2, 4, 5帧可见场景中的遗留物在传统GMM方法下均未能得到有效检测。图5(d)为本文所提方法的检测结果，和图5(c)进行对比可见包括站立交谈的人和遗留物在内的所有前景均得到更好的检测，和图5(b)对比可见算法检测到的非目标前景噪声要远少于基于自适应背景减的运动目标检测算法。

为了对以上3种算法进行量化比较，使用检出率(DR)和误检率(FAR)两个指标^[13]对3个试验结果进行分析，其中

$$DR = \frac{TP}{TP + FN} \quad (13)$$

$$FAR = \frac{FP}{TP + FP} \quad (14)$$

其中 TP 为检测出来的属于运动目标区域的像素数， FP 为检测出来的不属于运动目标区域的像素数， FN 为未被检测出来的运动目标区域的像素数，实验中真实的目标区域像素为手动计算得来，虽可能存在误差，但可用来进行参考对比。对比的结果见表2，可见在3个实验中，传统GMM算法的平均检出率在60%左右，但本文方法的平均检出率接近74%，虽然自适应背景减方法也可获得较高的检出率，但其检测结果带有较多非目标噪声，所以误检率高于另两种方法1个数量级。本文方法在较传统方法大幅提高检出率的同时虽然付出一定的误检代价，其性能取舍完全合理，算法总体的性能提升显而易见。

5 结束语

传统高斯混合建模方法中背景匹配失败时生成的前景模型一直以来被人们所忽视，但事实上其包含了重要的前景目标信息。本文在传统高斯混合建模方法的基础上提出一种新的运动目标检测算法，算法对每次背景匹配失败时生成的前景模型加以利用，考察当前像素点和前景模型的匹配程度并使用

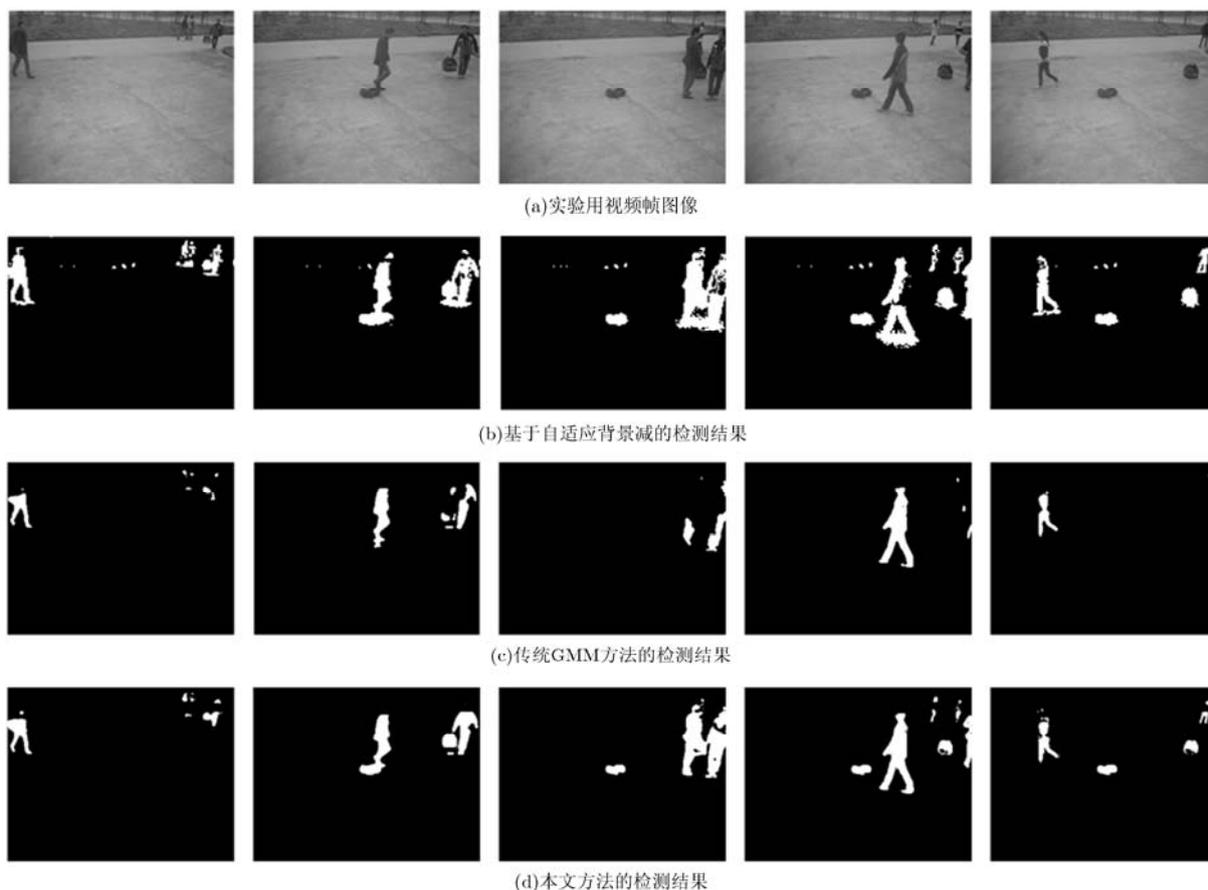


图 5 实验 3 中运动目标检测试验结果对比

短时稳定度指标对像素序列的稳定度进行度量，以此精确判断像素序列是否处于前景状态。该方法能一定程度上克服传统高斯混合建模方法无法有效应对慢速目标的局限，实现较复杂场景下的运动目标检测。

表 2 3 个实验中 3 种方法的检出率和误检率对比

		传统 GMM 方法	自适应背景减方法	本文所提方法
实验 1	检出率	0.67060	0.7447	0.7587
	误检率	0.00530	0.0308	0.0103
实验 2	检出率	0.55380	0.6752	0.7265
	误检率	0.00012	0.0258	0.0029
实验 3	检出率	0.59560	0.7226	0.7396
	误检率	0.00050	0.0218	0.0016

参 考 文 献

[1] McKenna S, Jabri S, Duric Z, et al. Tracking groups of people[J]. *Computer Vision and Image Understanding*, 2000, 80(1): 42-56.

[2] Lipton A J, Fujiyoshi H, and Patil R S. Moving target classification and tracking from real-time video[C]. *Proceedings of IEEE Workshop on Applications of Computer Vision*, Princeton, NJ, 1998: 8-14.

[3] Meyer D, Denzler J, and Niemann H. Model based extraction of articulated objects in image sequences for gait analysis[C]. *Proceedings of IEEE International Conference on Image Processing*, Santa Barbara, CA, 1998: 78-81.

[4] Stauffer C and Grimson W E L. Adaptive background mixture models for real-time tracking[C]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Fort Collins, CO, 1999: 246-252.

[5] Gao X, Boul T, Coetzee F, et al. Error analysis of background adaption[C]. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Hilton Head Island, SC, USA, 2000: 503-510.

[6] Lee D. Effective Gaussian mixture learning for video background subtraction[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(5): 827-832.

[7] Quast K and Kaup A. Real-time moving object detection in video sequences using spatio-temporal adaptive gaussian

- mixture models[C]. Proceedings of the International Conference on Computer Vision Theory and Applications, Angers, France, 2010: 413-418.
- [8] Yang S and Hsu C. Background modeling from GMM likelihood combined with spatial and color coherency[C]. IEEE International Conference on Image Processing, Atlanta, GA, 2006: 2801-2804.
- [9] Choi J, Yoo Y J, and Choi J Y. Adaptive shadow estimator for removing shadow of moving object[J]. *Computer Vision and Image Understanding*, 2010, 114(9): 1017-1029.
- [10] Li Li-yuan, Huang Wei-min, *et al.*. Statistical modeling of complex backgrounds for foreground object detection[J]. *IEEE Transactions on Image Processing*, 2004, 13(11): 1459-1472.
- [11] Salas-Gonzalez D, Kuruoglu E E, and Ruiz D P. Modelling with mixture of symmetric stable distributions using Gibbs sampling[J]. *Signal Processing*, 2010, 90(3): 774-783.
- [12] Collins R T, Lipton A, Kanade T, *et al.*. A system for video surveillance and monitoring: VSAM final report[R]. Technical Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May 2000.
- [13] Fakharian A, Hosseini S, and Gustafsson T. Hybrid object detection using improved gaussian mixture model[C]. International Conference on Control, Automation and Systems, KINTE X, Gyeonggi-do, Korea, 2011: 1475-1479.
- 张超: 男, 1983年生, 博士生, 研究方向为图像处理、盲信号处理.
- 吴小培: 男, 1966年生, 教授, 博士生导师, 研究方向为生物医学信号处理、统计信号处理.
- 周建英: 女, 1986年生, 硕士生, 研究方向为图像处理.