

## 基于自适应混合高斯模型的时空背景建模

王永忠<sup>1</sup> 梁彦<sup>1</sup> 潘泉<sup>1</sup> 程咏梅<sup>1</sup> 赵春晖<sup>1</sup>

**摘要** 提出了一种基于自适应混合高斯模型的时空背景建模方法,有效地融合了像素在时空域上的分布信息,改善了传统的混合高斯背景建模方法对非平稳场景较为敏感的缺点.首先利用混合高斯模型学习每个像素在时间域上的分布,构造了基于像素的时间域背景模型,在此基础上,通过非参数密度估计方法统计每个像素邻域内表示背景的高斯成分在空间上的分布,构造了基于像素的空间域背景模型;在决策层融合了基于时空背景模型的背景减除结果.为了提高本文时空背景建模的效率,提出了一种新的混合高斯模型高斯成分个数的自适应选择策略,并利用积分图实现了空间域背景模型的快速计算.通过在不同的场景下与多个背景建模方法比较,实验结果验证了本文算法的有效性.

**关键词** 时空背景模型,信息融合,混合高斯模型,非参数密度估计  
**中图分类号** TP391

### Spatiotemporal Background Modeling Based on Adaptive Mixture of Gaussians

WANG Yong-Zhong<sup>1</sup> LIANG Yan<sup>1</sup> PAN Quan<sup>1</sup> CHENG Yong-Mei<sup>1</sup> ZHAO Chun-Hui<sup>1</sup>

**Abstract** The background model of traditional mixture of Gaussians is less robust to non-stationary scenes. This paper presents an adaptive spatiotemporal background model, combining the temporal information of per-pixel and the spatial information in the local region. Based on the temporal distribution model learned by mixture of Gaussians, the spatial background model of per-pixel is utilized to construct the spatial distribution of background in the local region by non-parametric density estimation. The robust detection is achieved by fusing the subtraction results separately based on the temporal and spatial background models. Additionally, to improve the computation efficiency, an adaptive selection strategy of the number of components of mixture of Gaussians model is proposed and integral image method is applied to calculate the spatial background model. Experimental comparisons demonstrate the effectiveness of the proposed method.

**Key words** Spatiotemporal background model, information fusion, mixture of Gaussians, non-parametric density estimation

基于序列图像的运动目标检测是智能视频监控、基于视觉的人机交互、视频压缩及步态识别等领域的重要研究内容.背景减除法近年来在运动目标检测中获得了大量的应用,背景减除法是通过比较当前帧与背景模型之间的差别来实现对运动目标的检测,其性能依赖于所使用的背景建模技术.背景建模遇到的主要问题是在自然场景中背景往往是动态变化的,例如晃动的树叶、变化的光照及波动的水面等,即使背景是静止的,摄像机的颤动及信号噪声等也会给建立鲁棒的背景模型提出挑战性的要求.常用的背景建模方法一般可分为基于像素的背景建

模方法和基于块的背景建模方法.

基于像素的背景建模方法主要是根据每个像素在时间域上的分布信息构造其背景模型.如 Friedman 等<sup>[1]</sup>在交通场景监控中提出了混合高斯(Mixture of Gaussians, MoG)背景建模方法,图像中每个像素的混合高斯概率模型由三个高斯成分组成,分别对应于道路、车辆及阴影.随后,Stauffer 等<sup>[2]</sup>提出了一种更为一般的混合高斯背景建模方法,在学习模型参数的过程中,用在线 K 均值近似的方法代替了严格的期望最大(Expectation maximization, EM)算法,提高了混合高斯模型学习的效率.此后的研究者又对混合高斯背景建模方法做了各种的改进<sup>[3-6]</sup>.此外,Elgammal 等<sup>[7]</sup>利用核密度估计提出了一种非参数背景建模算法,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.Stenger 等<sup>[8]</sup>在背景建模中使用了一种拓扑自由的隐马尔可夫模型,用于处理光照的突然变化.另外,维纳滤波<sup>[9]</sup>、条件随机场<sup>[10]</sup>等方法也被用于基于像素的背景建模中.利用基于像素的背景建模方法能获得较细致的运动目标形状,但当背景发生非平稳变化时(如突然一阵风引起的树木的晃动),现有的基于像素的背景减除

收稿日期 2007-12-07 收修改稿日期 2008-03-30  
Received December 7, 2007; in revised form March 30, 2008  
国家自然科学基金重点项目(60634030),航空科学基金(2007ZC53037),高等学校博士学科点专项基金(20060699032),教育部新世纪优秀人才支持计划(NCET-06-0878),西北工业大学科技创新基金资助  
Supported by National Natural Science Foundation of China (60634030), Air Science Foundation (2007ZC53037), Research Fund for the Doctoral Program of Higher Education of China (20060699032), Program for New Century Excellent Talents in University of Ministry of Education of China (NCET-06-0878), and Northwestern Polytechnical University Teachers Scientific and Technological Innovation Foundation  
1. 西北工业大学自动化学院 西安 710072  
1. College of Automation, Northwestern Polytechnical University, Xi'an 710072  
DOI: 10.3724/SP.J.1004.2009.00371

方法依然会产生大量的虚警. 其原因是现有的基于像素的背景建模方法大都仅利用了每个像素在时间域上的分布信息, 并没有利用像素在空间域上的分布信息, 受模型学习速度的影响, 场景的突然变化不能被立即学入背景模型中, 从而使得此类方法对场景的非平稳变化依然比较敏感.

基于块的背景建模方法是将每帧图像分割为多个重叠或非重叠的小块, 然后使用每个块的特征构造背景模型. 与单个像素相比, 块能够获得更多的空间分布信息, 使得基于块的背景建模方法对场景的局部变化不敏感. 如 Seki 等<sup>[11]</sup> 利用相邻块之间的图像变化在空间上的同现 (Cooccurrence) 现象, 改善对前景目标的检测效果. Matsuyama 等<sup>[12]</sup> 使用归一化向量距离 (Normalized vector distance) 度量块之间的相关性. Heikkilä 等<sup>[13]</sup> 用每个像素邻域内的 LBP (Local binary pattern) 直方图去描述背景模型. 基于块的背景减除方法最大缺点是获得的运动目标比较粗糙, 不适于对运动目标分割要求较高的应用场合. Chen 等<sup>[14]</sup> 提出了一种分级的背景建模方法, 在利用基于块的背景模型对运动目标检测的基础上, 用基于像素的背景模型对运动目标做进一步检测, 以获得准确的目标外形, 但块的划分依然会影响到对弱小运动目标的检测.

显然, 如果能够充分地融合背景在时间及空间上的分布信息, 将可以获得更为鲁棒的背景模型. 但简单地将基于像素和基于块的背景建模方法组合起来, 机理上会造成计算量的大量增加. 针对上述问题, 本文提出了一种基于自适应混合高斯模型的时空背景建模方法, 在利用混合高斯模型学习每个像素的时间域背景模型之后, 使用非参数密度估计方法构造了基于像素的空间域背景模型, 从而有效地融合了每个像素在时空域上的分布信息; 同时利用提出的混合高斯模型高斯成分个数的自适应选择策略及基于积分图的空间域背景模型快速计算方法, 改善了本文时空背景建模的效率; 最后在所提出的时空背景模型的基础上, 利用决策融合实现了鲁棒的背景减除. 在实验部分比较了本文方法与传统的混合高斯背景建模方法<sup>[2]</sup> 及文献 [13] 的 TBMOD 方法在多个典型场景下的运动目标检测性能.

## 1 自适应时空背景模型

由于混合高斯模型可以表示多模分布的背景, 并且具有存储量小及可以自适应学习背景的优点, 因此我们利用混合高斯模型构造本文的时空域背景模型. 时空背景建模的流程如图 1 所示, 首先利用自适应混合高斯模型统计每个像素在时间域上的分布信息, 构造每个像素的时间域背景模型; 然后在时间域背景模型的基础上, 用非参数方法统计每个像

素邻域内表示背景的高斯成分在空间域上的分布信息, 构造了背景的空间域模型, 实现了基于像素的时空背景建模.

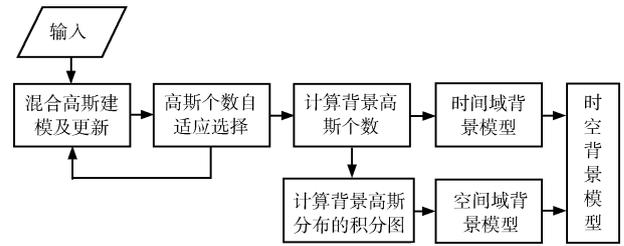


图 1 时空背景建模流程图

Fig. 1 Scheme of spatiotemporal background modeling

这样做的好处是空间域背景模型是建立在时间域背景模型对场景学习的基础上, 因此不需要对空间域模型重新学习. 由于是基于每个像素构造时空域背景模型, 因而造成计算量较大. 为了提高背景建模的效率, 本文提出了一种新的混合高斯模型的高斯成分个数的自适应选择策略, 并提出了基于积分图的空间域背景模型计算方法.

### 1.1 时间域背景模型

#### 1.1.1 混合高斯模型

文献 [2] 用由  $K$  个高斯成分组成的混合高斯模型表示像素在时间域上的概率分布模型, 即图像中像素  $j$  在时刻  $t$  取值为  $\mathbf{x}_j$  的概率为

$$P(\mathbf{x}_j) = \sum_{i=1}^K \omega_{j,t}^i \cdot \eta(\mathbf{x}_j; \boldsymbol{\mu}_{j,t}^i, \boldsymbol{\Sigma}_{j,t}^i) \quad (1)$$

其中,  $\omega_{j,t}^i$  表示在时刻  $t$  像素  $j$  的混合高斯模型中第  $i$  个高斯成分的权重,  $\boldsymbol{\mu}_{j,t}^i$  及  $\boldsymbol{\Sigma}_{j,t}^i = (\sigma_{j,t}^i)^2 I$  分别表示第  $i$  个高斯成分的均值及协方差,  $\sigma$  表示标准差,  $I$  表示单位矩阵,  $\eta$  表示高斯概率密度函数

$$\eta(\mathbf{x}_j; \boldsymbol{\mu}_{j,t}^i, \boldsymbol{\Sigma}_{j,t}^i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}_{j,t}^i|^{\frac{1}{2}}} \times \exp \left[ -\frac{1}{2} (\mathbf{x}_j - \boldsymbol{\mu}_{j,t}^i)^T (\boldsymbol{\Sigma}_{j,t}^i)^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_{j,t}^i) \right] \quad (2)$$

其中  $d$  为  $\mathbf{x}_j$  的维数.

随着场景的变化, 每个像素的混合高斯模型都需要被不断地学习更新, 方法是将混合高斯模型中的  $K$  个高斯成分按照  $\omega_{j,t}^i / \sigma_{j,t}^i$  由大到小排序, 然后用像素的当前值  $\mathbf{x}_j$  与其混合高斯模型中的  $K$  个高斯成分逐一比较, 若  $\mathbf{x}_j$  与第  $i$  个高斯成分的均值  $\boldsymbol{\mu}_{j,t}^i$  之间的差小于  $\delta$  倍的该高斯成分的标准差  $\sigma_{j,t}^i$  (通常  $\delta$  设为 2.5 ~ 3.5), 则该高斯成分被  $\mathbf{x}_j$  更新, 其余高斯成分保持不变, 更新方程<sup>[6]</sup> 如下

$$\omega_{j,t+1}^i = (1 - \alpha) \omega_{j,t}^i + \alpha \cdot (M_{j,t}^i) \quad (3)$$

$$\boldsymbol{\mu}_{j,t+1}^i = (1 - \rho)\boldsymbol{\mu}_{j,t}^i + \rho \cdot \mathbf{x}_j \quad (4)$$

$$(\sigma_{j,t+1}^i)^2 = (1 - \rho)(\sigma_{j,t}^i)^2 + \rho(\mathbf{x}_j - \boldsymbol{\mu}_{j,t}^i)^\top (\mathbf{x}_j - \boldsymbol{\mu}_{j,t}^i) \quad (5)$$

$$\rho = \frac{\alpha}{\omega_{j,t}^i} \quad (6)$$

其中  $\alpha$  是模型的学习率, 当第  $i$  个高斯成分与  $\mathbf{x}_j$  匹配时,  $M_{j,t}^i$  为 1, 否则为 0; 如果  $\mathbf{x}_j$  与像素  $j$  的混合高斯模型中的  $K$  个高斯成分都不匹配, 则该像素混合高斯模型中排在最后的高斯成分被新的高斯成分取代, 新的高斯成分的均值为  $\mathbf{x}_j$ , 初始标准差及权重设为  $\sigma_{\text{init}}$  和  $\omega_{\text{init}}$ . 在更新完成后, 各高斯成分的权重被归一化, 以使  $\sum_{i=1}^K \omega_{j,t+1}^i = 1$ .

像素  $j$  的混合高斯模型描述了其特征值  $\mathbf{x}_j$  在时间域上的概率分布, 为了确定像素的混合高斯模型里哪些高斯成分是由背景产生的, 根据每个高斯成分的权重与其标准差之比排序后, 取前  $B_j$  个高斯成分作为背景的概率分布,  $B_j$  根据下式计算

$$B_j = \arg \min_b \left( \sum_{i=1}^b \omega_{j,t+1}^i > T \right) \quad (7)$$

阈值  $T$  度量了背景高斯成分在像素的整个概率分布中所占的最小比例, 当  $T$  取值较小时, 背景用单模分布表示; 当  $T$  取值较大时, 可以表示多模分布的背景.

### 1.1.2 高斯成分个数自适应选择

现有的混合高斯背景建模方法大多采用固定高斯成分个数的方法, 像素混合高斯模型中的每个高斯成分对应场景的一个状态. 但在实际场景中, 在不同的区域背景的状态个数通常是不一样的. 此外, 随着场景的变化, 同一个区域的状态个数也会经常改变. 在每帧中对所有像素的混合高斯模型都保持相同的固定不变的高斯成分个数 (通常为 3~5 个), 会造成系统运算资源的大量浪费. 因此有必要根据场景自适应选择每个像素混合高斯模型的高斯成分个数. Zivkovic 等<sup>[4]</sup> 利用最大似然估计提出了一种高斯成分个数的选择方法, 由于该方法人为地引入了负的先验系数, 使得在更新过程中高斯成分的权重有可能被不合理地负更新. 与该方法不同的是, 本文提出了一种更为直观的高斯成分个数自适应选择策略.

从式 (3) 可以看出, 像素混合高斯模型中长时间与场景匹配的高斯成分的权重会越来越大, 而权重归一化后, 那些不匹配的高斯成分的权重会越来越小, 根据式 (7), 这些高斯成分会逐渐落入表示前景的部分, 当某个高斯成分的权重  $\omega_{j,t}^i$  小于初始权重  $\omega_{\text{init}}$ , 且该高斯成分的  $\omega_{j,t}^i/\sigma_{j,t}^i$  小于初始的  $\omega_{\text{init}}/\sigma_{\text{init}}$  时, 经过排序, 该高斯成分将会被排在新

初始化的高斯成分之后. 继续保留该高斯成分, 将使得再次出现与这个高斯成分匹配的场景时, 利用该高斯成分学习此场景比用一个新的高斯成分学习此场景所花费的时间更长, 我们将这些高斯成分称为“过期的”高斯成分. 显然, 过期高斯成分的存在既影响模型的学习收敛速度, 又浪费系统的计算资源, 应该被删除. 过期高斯成分的判别公式如下

$$\omega_{j,t}^i < \omega_{\text{init}} \quad \text{且} \quad \frac{\omega_{j,t}^i}{\sigma_{j,t}^i} < \frac{\omega_{\text{init}}}{\sigma_{\text{init}}} \quad (8)$$

根据以上分析, 我们提出如下的高斯成分个数自适应选择策略: 1) 初始化时, 给场景中每个像素的混合高斯模型都只设置一个高斯成分; 2) 增加新的高斯成分: 随着场景的变化, 当某个像素的混合高斯模型不能与当前像素值匹配时, 如果该像素混合高斯模型中的高斯成分个数没有达到设定的最大值, 则自动地增加一个以当前值为均值的初始高斯成分, 否则用以当前像素值为均值的新的的高斯成分代替该像素混合高斯模型中排在最后的一个高斯成分; 3) 删除过期的高斯成分: 在每次模型更新完成后, 根据式 (8) 判断每个像素混合高斯模型中的最后一个高斯成分是否过期, 如果过期则删除.

图 2 (见下页) 显示了本文自适应选择高斯成分个数的混合高斯背景建模方法 (Adaptively selecting components number of MoG, ACMoG) 在 PETS 2000 测试序列上的运行结果, 第一列分别显示了第 345 帧、第 750 帧及第 1450 帧的原始图像, 第二列分别显示了在这些帧上每个像素混合高斯模型中高斯成分个数的分布. 实验中为每个像素设置的最大高斯成分个数  $K_{\text{max}} = 3$ , 高斯成分个数分布图中的黑色、灰色及白色分别代表像素的混合高斯模型由一个、两个及三个高斯成分组成, 初始化时在第 1 帧为场景中每个像素的混合高斯模型只设置一个高斯成分.

从图 2 中可以看出场景中大多数区域只需要一个高斯成分就可以表示, 如草地等地方; 而在那些频繁变化的区域, 如车辆及行人经过的地方, 像素混合高斯模型中的高斯成分个数会自适应地由一个增加为两个甚至三个; 此外, 由于摄像机的抖动, 也会造成在那些纹理丰富区域像素的高斯成分个数增加, 如在第 1450 帧中道路的边缘及右上角的建筑物等处; 最后可以看出在一些区域高斯成分个数会自适应地减少, 如第 750 帧所示, 在右下方道路区域由于长时间没有车辆经过, 一些像素的高斯成分个数会自适应地由第 345 帧的三个减少为了两个甚至一个; 同样的现象可以从场景的右上角看到, 在第 345 帧鸟飞过的区域到第 750 帧时一些像素的高斯成分个数会自适应地由两个减少为一个. 需要指出的是, 对

那些运动较慢的前景目标所对应的高斯成分, 因为匹配次数较多而使权重变得较大, 因而删除的速度会较慢, 如行人经过的区域. 上述实验结果表明, 本文提出的高斯成分个数自适应选择策略能够根据场景的变化自适应地选择每个像素混合高斯模型中的高斯成分个数.

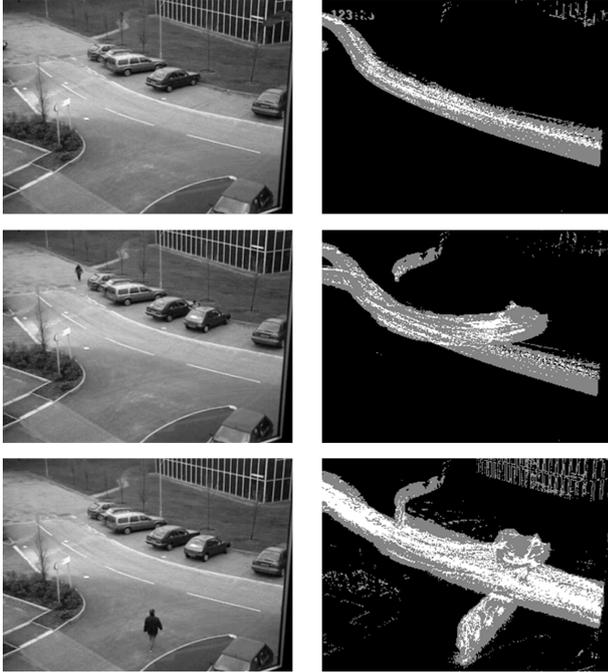


图 2 混合高斯模型的高斯成分个数分布

Fig. 2 Distribution of MoG's component numbers

表 1 给出了本文自适应选择高斯成分个数的混合高斯背景建模方法 (ACMoG) 与传统的混合高斯背景建模方法 (MoG)<sup>[2]</sup> 在不同测试序列上的处理速度. 对各测试序列的详细说明见本文第 3 节. 可以看出通过自适应地选择每个像素混合高斯模型中的高斯成分个数, 可以显著地提高背景建模的速度, 从而有效地节约了运算资源; 此外, 随着场景复杂程度的增加, 每个像素的混合高斯模型需要更多的高斯成分, 因此复杂场景的背景建模 (如在 Suburb 序列) 所需的处理时间也会相应增加.

表 1 ACMoG 与 MoG 算法的处理速度 (ms/f)

Table 1 Processing times of MoG and ACMoG (ms/f)

算法	PETS 2000	Campus	Suburb	Pond	平均
MoG	43.0	50.1	50.4	49.7	48.3
ACMoG	29.7	32.2	33.5	27.5	30.7

## 1.2 空间域背景模型

### 1.2.1 空间域背景模型的非参数估计

文献 [13, 15] 已经表明每个像素邻域内的空间

分布特征可以用来建立背景模型. 文献 [13] 利用像素邻域内的多个局部二值模式 (Local binary pattern, LBP) 纹理特征直方图描述动态背景, 为了适应场景的变化, 需要不断地学习表示背景的直方图. 文献 [15] 将不包含目标的初始帧作为背景, 利用颜色直方图统计图像中每个块内的像素分布, 但该方法对场景的变化不具有自适应性.

与此前的方法不同, 本文在利用时间域背景模型对场景学习之后, 已经获得了每个像素的混合高斯模型中表示背景的前  $B_j$  个高斯成分, 这些高斯成分的均值对应着背景的状态, 权重表示该背景状态出现的相对频率. 即通过基于混合高斯模型的时间域背景模型对场景的学习, 我们已经获得了表示背景的一组样本. 因而可以直接统计这些表示背景的样本在空间上的分布, 作为像素的空间域背景模型. 由于每个样本 (即背景高斯成分) 在时间上出现的频率不一样, 因此在统计像素的空间域背景模型时, 需要用背景高斯成分的权重对相应的样本进行加权. 本文利用颜色直方图来统计每个像素邻域内背景高斯成分的分布. 颜色直方图是一种简单的非参数概率密度估计方法, 它具有旋转及平移不变性, 用其统计像素邻域内的背景分布能够较好地克服背景局部运动的干扰. 设像素  $j$  用颜色直方图表示的空间域背景模型为  $\mathbf{q}(\mathbf{x}_j) = \{q_v(\mathbf{x}_j)\}_{v=1, \dots, m}$ ,  $m$  为直方图区间的个数,

$$q_v(\mathbf{x}_j) = C_q \sum_{l \in x_j^N} \sum_{i=1}^{B_l} \omega_{l,t}^i \delta[b(\boldsymbol{\mu}_{l,t}^i) - v] \quad (9)$$

其中,  $x_j^N$  表示以像素  $j$  为中心的  $N \times N$  邻域,  $B_l$  为该邻域内第  $l$  个像素的混合高斯模型中背景高斯成分的个数,  $\omega_{l,t}^i$  为像素  $l$  的第  $i$  个高斯成分的权重,  $b(\boldsymbol{\mu}_{l,t}^i)$  表示该高斯成分的均值  $\boldsymbol{\mu}_{l,t}^i$  对应的直方图颜色区间,  $v$  为相应的颜色区间,  $\delta$  为 Kronecker 函数,  $C_q$  为归一化系数. 由于  $\omega_{l,t}^i$  表示了相应的背景状态在时间上出现的相对频率, 反映了该背景状态在时间上的分布信息, 所以本文的空间域背景模型也同时反映了背景在时间域上的分布信息, 因而用一个直方图即可表示时间上动态变化的背景.

在获得了像素的背景空间域模型后, 为了实现背景减除, 还需要计算当前场景中像素在空间上的颜色分布. 同样, 本文利用非参数概率密度估计方法统计像素  $j$  的邻域  $x_j^N$  在当前帧中的颜色分布直方图  $\mathbf{p}(\mathbf{x}_j) = \{p_v(\mathbf{x}_j)\}_{v=1, \dots, m}$ . 由于图像中获得的像素值通常会受到噪声的干扰, 如果直接利用当前像素值作为样本统计每个像素邻域内的颜色直方图, 会使颜色直方图易受噪声的影响. 因此当像素  $l$  的当前值  $x_l$  与其混合高斯模型的某个高斯成分匹配

时, 我们用该高斯成分的均值作为当前空间颜色直方图的统计样本, 否则直接用该像素的当前值作为统计样本, 于是

$$p_v(\mathbf{x}_j) = C_p \sum_{l \in x_j^N} \sum_{i=1}^{K_l} \delta\{b[\boldsymbol{\mu}_{l,t}^i \cdot (M_{l,t}^i)] - v\} \quad (10)$$

其中,  $K_l$  为像素  $l$  的混合高斯模型中的高斯成分个数,  $C_p$  为归一化系数.

在背景减除过程中, 需要判断在当前帧中像素的空间分布是否与其空间域背景模型相似, 即判断两个直方图之间的相似性. 为了计算简单, 本文采用直方图相交作为空间域背景模型与当前场景的相似性度量. 直方图相交计算的是两个直方图共有的部分, 忽略了仅在其中一个直方图出现的特征, 其计算复杂度为  $O(m)$ , 因而计算复杂度较低. 像素  $j$  的空间域背景模型与当前场景的相似性  $\rho$  表示为

$$\rho(\mathbf{x}_j) = \sum_{v=1}^m \min(p_v, q_v) \quad (11)$$

需要指出的是, 本文的空间域背景模型利用了时间域背景模型对场景的学习结果, 不需要单独保留空间域的背景历史信息, 从而节省了存储空间. 与基于块的背景建模方法是基于每个块构造背景模型不同, 本文是基于每个像素建立空间域背景模型, 因而提高了背景模型在空间上的分辨率, 但同时会造成计算量的增加.

### 1.2.2 空间域背景模型的积分图计算

计算空间域背景模型, 需要计算每个像素的  $N \times N$  邻域内的直方图, 假设每个像素的混合高斯模型平均有  $A$  个高斯成分表示背景, 则计算该像素空间域背景模型的复杂度为  $O(N^2A + m)$ , 其中  $m$  为直方图区间的个数. 每帧图像的空间域背景建模的计算复杂度为  $O(LN^2A + Lm)$ ,  $L$  为每帧图像所包含的像素个数. 可以看出, 为每帧图像建立空间域背景模型的计算复杂度还是较高的, 这会影响运动目标检测的实时性. 造成复杂度较高的主要原因是: 在为每帧图像建立空间域背景模型时, 图像中的每个像素的背景高斯成分作为样本会多次出现在其邻域内不同像素的空间域背景模型中, 使得每个像素的背景高斯成分被重复计算, 重复次数可以达到  $N^2$  次, 因而造成了运算资源的大量浪费.

为了提高空间域背景建模的效率, 我们采用文献 [16] 提出的积分图 (Integral image) 方法计算场景的空间域背景模型. 文献 [16] 用每个像素的灰度值表示其特征, 与该方法不同的是, 本文用  $m$  维颜色直方图表示每个像素的特征, 积分图的具体计算过程详见文献 [16]. 在获得了整幅图像的积分图之

后, 可以通过简单的加减运算获得任意大小的矩形区域内背景高斯成分的颜色分布直方图. 利用积分图计算空间域背景模型, 避免了多次重复统计每个像素的背景高斯成分, 计算复杂度为  $O(LA + Lm)$ . 在本文实验中  $m$  一般取为 64, 而  $N$  取为 9~15. 可以看出采用积分图方法计算空间域背景模型的复杂度  $O(LA + Lm)$  要远小于不采用积分图计算空间域背景模型的复杂度  $O(LN^2A + Lm)$ , 因此可以有效地提高空间域背景建模的速度.

## 2 背景减除

利用本文的时空背景模型, 我们提出了一种基于决策融合的运动目标检测方法, 其主要思想是用空间域背景模型进行粗尺度判断, 当像素的当前邻域与其空间域背景模型极为相似时, 即  $\rho > \tau_1$  ( $\tau_1$  为判断当前场景为背景的相似性度量的下界) 时, 可以判定该像素的邻域来自于背景, 从而也可以判定该像素是来自于背景; 当该像素的当前邻域与其空间域背景模型极为不相似时, 即  $\rho < \tau_2$  ( $\tau_2$  为判断当前场景为前景的相似性度量的上界) 时, 则可以断定该像素的当前邻域来自于前景, 进而判定该像素也是来自于前景; 当基于空间域背景模型不能准确地判断像素的当前邻域是属于前景还是背景时, 即  $\tau_1 \leq \rho \leq \tau_2$ , 我们利用时间域背景模型对该像素进行细尺度判断, 即当该像素的混合高斯模型里存在与其当前值匹配的背景高斯成分时, 则像素为背景, 否则为前景. 决策公式如下

如果  $\tau_1 \leq \rho \leq \tau_2$

$$D(\mathbf{x}_j) = \begin{cases} 0, & \exists i \leq B_j \text{ 且 } M_{j,t}^i = 1 \\ 1, & \text{其他} \end{cases} \quad (12)$$

否则

$$D(\mathbf{x}_j) = \begin{cases} 0, & \rho > \tau_1 \\ 1, & \rho < \tau_2 \end{cases} \quad (13)$$

其中,  $D(\mathbf{x}_j)$  为 0 时表示像素  $j$  为背景, 为 1 时表示像素  $j$  为前景. 与文献 [2] 相比, 我们是根据像素在时间及空间域上的分布, 对每个像素进行检测的, 由于利用了像素邻域内空间分布信息, 因而可以减轻传统的混合高斯背景模型对非平稳变化比较敏感的缺点. 与基于块的方法<sup>[15]</sup> 相比, 本文方法依然是一种基于像素的背景减除方法, 因而能保持较好的运动目标轮廓.

## 3 实验结果及分析

为了测试本文基于混合高斯的时空背景建模 (Spatiotemporal modeling based on mixture of

Gaussians, STMoG) 方法的有效性, 我们在多个视频上进行了测试, 并与经典的混合高斯背景建模 (MoG)<sup>[2]</sup> 方法及文献 [13] 提出的 TBMOD 算法进行了比较. 在所有测试序列中 STMoG 及 MoG 算法所采用的参数为:  $K_{\max} = 4$ ,  $\alpha = 0.001$ ,  $\sigma_{\text{init}} = 30$ ,  $\omega_{\text{init}} = 0.05$ ,  $T = 0.7$ ,  $\delta = 2.5$ ,  $\tau_1 = 0.9$ ,  $\tau_2 = 0.1$ , 每个像素邻域的大小为  $11 \times 11$ , 表示空间域背景模型的颜色直方图的区间数为 64 个, TBMOD 算法的参数选择见文献 [13].

本文所用的测试视频如图 3 所示, 序列 1 为 Wavetrees 视频<sup>[9]</sup>, 长 286 帧, 包含了动态变化的背景; 序列 2 为在郊外拍摄的 Suburb 视频, 长 1747 帧, 该视频由于包含了大量不断变化的杂乱背景, 代表了动态变化场景的一种极端情形; 序列 3 为 PETS 2000 视频, 长 1450 帧, 该视频在第 819 帧会发生摄像机的轻微抖动; 序列 4 是实际拍摄的 Campus 视频, 长度为 950 帧, 该视频系列的背景包含了大量的纹理, 同时在 892 帧会发生较大的摄像机抖动; 序列 5 为 Pond 视频, 长 1100 帧, 在该视频场景中包含了不断波动的水面. 除序列 1 的图像大小为  $160 \times 120$  像素, 其余四个视频图像大小均为  $320 \times 240$  像素.

图 3 显示了本文 STMoG 背景建模算法与经典

的 MoG 算法及 TBMOD 算法在多个测试视频上的检测结果, 图中的测试帧分别为 Wavetrees 序列的第 247 帧, Suburb 序列的 805 帧, PETS 2000 序列的 819 帧, Campus 序列的 892 帧及 Pond 序列的第 633 帧, 所有的检测结果均未做后处理. 可以看出, 在 Wavetrees 序列中, 虽然 MoG 算法对背景具有一定的自适应性, 但由于背景运动的不规则, 依然产生了较多的虚警; 特别是在 Suburb 序列中, MoG 检测算法产生了大量的虚警. TBMOD 算法虽然能够较好地消除动态背景的影响, 并能降低对运动目标的漏检率, 但该方法会在目标周围产生大量的虚警, 使得对目标的分割不够准确. 与上述两种方法相比, 本文 STMoG 算法在 Wavetrees 及 Suburb 序列上都能够很好地克服动态背景的干扰, 虚警大大低于 MoG 算法, 并获得比 TBMOD 算法更为准确的目标轮廓. 从对 Pond 序列的检测结果可以看出, 与经典的 MoG 算法相比, 本文 STMoG 算法能够较好地克服波动水面的影响, 而 TBMOD 在此序列上仍然存在对目标分割不够准确的缺点. 此外, 由于 Campus 序列在 892 帧摄像机发生了较大的抖动, MoG 算法该序列的纹理边缘区域产生了大量的虚警, 同样的情形发生在 PETS 2000 序列的第 819 帧, 由于摄像机的抖动, MoG 算法也在图像的右上



图 3 运动目标检测结果

Fig. 3 Some detection results of our method and the previous methods

角的多纹理区域也产生了许多虚警. TBMOD 算法虽然能够在纹理较为稀疏的 Campus 序列上克服摄像机抖动造成的影响, 但在 PETS 2000 视频的左上角纹理密集区域, TBMOD 算法产生了比 MoG 算法更多的虚警. 而本文 STMoG 算法在这两个视频上的检测结果要优于 MoG 及 TBMOD 算法. 本文 STMoG 算法虽然结合了背景在时空域上的分布信息, 但由于该方法依然是一种基于像素的背景建模方法, 在降低虚警的同时, 即使在 Campus 序列中目标较小的情况下, 依然保持了较好的目标外形.

图 4 给出了 MoG 算法、TBMOD 算法及本文 STMoG 算法对上述五个视频检测结果的定量评价, 评价指标采用了文献 [13] 的方法, 分别统计了三种检测算法的虚警数 FP (False positive) 和漏检数 FN (False negative), 漏检数是指前景被检测为背景的像素个数, 虚警数是指背景被检测为前景的像素个数. 除序列 1 本身提供基准图 (Benchmark) 外, 其他四个序列的基准图均通过手工选定. 可以看出在所有的测试视频上本文 STMoG 算法的虚警数都远低于 MoG 及 TBMOD 算法, 而漏检数与 MoG 算法相当; TBMOD 算法的漏检数在三个序列上要低于 MoG 及 STMoG 算法, 在两个序列上漏检数略高于 MoG 及 STMoG 算法; 在总的检测效果上, 本文 STMoG 算法在所有测试序列上都要优于 MoG 及 TBMOD 算法, 而 MoG 及 TBMOD 在不同的测试序列上各有优劣. 从上述实验结果可以看出, 本文 STMoG 算法在保持较低的漏检数的前提下, 能有效地降低虚警数, 而通常高虚警率也是许多现有的视频应用系统所面对的主要问题.

表 2 比较了 TBMOD 算法及本文 STMoG 算法在同样的测试平台下的处理速度, 采用的硬件测试平台为 P4 3.0 GHz 768 MB 计算机. 为了比较, 所有的测试序列被规整到  $160 \times 120$  大小, 可以看出本文 STMoG 算法处理速度明显要快于 TBMOD 算

法, STMoG 算法检测每帧所需的平均处理时间为 64.9 ms, 即可以达到 15 fps 的处理速度. 显然, 由于增加了空间域背景模型, 使得本文 STMoG 算法的处理速度要低于传统的基于 MoG 的背景减除算法. 但基于 MoG 检测算法在动态背景下往往会产生大量的虚警, 从而需要花费大量的时间用于运动检测的后处理 (如形态学滤波), 并给准确的目标分割带来困难.

表 2 STMoG 及 TBMOD 算法的处理速度 (ms/f)

Table 2 Processing times of TBMOD and STMoG (ms/f)

算法	Wavetrees	PETS 2000	Campus	Suburb	Pond	平均
TBMOD	135.7	123.5	133.7	142.2	129.9	133.0
STMoG	72.1	56.9	62.1	75.9	57.5	64.9

虽然本文采用了一种比较简单的融合策略, 但实验结果显示了本文 STMoG 算法能够明显改善检测效果, 表明本文所提出的时空背景建模方法是一种比较有效的背景建模方法. 当然如果采用更复杂且有效的融合策略, 应该可以获得更好的检测结果, 这将是我们的进一步的研究内容.

## 4 结论

本文提出了一种基于自适应混合高斯模型的时空背景建模方法, 该方法同时利用了像素在时间域及空间域上的分布信息, 改善了经典的混合高斯背景建模方法对非平稳背景变化较为敏感的缺点. 基于像素的决策级融合检测算法在提高检测效果的同时, 保持了较好的运动目标外形. 此外, 通过自适应选择每个像素混合高斯模型中的高斯成分个数及利用积分图计算背景空间域模型, 保证了本文算法具有一定的实时性. 经过在多个视频序列上与经典的混合高斯背景方法及 TBMOD 算法比较, 实验结果

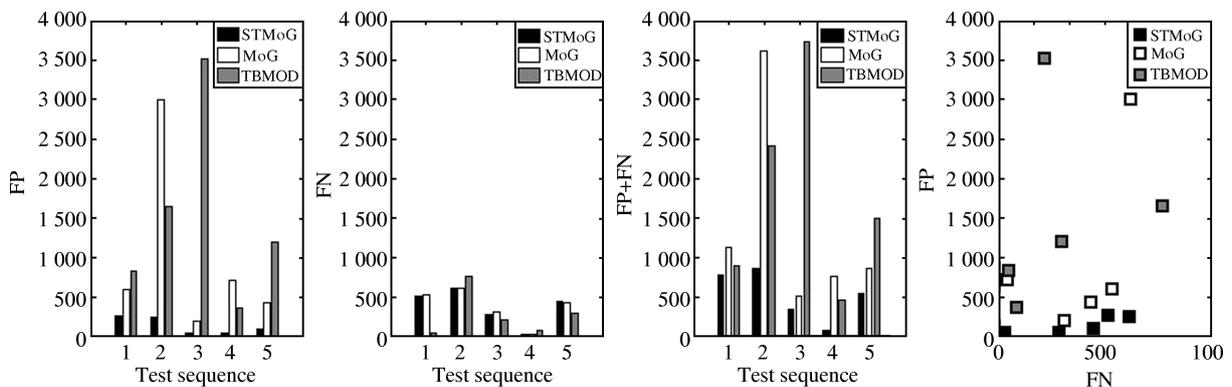


图 4 检测结果的定量比较

Fig. 4 Comparison of our method to the previous methods

显示了本文提出的时空背景建模方法在保持较低漏检数的情况下,能显著地降低虚警数,证明了本文算法的有效性。

## References

- 1 Friedman N, Russell S. Image segmentation in video sequences: a probabilistic approach. In: Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence. Providence, USA: Morgan Kaufmann, 1997. 175–181
- 2 Stauffer C, Grimson W E L. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 747–757
- 3 Kaewtrakulpong P, Bowden R. An improved adaptive background mixture model for real-time tracking with shadow detection. In: Proceedings of the 2nd European Workshop on Advanced Video Based Surveillance Systems. Providence, USA: Kluwer Academic Publishers, 2001. 1–5
- 4 Zivkovic Z, van der Heijden F. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 2006, **27**(7): 773–780
- 5 Lee D S. Effective Gaussian mixture learning for video background subtraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(5): 827–832
- 6 Power P W, Schoonees J A. Understanding background mixture models for foreground segmentation. In: Proceedings of Image and Vision Computing New Zealand. Auckland, New Zealand: Auckland University Press, 2002. 267–271
- 7 Elgammal A, Duraiswami R, Harwood D, Davis L S. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proceedings of IEEE*, 2002, **90**(7): 1151–1163
- 8 Stenger B, Ramesh V, Paragios N, Coetzee F, Buhmann J M. Topology free hidden Markov models: application to background modeling. In: Proceedings of the 8th International Conference on Computer Vision. Vancouver, Canada: IEEE, 2001. 294–301
- 9 Toyama K, Krumm J, Brumitt B, Meyers B. Wallflower: principles and practice of background maintenance. In: Proceedings of the 7th International Conference on Computer Vision. Kerkyra, Greece: IEEE, 1999. 255–261
- 10 Wang Y, Loe K F, Wu J K. A dynamic conditional random field model for foreground and shadow segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(2): 279–289
- 11 Seki M, Wada T, Fujiwara H, Sumi K. Background subtraction based on cooccurrence of image variations. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE, 2003. 65–72
- 12 Matsuyama T, Ohya T, Habe H. Background subtraction for non-stationary scenes. In: Proceedings of the 4th Asian Conference on Computer Vision. Taipei, Taiwan: University Trier Press, 2000. 114–116
- 13 Heikkilä M, Pietikäinen M. A texture-based method for modeling the background and detecting moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(4): 657–662
- 14 Chen Y T, Chen C S, Huang C R, Hung Y P. Efficient hierarchical method for background subtraction. *Pattern Recognition*, 2007, **40**(10): 2706–2715
- 15 Mason M, Duric Z. Using histograms to detect and track objects in color video. In: Proceedings of the 30th Conference on Applied Imagery Pattern Recognition Workshop. Washington D. C., USA: IEEE, 2001. 154–159
- 16 Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE, 2001. 511–518

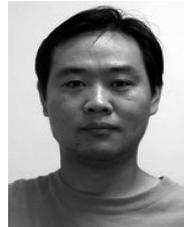


**王永忠** 博士研究生. 主要研究方向为视觉目标跟踪, 序列图像处理和智能视频监控. 本文通信作者.

E-mail: wang\_yzhong@163.com

(**WANG Yong-Zhong** Ph. D. candidate. His research interest covers visual object tracking, sequence images processing, and intelligent video

surveillance. Corresponding author of this paper.)



**梁彦** 副教授, 博士. 主要研究方向为非线性滤波, 多目标跟踪, 信息融合和图像分析. E-mail: liangyan@nwpu.edu.cn

(**LIANG Yan** Associate professor, Ph. D.. His research interest covers nonlinear filter, multi-target tracking, information fusion, and image analysis.)



**潘泉** 教授. 主要研究方向为信息融合理论及应用, 多目标跟踪, 智能信息处理, 模式识别和图像与视频处理.

E-mail: quanpan@nwpu.edu.cn

(**PAN Quan** Professor. His research interests covers theory and application of information fusion, multi-target tracking, intelligence information process, pattern recognition, and image and video process.)



**程咏梅** 教授. 主要研究方向为信息融合, 主/被动跟踪, 信号与信息处理和认知计算.

E-mail: chengym@nwpu.edu.cn

(**CHENG Yong-Mei** Professor. Her research interest covers information fusion, active/passive target tracking, signal and information processing, and cognition computation.)



**赵春晖** 讲师, 博士. 主要研究方向为机器视觉, 模式识别和图像分析.

E-mail: zhaochunhui@nwpu.edu.cn

(**ZHAO Chun-Hui** Lecturer, Ph. D.. His research interest covers computer vision, pattern recognition, and image analysis.)