

# 基于场景感知的运动目标检测方法

宋涛<sup>1</sup>,李鸥<sup>1</sup>,崔弘亮<sup>2</sup>

(1. 信息工程大学信息工程学院,河南郑州 450002;2. 南京理工大学自动化学院,江苏南京 210094)

**摘要:** 背景减除法是一种主要的运动目标检测框架,但在复杂环境中构建一种初始模型建立周期短、可靠性高、鲁棒性好的模型仍是一大难题. 本文从场景感知的角度出发,在背景减除框架的基础上提出一种目标检测方法. 该方法根据前两帧中稳定的结构信息感知背景中潜在的前景区域,在第二帧建立初始模型时利用最近邻域背景像素点代替可能的前景像素点,提高了初始模型可靠性;结合颜色信息和二进制特征提出了像素点二级分类判决机制,并通过感知像素点邻域内的纹理复杂度自适应调整局部判决阈值和更新频率;在模型更新阶段提出处理误判的反馈机制. 在公开视频序列上同几种流行检测算法的实验对比结果证明了本文算法的有效性和优越性.

**关键词:** 目标检测; 场景感知; 二进制特征; 反馈机制

**中图分类号:** TP319 **文献标识码:** A **文章编号:** 0372-2112 (2016)11-2625-08

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2016.11.009

## A Moving Object Detection Method Based on Scene Perception

SONG Tao<sup>1</sup>, LI Ou<sup>1</sup>, CUI Hong-liang<sup>2</sup>

(1. Institute of Information System Engineering, Information Engineering University, Zhengzhou, Henan 450002, China;

2. School of Automation, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China)

**Abstract:** Background subtraction algorithm is a kind of main moving object detection framework, but it is too difficult to build a model with short establishing period, high reliability and good robustness. From the perspective of scene perception, a technique for object detection based on the framework of background subtraction is proposed. To improve the reliability of initial model, the potential foreground pixels in background, which are obtained on the basis of stable structural information in the former two frames, are replaced by the nearest neighbor pixels belonging to background, when the initial model is being established in the second frame. Integrating color information with binary feature, a two-stage classification decision mechanism is proposed, meanwhile the local decision threshold and update frequency are adaptively adjusted in accordance with the texture complexity of pixels neighborhood. Subsequently, a feedback mechanism for misclassification is presented in the update model phase. Experimental results using challenging public video sequences show the effectiveness and superiority of the proposed method, compared with other state-of-the-art tracking approaches.

**Key words:** object detection; scene perception; binary feature; feedback mechanism

## 1 引言

在计算机视觉领域,运动目标检测是指在视频序列上发现并捕捉到感兴趣目标的行为,是目标跟踪、分类和行为理解等更高层次处理的关键底层技术,其检测效果将直接影响系统后续处理的各项性能. 当前,越来越多的检测算法被开发以满足不同的应用需求(比如,智能机器人、自动监控系统和人机交互等),其中应用比较广泛的是以背景减除法<sup>[1,2]</sup>为核心的运动目标

检测框架. 背景减除法通过对比背景模型和当前帧图像来检测运动目标,对简单场景有很好的检测效果. 但是,在实际复杂应用环境中,由于光照变化、噪声和动态背景等干扰因素,使得构建性能稳健的基于背景减除的运动目标检测方法成为一项极具挑战性的研究课题.

基于背景减除的目标检测方法必须要处理好以下四个关键问题<sup>[3]</sup>: (1) 如何定义并建立背景模型? 文献[4]中, Stauffer 和 Grimson 根据图像序列中单个像素点的时间分布特性提出了一种经典的高斯混合参数模型

(GMM, Gaussian Mixture Models), 但固定的模型参数  $K$  不能真实反映复杂的检测环境. 文献[5]提出了一种基于核密度估计(KDE, Kernel Density Estimation)的非参数统计模型, 该方法根据前  $N$  帧的像素点构建背景模型, 能够真实反映背景像素的概率密度分布情况, 但这种建模方法需要消耗大量存储空间和运算成本. 文献[6]利用图像分割技术将背景划分为多个子块, 构建像素级和区域级相结合的阶层背景模型, 但该方法要求初始化图像帧内不能包含运动目标, 限制了其应用范围; (2) 如何在没有任何附加条件的情况下快速有效地初始化背景模型? 近年来, 仅有少数文献讨论了背景模型的初始化问题, 文献[4]和[6]要求在模型的初始化周期内没有或仅有少量的运动目标, 这在实际应用中很难做到; 文献[3]、[7]、[8]均采用单帧初始化模型的方法, 利用邻域像素构建背景模型, 缩短了初始背景模型的建立周期, 但是这种方法不适用于纹理丰富的复杂场景, 且当初始帧包含运动目标时, 会产生错误的 ghost 目标(即不属于任何真实运动目标的前景区域); (3) 如何在分类判决阶段设计合理的判决和反馈机制? 文献[3]和[4]均采用全局阈值来分类判决像素点, 没有考虑不同局部环境存在不同程度的敏感度, 易产生误判; 且两类算法都缺乏处理误判的反馈机制, 不能迅速消除 ghost 目标问题; (4) 在模型更新阶段, 采用什么样的更新策略和更新频率? 模型更新是算法适应背景变化的关键步骤, 文献[9]采取“先入先出”的更新策略, 无法处理长期存在的目标. 文献[3]和[10]均采用固定频率的随机抽样方法来更新背景模型中的样本点, 使得采样点可以长期存在, 能够处理短期或长期存在的目标, 但该方法不适用于帧率变化较大的应用场景.

针对上述问题, 本文在背景减除框架的基础上提出一种基于场景感知的非参数化目标检测方法. 在模型初始化阶段, 不同于大多数文献的背景模型构建原则(粗建立, 细调整), 该方法通过感知前两帧中可能存在的运动目标, 利用邻域背景信息在线建立更加可靠、有效的初始模型. 在像素点分类判决阶段, 该方法以颜色信息作为第一级触发器, 通过计算邻域二进制特征给出第二级判决, 并根据像素点局部邻域的纹理复杂度自适应调整两级判决的阈值, 提高了像素点的分类精度; 同时, 根据判决结果提出了处理 ghost 目标的反馈机制. 在模型更新阶段, 根据应用场景自适应调整更新频率.

## 2 基于场景感知的背景模型初始化

为了适应背景动态变化, 许多检测方法需要训练大量的视频帧来初始化模型, 文献[4]统计每个背景像素点的时间分布模型, 当模型趋于稳定时, 初始化完成:  $B = \operatorname{argmin}_b [\sum(w_{k=1, \dots, b}) > T]$ ; 但这类方法的初始化周期过长. 另外一类是以文献[3]为代表的单帧初始化方

法, 将第一帧图像视为背景并建立模型, 即  $B = \text{Frame}_0$ ; 该类方法会使大量的前景目标融入背景模型, 在模型稳定之前会产生许多错误检测. 这两类初始化算法都属于盲分析, 只将图像帧视为一个个单独的像素点进行处理, 并没有对应用场景进行任何分析. 本文利用空时结构信息对检测场景进行感知<sup>[11]</sup>, 分析场景中是否包含运动目标及场景的复杂度, 指导背景模型的初始化.

### 2.1 基于空时结构信息的场景感知方法

本文通过 Canny 算子检测初始帧的空间结构信息, 判断背景图像的纹理复杂程度, 定位背景中的平坦区域和复杂区域, 为背景像素点建模提供先验信息.

$$\begin{cases} \varphi_k^t(x, y) = \sqrt{\varphi_1^2(x, y) + \varphi_2^2(x, y)} \\ \theta_k^t = \tan^{-1} \frac{\varphi_2(x, y)}{\varphi_1(x, y)} \end{cases} \Rightarrow \begin{cases} c_k^t(x, y) = 1, & \text{complex region} \\ c_k^t(x, y) = 0, & \text{flat region} \end{cases} \quad (1)$$

其中,  $\varphi_k^t$  和  $\theta_k^t$  分别表示  $t$  帧中第  $k$  个像素点的梯度幅值和方向;  $c_k^t$  是边缘检测结果,  $c_k^t = 1$  表示该像素点属于边缘结构信息, 其邻域的纹理比较复杂; 反之,  $c_k^t = 0$  表示该像素点位于纹理较为平坦的区域, 且该像素点与其邻域内像素点有相似的时间分布模型<sup>[7]</sup>.

从图 1(b) 中可以看出, Canny 算子不但可以检测到初始帧图像精确的结构信息, 而且通过计算  $\text{Sum}^0 = \sum(c_{k=1, \dots, N}^t)$  还可以推断出背景图像的纹理复杂程度(图 1 的第一行序列中  $\text{Sum}^0 < \text{threshold}$  为简单场景, 第二行序列中  $\text{Sum}^0 > \text{threshold}$  为复杂场景), 为模型的初始化提供先验信息. 与初始帧相比, 第二帧图像内的真实背景结构信息仅发生了细微变化, 而运动目标所在区域的局部结构则产生了较大变化, 如图 1(c) 所示. 通过对前两帧检测结果进行异或运算, 去除背景图像中稳定的结构信息, 得到运动目标可能存在的区域, 见图 1(d); 异或运算体现了融合的概念, 异或图像中既有单帧图像的空间信息, 又包含连续帧图像的时间信息, 本文称之为空时结构信息. 对异或图像进行后期处理, 突出显示运动目标可能存在的区域, 如图 1(e) 中的白色区域所示.

### 2.2 背景模型的初始化

参数背景建模方法存在一定的理论缺陷: 参数统计需要较长的统计周期、无法自适应确定所需参数的个数、通常需要假设背景像素值比前景值出现的更加频繁; 不适用于复杂的应用环境. 本文在场景感知的基础上, 采用样本集合的非参数方法构建背景模型  $B(\mathbf{x})$ , 依据背景复杂度  $\text{Sum}^0$  自适应决定所需样本个数  $N$  (场景越复杂  $N$  越大, 反之  $N$  越小).

$$B(\mathbf{x}) = \{b_1(\mathbf{x}), b_2(\mathbf{x}), \dots, b_N(\mathbf{x})\} \quad (2)$$

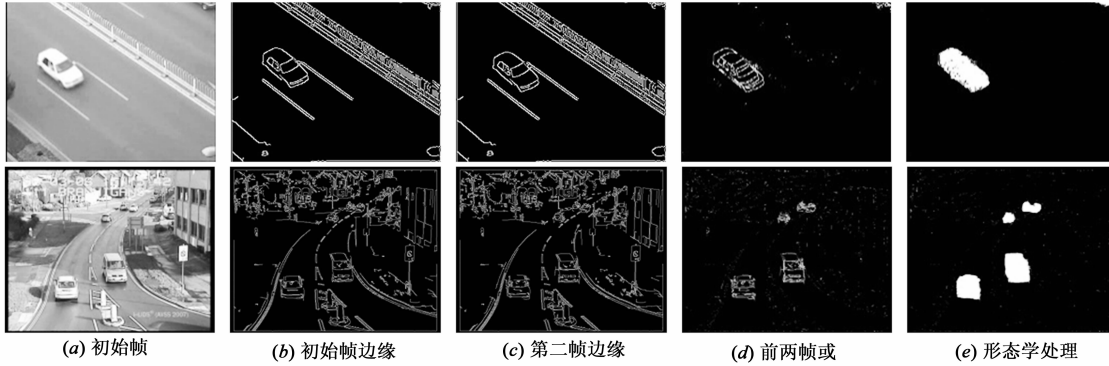


图1 场景感知示意图(第一行序列属于简单应用背景;第二行序列属于复杂应用背景)

其中,  $\mathbf{x}$  表示背景像素点的位置坐标; 每个样本  $\mathbf{b}_i(\mathbf{x}) = [v_i, e_i, w_i]_{i=1, \dots, N}$  分别表示该样本点的颜色信息、二进制特征(本文采用文献[8]的思想改进 LBP 纹理特征, 利用邻域像素之间的相似度描述像素点  $\mathbf{x}$  与其周围像素的空间信息)和样本点的权值。

本文根据应用场景的空时结构信息指导背景模型初始化, 如图 1(e) 所示: 第 2 帧内的白色区域表示运动目标可能出现的前景区域, 黑色区域表示背景, 分别用  $l_x^2 = 1$  和  $l_x^2 = 0$  表示。为了防止在背景模型中引入前景信息, 本文在第 2 帧中将所有可能存在运动前景和噪声干扰的区域均赋予  $l_x^2 = 1$ , 并利用其最近邻域内  $l_x^2 = 0$  的像素值建立初始模型; 对于  $l_x^2 = 0$  的像素点, 根据其自身的像素值构建模型。初始背景模型如下式所示:

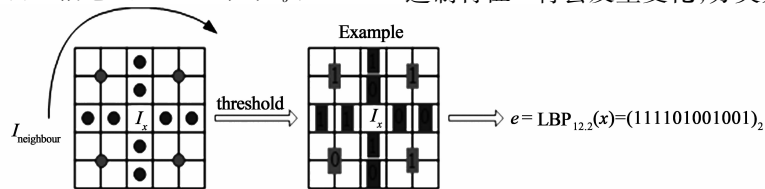
$$\begin{aligned} \mathbf{B}^0(\mathbf{x}) &= \{\mathbf{b}_1(\mathbf{x}) = [v_1, e_1, w_1], \\ &\mathbf{b}_2(\mathbf{x}) = 0, \dots, \mathbf{b}_N(\mathbf{x}) = 0\} \end{aligned} \quad (3)$$

其中,  $v_1$  为第 2 帧中像素点  $\mathbf{x}$  或其最近邻域像素点的原始颜色信息, 对像素值的变化比较敏感, 能精细捕捉颜色变化;  $w_1 = 1/N$  是样本点  $\mathbf{b}_1(\mathbf{x})$  的初始权值;  $e_1$  为第 2 帧中像素点  $\mathbf{x}$  或其最近邻域像素点的 12bit 局部二进制特征信息  $LBP_{12,2}$ , 对噪声、光照变化有较强的鲁棒性。

如图 2 所示, 在中心像素点  $I_x$  的  $5 \times 5$  邻域内, 用 12bit 二进制特征  $e$  近似表示 25bit 的局部邻域信息  $I_{\text{neighbour}}$ , 使特征的表达式更加简洁有效。其中, 四个对角上的灰色像素点是通过其 4 邻域像素利用双线性插值法确定其值; 阈值决定过程由式(4)计算得出:

$$\text{Binary}(i)_{i=0, \dots, 11} = \begin{cases} 1 & \text{if } |I_x - I_i| \leq T_b \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

其中,  $I_{i=0, \dots, 11}$  表示中心点像素周围的 12 个邻域像素;  $| \cdot |$  表示两个像素点之间颜色信息的绝对差值;  $T_b$  是

图2 局部二进制模式  $LBP_{12,2}$  示意图

度量邻域像素点之间相似度的阈值, 通常设为  $T_b \approx 0.5v_x$ 。

### 3 像素点的分类判决过程

本文采用与背景像素建模相同的方法表示第  $t$  输入帧像素点  $\mathbf{x}$ , 如式(5)所示:

$$\mathbf{p}^t(\mathbf{x}) = [v_x^t, e_x^t, c_x^t, l_x^t] \quad (5)$$

其中,  $v_x^t$  表示像素点  $\mathbf{x}$  的颜色信息;  $e_x^t = LBP_{12,2}^t(\mathbf{x})$  表示像素点  $\mathbf{x}$  的二进制特征;  $c_x^t = \text{sum}(c_{i=1, \dots, 25}^t)$  统计了像素点  $5 \times 5$  邻域内的边缘信息, 反映像素点  $\mathbf{x}$  邻域的纹理复杂程度;  $l_x^t$  是表示像素点  $\mathbf{x}$  属于前景或背景的标志(前景  $l_x^t = 1$ , 背景  $l_x^t = 0$ )。

文献[3~6]对比输入帧与背景模型像素点之间的颜色信息, 能够捕捉细微的前景变化区域, 但该方法无法处理复杂场景下动态干扰和光照变化等问题。文献[8]、[11]和[12]利用二进制特征来检测前景运动目标, 对光照变化有较强的鲁棒性, 但该方法无法处理纹理比较平坦的应用环境。本文采用颜色信息  $v$  和二进制特征  $e$  相结合的方法检测前景变化, 需要解决图 3 所示的几个难点。

一般情况下, 像素点分类过程主要是为了判断单独背景像素点、靠近前景目标的背景像素点、位于前景目标边缘的前景像素点和前景目标内像素点(分别对应图 3 中的像素点 1~4)是否属于背景模型的方法。其中, 像素点 1 是位于平坦区域的背景像素点, 在分类判决中易受到光照影响, 使得颜色信息  $v$  的变化大于  $T_v$ , 但其二进制特征  $e$  变化较小, 类似于像素点 4, 易被误判为前景; 像素点 2 的  $5 \times 5$  邻域内引入了前景信息, 其二进制特征  $e$  将会发生变化, 分类判决结果取决于阈值

$T_e$  的大小;像素点 3 的特征  $v$  和  $e$  均发生了变化,判决结果取决于阈值  $T_v$  和  $T_e$ ;像素点 4 位于平坦的前景区域内,如果仅依赖分类方法,易与受到光照变化影响的

像素点 1 发生混淆,产生误判. 所以,阈值  $T_v$  和  $T_e$  的设置及辅助机制是分类过程的关键.

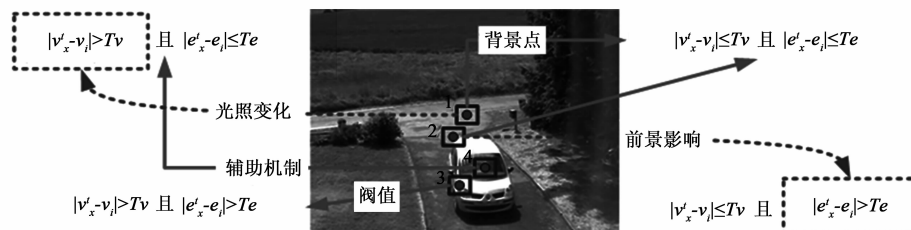


图3 像素点分类所要解决的问题

为了使背景模型更加精确有效,本文将特征  $v$  的变化检测视为一级触发机制,将  $T_v$  赋予较小的值,只有满足  $\#\{ |v'_x - v_x| \leq T_v, \forall_{i=1, \dots, N} \} \geq \#_{\min}$  的像素点才可以作为候选背景像素点,否则为候选前景像素点,其中  $\#_{\min}$  表示所需满足条件的最小样本个数. 将特征  $e$  的变化检测视为二级验证机制  $|e'_x - e_x| \leq T_e$  (对于二进制特征  $e$  而言,  $||$  表示汉明距离), 阈值  $T_e$  的大小直接影响检测结果: 如果  $T_e$  过大,则无法检测到与背景颜色相似的前景目标;如果  $T_e$  过小,图像噪声和动态背景干扰将被判为前景目标,无法融入背景模型. 文献[3]和[10]采用全局阈值,但实际应用场景中背景像素的变化并不一致,全局阈值会导致误判. 本文利用场景感知方法自适应地调整局部  $T_e$ , 计算输入帧各个像素点的纹理复杂度统计  $c'_x$ : 如果  $c'_x \leq a$  表示像素点  $x$  处于平坦区域,此时背景模型比较稳定可靠,通过减小  $T_e$  来增强特征分类的敏感性,以检测出与背景相似的前景目标;如果  $c'_x > a$  则表示像素点  $x$  所处环境比较复杂,通过增大  $T_e$  来提高背景模型的包容能力.

本文所提的分类方法:(1)当像素点  $x$  为候选背景像素点时,如果  $|e'_x - e_x| \leq T_e$ , 则  $l'_x = 0$  为背景像素点;如果  $|e'_x - e_x| > T_e$ , 则  $l'_x = 1$  为前景像素点,且此前景像素点是与背景相似的伪目标.(2)当像素点  $x$  为候选前景像素点时,如果  $|e'_x - e_x| > T_e$ , 则  $l'_x = 1$  为前景像素点;如果  $|e'_x - e_x| \leq T_e$  且  $|c'_x - c_{x-1}'| \leq T_c$ , 则  $l'_x = 0$  为背景像素点,此颜色信息  $v$  的变化是由光照变化引起;如果  $|e'_x - e_x| \leq T_e$  且  $|c'_x - c_{x-1}'| > T_c$ , 则  $l'_x = 1$  为前景像素点.

#### 4 背景模型的更新方法

本文采用类似于文献[3]的保守更新策略,只将被判为背景的像素点 ( $l'_x = 0$ ) 融入模型,当  $l'_x = 1$  时则保持原有的背景模型.

##### 4.1 保守更新策略存在的问题和解决方案

保守更新策略使得背景模型中不存在任何前景像素点,其可靠性较其他盲更新策略更高;但保守更新策略没有解决像素点误判的机制,背景像素点一旦被误

判为前景就不能再融入背景模型、无法处理 ghost 目标和长期静止的目标. 为了解决上述问题,本文根据像素点的分类判决结果和邻域像素点值域分布一致性原则<sup>[7]</sup> 提出一种处理像素点误判的反馈机制.

图4显示了 ghost 目标的形成过程(第一行是视频序列,第二行是对应的检测结果). 在模型初始化阶段,处于静止状态的前景目标将作为背景融入模型;当该目标开始运动时,其原来覆盖的背景和目标本身都将被检测为前景目标,分别对应图4第56帧中的灰色和黑色矩形框;保守更新策略无法将前景信息融入模型(如图4第175帧中的灰框区域),ghost 目标将一直存在.

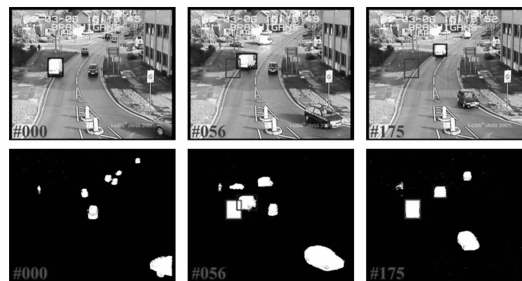


图4 ghost 目标(黑框为前景目标,灰框为ghost目标)

文献[7]从实验和理论上证明了背景局部邻域像素点的值域具有连续性. 基于此,本文对下一帧中被判为前景的像素点  $\mathbf{p}^{t+1}(\mathbf{x}) = [v_x^{t+1}, e_x^{t+1}, c_x^{t+1}, l_x^{t+1} = 1]$  进行二次分析,给出一种反馈机制来解决 ghost 目标问题. 首先,利用  $c_x^{t+1}$  确定目标边界,根据前景像素点的位置坐标由外向内逐步计算该前景像素点与其最近邻背景像素点  $\mathbf{p}^{t+1}(\mathbf{x}^*)$  之间的相似度: 如果两像素点间相似度大于阈值  $T_g$ , 则判断该前景像素点是属于背景的 ghost 目标,此时令  $l_x^{t+1} = 0$  并继续按该方向向内计算;如果像素点间相似度小于  $T_g$  且  $c_x^{t+1} = 1$ , 则判断该前景像素点属于目标且为目标边界,停止计算. 根据计算得到的目标边界像素点近似确定前景目标的几何区域,区域内部的像素点均属于所要跟踪的前景运动目标.

$$\begin{aligned} p^{i+1}(x^*) &= \underset{I_{x^*}}{\operatorname{argmin}} d[p^{i+1}(x), I_{x^*}], \\ \text{s. t. } I_{x^*} &\in \text{背景} \end{aligned} \quad (6)$$

对于停止运动并长期静止的目标,本文给出一种折中的处理方法:为每个检测到的前景目标设定一个计时器,如果目标停留在同一个位置的时间超过预设阈值,就将该目标融入背景;反之,保留此前景目标。

#### 4.2 模型样本更替方法和更新频率

样本更替方案及更新频率是背景模型是否稳健的关键,本文认为样本的重要性与其出现的时间先后顺序无关,而与其出现次数有关;并为每个样本赋予权值  $w_i$  来度量其重要程度,权值大小与样本出现频率呈正比,所以权值最小的样本最先被替换。在整个背景图像区域内,不同位置像素点经历的变化程度不同,本文根据每个背景像素点的局部纹理复杂度自适应调整其更新频率  $F(x)$ :如果  $l'_x = 0$  且  $c'_x \leq T_c$ ,则  $F(x) = 1/(c'_x + a)^2$ ,其中  $a > 1$ ,该式表示对纹理平坦的背景像素点赋予较小的更新频率;如果  $l'_x = 0$  且  $c'_x > T_c$ ,则  $F(x) = 1$ ,该式表示对纹理复杂的背景像素点进行逐帧更新。

### 5 实验结果与分析

为了验证所提算法的有效性,本文选择 CDnet2014 (the ChangeDetection.net Dataset)<sup>[13]</sup> 提供的带有真实目标检测结果的公共标准图像序列进行测试。为了证明所提算法在模型初始化和抑制 ghost 目标方面的优越性,本文与两类具有代表性的目标检测算法 ViBe<sup>[3]</sup> 和 GMM<sup>[4]</sup> 进行了两组对比实验;并通过实验 3 证明了所提算法在不同复杂应用场景下的鲁棒性。所有相关实验都是在 Intel 双核 2.2GHz CPU、2.0GB 内存的 PC 平台上,利用 C/C++、Matlab 和 OpenCV 函数库实现的。实验中的参数设置:GMM 的模式数  $K = 2$ 、两个学习率  $a_1 = a_2 = 0.2$ 、模型阈值  $T = 0.8$ ;ViBe 的粒子数  $N = 20$ 、邻域范围  $R = 20$ 、阈值  $\#_{\min} = 2$ 、子抽样时间  $t = 1$ ;本文算法的粒子数  $N = 10$ 、阈值  $\#_{\min} = 1$ 、 $T_c = 6$ 、学习率  $a = 3$ ,其他参数  $T_v$ 、 $T_c$  由不同的实验环境具体计算。

本文利用常见的目标检测性能评价标准:召回率 Recall、精确度 Precision 和  $F_1$ -Measure,对上述对比实验进行定量分析。

$$\begin{cases} \text{Recall} = R = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{\text{ground truth}} \\ \text{Precision} = P = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{\text{TP}}{\text{detector}} \\ F_1 - \text{Measure} = F_1 = \frac{2 \cdot P \cdot R}{P + R} \\ = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \approx \frac{\text{ground truth} \cap \text{detector}}{\text{ground truth} \cup \text{detector}} \end{cases} \quad (7)$$

其中,TP 表示被正确检测为感兴趣目标的像素点个数;FN 表示被误判为背景的像素点个数;TP + FN 即为感兴趣目标的真实大小;FP 表示被误判为感兴趣目标的背景像素点个数;TP + FP 即为检测算法得到的前景目标大小。式(7)中,Recall 为检测算法得到的正确分类数占真实目标的比例;Precision 体现了检测算法的正确率;这两个指标均比较片面,无法真实反映检测算法的整体性能。从检测-跟踪的角度出发,本文认为检测算法首要任务是发现目标,即检测得到的前景区域中包含正确分类 TP 越大越好,为后续匹配跟踪、识别提供可靠的初始信息,所以对跟踪任务而言,Precision 是一个重要指标。 $F_1$ -Measure 是度量检测算法性能的综合指标,反映了检测算法的灵敏性和有效性。

**实验 1** 定性分析了本文算法、GMM 和 ViBe 在背景模型初始化阶段的检测性能,实验场景是 CDnet2014 的三段交通监控图像序列,检测目标包括车辆和行人;ViBe 直接以初始帧为背景,本文算法根据前两帧计算背景模型、GMM 根据预设的高斯模式数来计算背景。从图 5(c) 可以看出,GMM 无法在模型建成的初期有效检测运动目标,这是因为初始帧背景中存在运动目标,导致背景模型包含了大量前景目标像素。ViBe 以初始帧为背景,在模型建立初期近似于最简单的背景减法,会造成大量 ghost 目标,如图 5(d) 所示。本文算法根据前两帧的空时信息,能够准确定位潜在的运动目标,利用最近邻域的背景像素信息对可能存在运动目标的区域进行建模,从图 5(e) 可以看出,本文算法在模型形成初期就具有很好的检测效果。

为了进一步定量分析本文算法、GMM 和 ViBe 在实验 1 中检测结果的准确性和可靠性,本文根据召回率  $R$ 、精确度  $P$  和  $F_1$  对各算法的检测结果进行了统计平均,如表 1 所示。实验 1 是为了验证各算法在模型建立初期的检测性能,从表 1 可以看出:GMM 的  $P$  指标较高但其他两项指标均非常低,这是因为 GMM 的初期模型中包含了大量前景像素,导致检测结果存在空洞;ViBe 的三项指标都比较低,平均小于 30%,这是由 ghost 目标造成的;本文算法的各项指标均比较高,且  $P$  指标接近 85%,说明本文算法的初始模型具有较高的可靠性,能在应用初期快速有效地捕捉到运动目标。

对于采用背景减法框架的检测算法而言,如果背景中预先存在的静止目标开始运动,就一定会出现 ghost 目标问题。因此,能否快速有效地解决 ghost 目标是衡量一个检测算法优劣的重要标准。

**实验 2** 本文采用的三段 CDnet2014 图像序列均在初始帧中包含感兴趣目标,当目标开始运动时就会出现 ghost 目标(灰色矩形框)。从图 6 第一列关于“streetLight”的检测结果可以看出,GMM 和 ViBe 算法

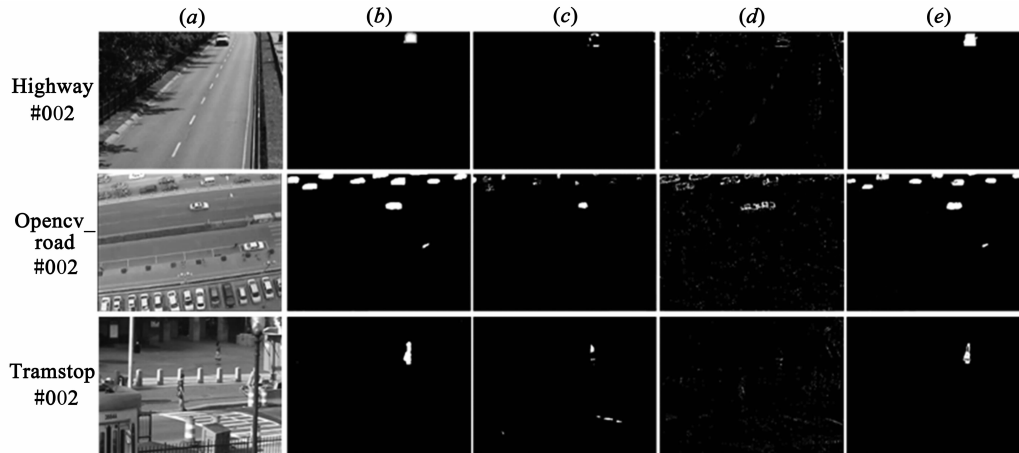


图5 实验1 三种算法在模型初始化阶段的检测结果对比图 (a) CDnet2014三个图像序列“Highway”、“Opencv\_road”和“Tramstop”的第二帧图像; (b) 真实目标检测结果; (c) GMM检测结果; (d) ViBe检测结果; (e) 本文算法检测结果

表1 实验1 中三种检测算法的  $R$ 、 $P$  和  $F_1$  指标对比

运动目标 检测算法	“Highway”			“Opencv_road”			“Tramstop”		
	$R$	$P$	$F_1$	$R$	$P$	$F_1$	$R$	$P$	$F_1$
GMM	0.137	0.819	0.235	0.373	0.695	0.485	0.468	0.563	0.511
ViBe	0.156	0.294	0.204	0.232	0.374	0.286	0.125	0.038	0.058
本文算法	0.834	0.857	0.845	0.821	0.845	0.833	0.786	0.819	0.802

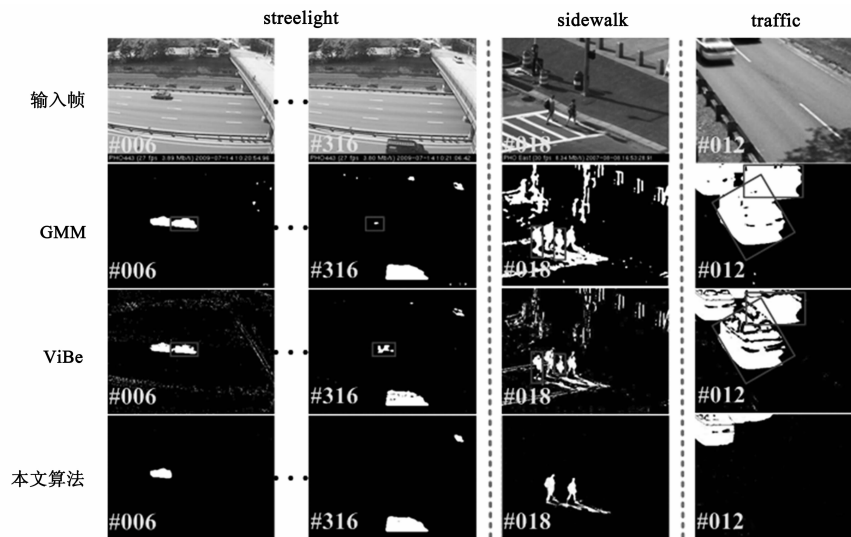


图6 实验2 三种检测算法处理ghost目标的对比图; 从左到右列依次为CDnet2014的三段序列“streetLight”、“sidewalk”和“traffic”; 从第一行到第四行依次为输入帧图像、GMM检测结果、ViBe检测结果和本文算法检测结果

都在第6帧出现了完整的ghost目标,其中GMM直到第316帧才将ghost目标的大部分区域清除,而ViBe的ghost问题在第316帧依然存在。图6第二和第三列关于“sidewalk”和“traffic”的检测结果同样验证了GMM和ViBe无法快速去除ghost目标,其中GMM和ViBe分别在“sidewalk”、“traffic”场景下直到第339、380帧和第343、1083帧才清除ghost目标。本文算法根据反馈机制对前景目标进行二次判决,能够直接解决ghost目标,

如图6第四行所示,本文算法在检测过程中并没有出现ghost目标。

**实验3** 定性分析了本文算法在不同复杂应用环境下的目标检测能力,实验场景是CDnet2014的四段图像序列,如图7所示。“blizzard”列显示了本文算法在恶劣天气条件下的检测结果,从图7第三行可以看出,虽然汽车与周围环境的相似度较高,但本文算法依然可以正确检测到目标;表2中的三项指标均大于80%。

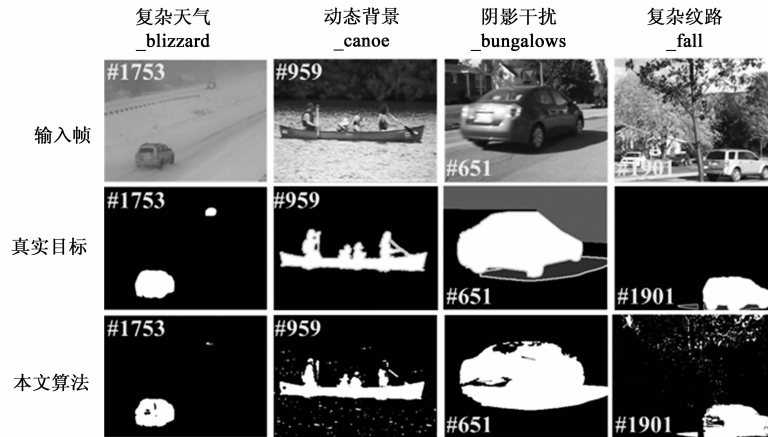


图7 实验3 本文算法在不同复杂应用场景下的检测结果; 从左到右列依次为CDnet2014的四段序列“blizzard”、“canoe”、“bungalows”和“fall”; 从第一行到第三行依次为输入帧图像、真实目标和本文算法检测结果

“canoe”和“fall”列均属于背景纹理复杂且存在动态背景干扰的应用场景,从图7第三行可以看出,本文算法可以检测到大部分目标( $R$ 指标较高),但检测结果中存在较多椒盐噪声干扰( $P$ 指标较低).本文算法在

“bungalows”列的检测结果中将目标阴影判决为前景目标,不能去除阴影干扰,导致表2中相关的 $P$ 和 $F_1$ 指标较低(平均值接近50%).

表2 实验3中本文算法在四种不同应用环境下的 $R$ 、 $P$ 和 $F_1$ 指标对比

“blizzard”			“canoe”			“bungalows”			“fall”		
$R$	$P$	$F_1$	$R$	$P$	$F_1$	$R$	$P$	$F_1$	$R$	$P$	$F_1$
0.811	0.863	0.836	0.847	0.772	0.807	0.829	0.598	0.694	0.848	0.532	0.653

## 6 结论

本文利用图像帧的空时信息提出一种基于场景感知的运动目标检测方法.该方法通过异或前两帧图像的结构信息捕捉到背景中可能存在的运动区域,去除了背景中的前景信息,提高了初始模型的可靠性.为了增强检测算法的有效性,该方法融合像素点的颜色和二进制特征提出了一种二级分类判决方案,并根据像素点邻域的纹理复杂程度自适应调整局部判决阈值和样本更新频率;根据判决结果给出了处理误判的反馈机制.实验结果表明,相较于GMM和ViBe两类算法,本文算法的背景模型在建立初期就具有较高的检测精度,有助于迅速发现并捕捉到较为完整的感兴趣目标,且算法能够有效抑制ghost目标问题;实验3证明了本文算法在四类不同复杂应用环境下依然能够有效检测运动目标,但在目标阴影和动态背景干扰的复杂场景中存在较多误判.在未来工作中,本文将重点研究解决目标阴影和动态背景干扰的方法,进一步提高算法的可靠性.

### 参考文献

- [1] Radke R J, Andra S, Al-Kofahi O. Image change detection algorithms: a systematic survey[J]. IEEE Transactions on Image Processing, 2005, 14(3): 294–307.
- [2] Wang B, Dudek P. A fast self-tuning background subtraction algorithm[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops[C]. Columbus OH: IEEE Press, 2014. 401–404.
- [3] Van Droogenbroeck M, Barnich O. ViBe: a disruptive method for background subtraction[A]. Proceedings of the Background Modeling and Foreground Detection for Video Surveillance[C]. USA: CRC Press, 2014. 1–23.
- [4] Stauffer C, Grimson W. Adaptive background mixture models for real-time tracking[A]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C]. Fort Collins: IEEE Press, 1999. 246–252.
- [5] Martins P, Caseiro R, Batista J. Non-parametric Bayesian constrained local models[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Columbus OH: IEEE Press, 2014. 1797–1804.
- [6] Chen S Y, Zhang J H, Li Y F. A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction[J]. IEEE Transactions on Industrial Informatics, 2012, 8(1): 118–127.
- [7] Jodoin P, 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.

- [8] St-Charles P, Bilodeau G, Bergevin R. SuBSENSE: a universal change detection method with local adaptive sensitivity[J]. IEEE Transactions on Image Processing, 2015, 24(1): 359–373.
- [9] Cuevas C, Mohedano R, Garcia N. VersatileBayesian classifier for moving object detection by non-parametric background-foreground modeling[A]. Proceedings of the 19<sup>th</sup> IEEE International Conference on Image Processing[C]. Orlando FL: IEEE Press, 2012. 313–316.
- [10] Van Droogenbroeck M, Paquot O. Background subtraction: experiments and improvements for ViBe[A]. Proceedings of the IEEE Computer Vision and Pattern Recognition Workshops[C]. Providence RI: IEEE Press, 2012. 32–37.
- [11] Wu J, Rehg J M. CENTRIST: a visual descriptor for scene categorization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1489–1501.
- [12] St-Charles P L, Bilodeau G A. Improving background subtraction using local binary similarity patterns[A]. Proceedings of the IEEE Winter Conference on Applications of Computer Vision[C]. Steamboat Springs CO: IEEE Press, 2014. 509–515.
- [13] Wang Y, Jodoin P M, Porikli F. CDnet 2014: an expanded change detection benchmark dataset[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops[C]. Columbus OH: IEEE Press, 2014. 393–400.

#### 作者简介



宋涛男, 1983年出生于河南焦作, 现为解放军信息工程大学信息工程学院博士研究生, 主要研究方向为计算机视觉和目标跟踪技术。

E-mail: taosong\_1983@126.com



李鸥男, 1961年出生于河南郑州, 现为解放军信息工程大学信息工程学院教授、博士生导师, 主要研究方向为无线通信网络、信息融合和目标跟踪技术。

E-mail: zzliou@126.com