

基于多特征的复杂场景运动目标检测

朱文杰, 王广龙, 田杰, 乔中涛, 高凤岐

陆军工程大学纳米技术与微系统实验室, 河北 石家庄 050003

摘要 为提高复杂场景下运动目标检测的完整性和准确性, 提出了一种多特征结合的运动目标检测方法。提出了一种自适应的高斯混合建模算法对颜色特征进行建模; 通过滞后多阈值建模的方法, 同时利用颜色和改进的局部二值模式纹理特征对环境背景进行了建模, 并采用邻域补偿策略将基于两种特征提取得到的目标区域进行了结合; 采用结合 Canny 思想改进的 Kirsch 方法进行了边缘提取, 消除了鬼影误识别像素, 改善了前景目标边缘。实验结果表明, 所提方法在运动目标检测的完整性、准确性等指标上优于传统算法的, 实时性也较好。

关键词 测量; 运动目标检测; 多特征; 背景建模; 纹理; 滞后多阈值

中图分类号 TP391

文献标识码 A

doi: 10.3788/AOS201838.0612004

Detection of Moving Objects in Complex Scenes Based on Multiple Features

Zhu Wenjie, Wang Guanglong, Tian Jie, Qiao Zhongtao, Gao Fengqi

Laboratory of Nanotechnology and Micro System, Army Engineering University,
Shijiazhuang, Hebei 050003, China

Abstract In order to enhance the integrity and accuracy of moving object detection in complex scenes, a multi-features-based moving object detection method is proposed. The color feature is modeled by using the proposed adaptive Gaussian mixture model (GMM) algorithm. A kind of hysteresis multi-thresholds modeling method is used to model the scene background by adopting the color and improved local binary pattern (LBP) texture feature simultaneously. A neighborhood compensation strategy is adopted to combine the object regions obtained by the two-features extraction. The improved Kirsch edge detection method combined with the Canny thoughts is adopted in the edge extraction which eliminates the mistakenly detected ghost pixels and improves the edges of foreground objects. The experimental results show that the proposed method is superior to the traditional algorithms in the detection integrity and accuracy, and the real-time performance is also better.

Key words measurement; moving objects detection; multiple features; background modeling; texture; hysteresis multi-thresholds

OCIS codes 120.1880; 110.2970; 110.4155

1 引 言

运动目标检测的目的是尽可能完整且准确地将感兴趣的运动前景目标像素区域从环境背景中分离出来。运动目标检测在安防监控、无人驾驶、人机交互、目标识别、跟踪、场景分析等领域^[1-2]发挥了重要作用, 且其检测结果决定了后续处理的难度和最终效果。

现有的主要检测方法^[3]包括帧差法、背景差法、光流法, 它们大都基于单个特征如颜色或纹理等进行分析处理。由于实际应用场景的复杂性, 基于单个特征的检测算法通常难以适应环境, 提取出的前景目标经常不完整, 或含有空腔区域和背景干扰像素, 或形成零散的像素块, 导致检测虚警, 影响高层应用的效果, 甚至导致高层应用失败或产生错误的结果。

收稿日期: 2017-12-27; 收到修改稿日期: 2018-01-31

基金项目: 国家自然科学基金(61274125, 61176012)

作者简介: 朱文杰(1988—), 男, 博士研究生, 主要从事运动目标检测方面的研究。E-mail: zwj_881218@126.com

导师简介: 王广龙(1964—), 男, 教授, 博士生导师, 主要从事精密仪器与微系统方面的研究。

E-mail: 815427360@qq.com(通信联系人)

Chen 等^[2]将连续两帧图像边缘差分,得到的前景与局域二值模式(LBP)背景相减,将得到的前景通过逻辑与进行结合,得到了最终的检测结果。该算法运算量小,便于硬件实现,但其颜色与纹理特征的结合策略过于简略,容易导致检测结果缺失,或者会扩大目标像素区域,导致检测结果不准确。Kumar 等^[4]将颜色特征和纹理特征组成一个长特征向量并进行背景建模,但其建模、更新及判定的过程实质上还是各个特征单独进行,特征间的结合不紧密。Chua 等^[5]采用模糊策略,将颜色特征相似度的权重和 LBP 纹理相似度的权重进行结合,相较于传统的定权重结合方法,该方法取得了更好的检测效果,但其自适应性较差。宋涛等^[6]利用颜色信息和 LBP 纹理信息进行二级分类判决,并以基于时空结构的空感知信息作为先验,辅助指导背景建模及模型参数的动态调整。该方法的模型初始化建立较快,但对高动态背景及阴影引起的干扰误判率较高。

颜色特征含有丰富的场景信息,是最常被采用的一类特征。基于颜色特征的背景建模方法大都根据单个像素的时间分布统计信息进行建模,通常假定观察得到的像素时间序列相互独立,忽略了像素间的空间关联,导致基于颜色特征的运动目标检测方法对光照变化、噪声及背景运动引起的干扰比较敏感,限制了其在动态背景中的应用。

相比于颜色特征,利用纹理基于区域特征进行建模的方法,对光照的变化不敏感,并能反映图像颜色的空间排列规律,可作为颜色信息的有力补充,在一定程度上能避免将动态变化的背景像素误识别为前景目标。而目标的边缘信息对目标的形状、轮廓的形成及目标的判定和识别具有重要的意义,通常被用作目标分析和识别的基础特征,同时其对光照变化及均匀纹理引起的干扰具有较强的稳健性。获得准确、连续的目标边缘对分割出准确的运动目标具有重要意义。

因此,本文从特征选择与结合运用的角度出发,提出了一种将颜色、纹理及边缘信息等多特征有机结合的运动目标检测方法。在空间上,利用滞后多阈值处理及邻域像素补偿策略,建立了基于颜色特征的改进的自适应高斯混合模型(GMM),并与改进的基于纹理特征的 LBP 建模及边缘特征进行结合,抑制了鬼影,优化了检测结果;在时间上,利用目标的运动信息进行了补偿。

2 多特征结合的运动目标检测方法

算法的整体框图如图 1 所示,其中 t 代表时间,SC 表示辅颜色前景,PC 表示主颜色前景,PT 表示主纹理前景。以下将结合该框图对每一部分进行详细阐述。

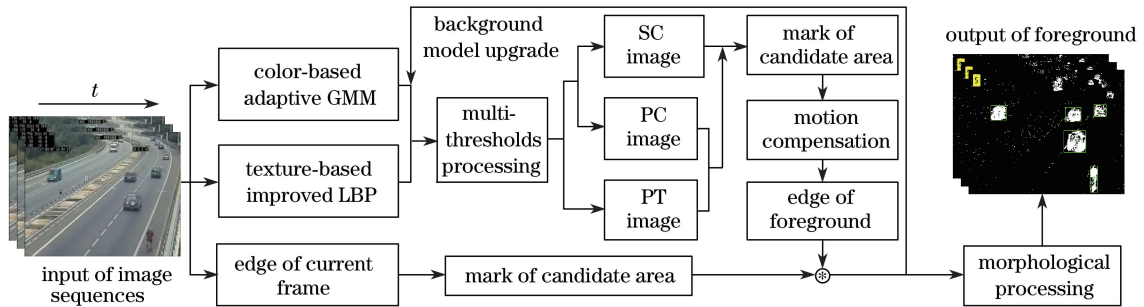


图 1 算法整体框图

Fig. 1 Overall block diagram of algorithm

2.1 基于颜色特征的 GMM 背景建模

位置 (i, j) 处的像素在 t 时刻取值 $I_{(i,j),t}$, 满足的概率分布用加权高斯函数和的形式描述, 表示为

$$P[I_{(i,j),t}] = \sum_{n=1}^K \omega_{(i,j),t,n} \eta[I_{(i,j),t}; \mu_{(i,j),t,n}, \Sigma_{(i,j),t,n}], \quad (1)$$

式中 K 为 GMM 中高斯分布的个数, 一般选为 3~5; $\omega_{(i,j),t,n}$ 、 $\mu_{(i,j),t,n}$ 、 $\Sigma_{(i,j),t,n}$ 分别为描述 t 时刻 (i, j) 处像素的第 n 个高斯分布的权重、均值和协方差矩阵; η 为高斯概率密度函数。为简化计算, 通常假定

协方差矩阵为

$$\Sigma_{(i,j),t,n} = \sigma_{(i,j),t,n}^2 \mathbf{E}, \quad (2)$$

式中 \mathbf{E} 为单位矩阵, $\sigma_{(i,j),t,n}$ 为高斯分布的标准差。

由于视频序列是一个动态变化的过程, GMM 模型也需要随之进行更新和维护, 即

$$|I_{(i,j),t} - \mu_{(i,j),t,n}| < D\sigma_{(i,j),t,n}, \quad (3)$$

式中 D 为置信度参数阈值。由 (3) 式可知, 若新输入的像素 $I_{(i,j),t}$ 与某一个高斯项均值之差在 $D\sigma_{(i,j),t,n}$ 范围内, 则 $I_{(i,j),t}$ 与该高斯项匹配。对匹配的高斯项的参数用无限冲激响应(IIR)滤波器^[7]

进行更新:

$$\omega_{(i,j),t,n} = (1 - \alpha)\omega_{(i,j),t-1,n} + \alpha, \quad (4)$$

$$\boldsymbol{\mu}_{(i,j),t,n} = (1 - \beta)\boldsymbol{\mu}_{(i,j),t-1,n} + \beta\mathbf{I}_{(i,j),t}, \quad (5)$$

$$\sigma_{(i,j),t,n}^2 = (1 - \beta)\sigma_{(i,j),t-1,n}^2 + \beta[\mathbf{I}_{(i,j),t} - \boldsymbol{\mu}_{(i,j),t,n}]^T[\mathbf{I}_{(i,j),t} - \boldsymbol{\mu}_{(i,j),t,n}], \quad (6)$$

式中 α, β 为参数学习率, T 代表转置。其余的高斯项只对权重进行衰减, 表达式为

$$\omega_{(i,j),t,n} = (1 - \alpha)\omega_{(i,j),t-1,n}。 \quad (7)$$

如果没有高斯项与 $\mathbf{I}_{(i,j),t}$ 匹配, 将所有高斯项中权重系数最小的高斯项对应的权重系数用较小的初值 ω_0 , 方差用较大的初值 σ_0 , 均值用 $\mathbf{I}_{(i,j),t}$ 取代。之后对 ω 进行归一化, 以保证所有 K 个高斯项权重之和为 1。然后对 K 个高斯项按照 $\omega_{(i,j),t,n} / \sigma_{(i,j),t,n}$ 降序排列, 得到

$$B = \arg \min_n \left[\sum_{n=1}^K \omega_{(i,j),t,n} > T_{\text{GMM}} \right], \quad (8)$$

式中 T_{GMM} 为阈值参数, 表示背景高斯项所占的比重。选择前 B 个满足(8)式的项作为背景高斯分布, 在目标像素判别阶段, 如果 $\mathbf{I}_{(i,j),t}$ 与 B 个背景高斯项都不能满足(3)式, 则该点为前景目标, 否则为背景像素。

2.2 GMM 方法的自适应性改进

传统 GMM 方法存在的缺点如下。

1) 高斯项数目 K 选取为 3~5 之间的固定值。对于长时间占据画面较大面积且变化不复杂的背景像素, 采用单模模型便足以模拟, 而采用过多的冗余高斯项来描述, 加大了运算量; 而对于变化较为频繁的运动目标所在区域, 小 K 值会导致建模不充分。

2) 参数学习率 α 和 β 选用固定值。较小的 α 和 β 会导致背景更新较慢, 出现拖影、虚假目标现象; 而较大 α 和 β 又会造成运动较慢的目标被更新融进背景中, 导致模型不稳定。

3) 只在时间维度上考虑单个像素的变化过程, 未对空间信息加以利用, 未考虑到邻域像素之间的空间相关性。这会导致检测结果的稳健性较差, 容易受到噪声及动态背景等影响, 前景目标不完整。

针对上述问题, 提出如下改进及优化方法。首先, 使 K 值能够随实际的场景复杂程度增大或减小。原算法每次只对首个匹配的高斯项进行更新, 这会导致模型的收敛速度较慢, 将其修正为

$$(\forall n = 1, \dots, K) p_n = \begin{cases} \omega_n \eta(\mathbf{I}_t; \boldsymbol{\mu}_n, \sigma_n), & |\mathbf{I}_t - \boldsymbol{\mu}_n| < \max(D\sigma_n, S) \\ 0, & \text{otherwise} \end{cases}, \quad (9)$$

$$q_n = \begin{cases} p_n / \sum_{n=1}^K p_n, & \sum_{n=1}^K p_n > 0 \\ 0, & \text{otherwise} \end{cases}, \quad (10)$$

式中 p_n 为每个高斯项的匹配度参数, q_n 为归一化匹配度参数。而后对所有匹配的高斯项都进行更新:

$$\omega_{(i,j),t,n} = (1 - \alpha)\omega_{(i,j),t-1,n} + \alpha q_n - \alpha c_T, \quad (11)$$

式中 c_T 为防止算法退化的修正参数。

增加敏感度参数 S 以防止当像素值长时间趋于稳定时(此时 σ 较小)产生非必需的重复的新高斯项。该参数也起到了基本阈值背景减除法与 GMM 融合的效果^[8]。较大的学习率 α 可以增强模型对光照突变等的快速适应性, 但也易产生冗余的背景高斯项 K 值, 导致前景目标被吸收入背景, 造成模型不稳定。引入参数 S 可以减弱高斯项对学习率 α 的依赖, 使得模型可以取较大的 α , 从而在加快模型对环境变化的适应性的同时, 保持稳定性。

其次, 当场景中有运动目标经过时, 对应的 K 值会增大; 而当运动目标经过后, 经过一段时间背景中的像素值趋于稳定, 此时对应的 K 值应该减小。将权重更新公式增加一项很小的值 $-\alpha c_T$, 如(11)式所示。每次 ω 更新之后对其值进行判断, 如果其值为零或负, 说明经过多帧的更新之后, 该像素值已趋于稳定, 该 ω 对应的高斯项已经不再需要, 此时将该项删除并将 K 值减 1。

对于学习率 α , 在初始背景建立阶段, 其值应取得较大, 以加快模型收敛速度; 一旦模型建立完成, 应使其保持较小值, 以保证模型的稳定。因此, 令 α 动态取值为

$$\alpha = 1 / \min(n, L_{\text{history}}), \quad (12)$$

式中 L_{history} 为设定的用于初始阶段保证模型快速收敛的帧数。另一方面, 为了利用像素的邻域空间冗余相关性的特点, 对于满足(9)式的匹配某一背景的像素, 对以该点为中心的邻域内的像素也进行一定程度的更新。选择邻域的半宽度为 1 pixel, 即对 3 pixel × 3 pixel 矩形区域内的像素进行更新。

而该区域内不同位置像素对应的学习率由学习率矩阵 $\boldsymbol{\alpha}'$ 确定, 即

$$\boldsymbol{\alpha}' = \frac{\alpha}{4} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \quad (13)$$

由(13)式确定的学习率可以确保匹配的像素所对应的 GMM 模型得到充分、快速的更新, 同时其邻域内的像素也能得到一定程度的更新, 从而充分利用空间相关性信息, 加快背景模型的收敛速度。

对于 β , 引入匹配度计数变量 $c_{(i,j),k}$ (k 表示某一高斯项), 初始值为 1, 当 $\mathbf{I}_{(i,j),t}$ 与某一高斯项匹配时, $c_{(i,j),k} = c_{(i,j),k} + q_{(i,j),k}$, 否则 $c_{(i,j),k}$ 值不变。将 (5)~(6) 式对应的 β 计算式修正为

$$\beta_{(i,j),n} = q_{(i,j),n} \left[\frac{1-\alpha}{c_{(i,j),n}} + \alpha \right]. \quad (14)$$

2.3 改进的 LBP 纹理特征背景建模

LBP 纹理分析算子是一种灰度级不变的纹理特征提取方法, 其计算简单, 对均匀光照变化不敏感, 且具有旋转不变性, 是纹理分析的有力工具。但是, 由于原始 LBP 算子采取的是计算像素间差分的形式, 当中心像素值受噪声影响之后, 对应的直方图不稳定, 导致最终的计算结果对噪声敏感, 误识别像素增多。

采用分块取均值的方式计算参考差值点, 充分利用邻域特征信息, 避免直接采用中心单像素计算抗噪声性能差的缺点。同时, 传统 LBP 算子需要取每一个像素的邻域直方图, 选定的邻域较小, 会造成求得的邻域直方图过于稀疏而失去统计意义, 且对每个像素求取直方图会导致计算量很大。采用一种新的纹理描述及处理方式, 可以避免求取直方图这一步骤。

利用中心为像素 $\mathbf{I}_{(i,j),t}$ 、半宽度为 R 的分块区域内的像素, 按照逆时针顺序, 计算得到一个二进制纹理描述字串, 记为修正的 LBP(MLBP) 算子, 其表达式为

$$M_{N,R}[\mathbf{I}_{(i,j)}] = \bigotimes_{k=0}^{N-1} b_{(i,j),k} = \bigotimes_{k=0}^{N-1} S_{th}[\mathbf{I}_{(i,j),k}, \bar{\mathbf{I}}_{(i,j)}], \quad (15)$$

$$S_{th}[\mathbf{I}_{(i,j),k}, \bar{\mathbf{I}}_{(i,j)}] = \begin{cases} 0, & \|\mathbf{I}_{(i,j),k} - \bar{\mathbf{I}}_{(i,j)}\| < T_h \\ 1, & \text{otherwise} \end{cases}, \quad (16)$$

式中 $M_{N,R}$ 为邻域内的 MLBP 特征描述子, S_{th} 为邻域像素与均值像素之间的阈值判定结果, $\bar{\mathbf{I}}_{(i,j)}$ 为以 $\mathbf{I}_{(i,j)}$ 为中心、半宽度为 R 的邻域内 N 个像素点的均值, $\mathbf{I}_{(i,j),k}$ 为邻域内除中心点外的其他像素点, T_h 为抗噪阈值, 根据 Lee 等^[9] 的研究结论选为 8, \bigotimes 表示二进制位连接, $b_{(i,j),k}$ 为 $M_{N,R}[\mathbf{I}_{(i,j)}]$ 中的一个二进制位。

每个像素位置用 K 个二进制字串描述。所有 K 个模型对应的权重 ω_k 之和为 1。当采集到新一帧图像后, 通过汉明距离计算, 将其得到的 M_{new} 与 K 个模型依次比较, 得到

$$D_{ist}(M_{new}, M_m) = \sum_{k=1}^{N-1} [b_{(i,j),k}^{new} \oplus b_{(i,j),k}^m], \quad (17)$$

式中 M_m 为背景帧的 MLBP 算子, $b_{(i,j),k}^{new}$ 为当前帧 MLBP 算子中的一个二进制位, $b_{(i,j),k}^m$ 为背景帧 MLBP 算子中的一个二进制位, m 取值为 $1 \sim K$, $D_{ist}(\cdot)$ 代表汉明距离计算, \oplus 表示异或运算。

模型更新借鉴 GMM 及 LBP 的思想, 并进行了改进。若 M_{new} 与某一项 M_m 的距离小于一定的阈值 T_{dist} , 则 M_{new} 与该背景模型匹配。在这些匹配模型中, 距离值最小的为最佳匹配模型, 对其权重进行更新:

$$\omega_{(i,j),t,m}^{MLBP} = (1 - \alpha_{LBP}) \omega_{(i,j),t-1,m}^{MLBP} + M_{(i,j),t-1,m} \alpha_{LBP}, \quad (18)$$

式中 $\omega_{(i,j),t,m}^{MLBP}$ 为当前帧权重系数, $M_{(i,j),t-1,m}$ 为前一帧权重系数, α_{LBP} 为学习速率, 最佳匹配项的 $M_{(i,j),t-1,m}$ 为 1, 其他项的为 0。对背景模型的二进制字串中的每一位进行更新:

$$p'_{(i,j),k} = \left(1 - \frac{1}{t_F}\right) p_{(i,j),k} + \frac{1}{t_F} b_{(i,j),k}, \quad (19)$$

$$b'_{(i,j),k} = \begin{cases} 0, & p'_{(i,j),k} \leq 0.5 \\ 1, & \text{otherwise} \end{cases}, \quad (20)$$

式中 $p'_{(i,j),k}$ 为当前帧二进制字串判定参数, $b'_{(i,j),k}$ 为当前帧二进制字串位更新值, t_F 为帧序号, $p_{(i,j),k}$ 为 $b_{(i,j),k} = 1$ 的概率, 初始 $b_{(i,j),k} = 0$ 。 t_F 的取值在小于阈值 T 时随帧数增大而增大, 以便模型更快地收敛; 当帧数大于 T 时, 恒定取为 T , 以保持模型的稳定。实验中 T 取为 200。

若 M_{new} 与 K 个模型中的任意一个都不匹配, 则用 M_{new} 替代 K 个模型中权重最小的一项, 且令该项的权重为一小的初始值, 实验中取为 0.01, 随后对 K 项的权重进行归一化处理。之后将 K 个模型按照权重进行降序排列, 并选取其中的 B_{MLBP} 项作为背景模型:

$$B_{MLBP} = \arg \min_m \left[\sum_{m=1}^K \omega_{(i,j),t,m} > T_{MLBP} \right], \quad (21)$$

式中 T_{MLBP} 为背景模型项所占比重的阈值。

2.4 颜色与纹理建模多阈值处理及结合策略

对于 GMM 和 MLBP 算法, 前景检测结果的质量高度依赖于 (3) 式和 (17) 式计算判定时选定的阈值, 较大的阈值会导致检测出的目标出现分裂、不完整现象; 而较小的阈值会产生由很多噪声引起的伪前景目标像素。因此, 在实际应用时, 很难对动态变化的每一帧选取最恰当的单阈值。对以上 GMM 和 MLBP 建模均采用滞后多阈值处理策略, 充分结合大小阈值检测的优点。

通过设置多个不同大小的阈值, 对应可得到多

个前景二值图像。对于基于颜色特征的 GMM 算法,置信度参数取 5 个不同大小的阈值 $D=[3.5,3,2.5,2,1.5]$, D 最大时得到的称为主颜色前景图(用 PC 表示),其余为辅颜色前景图(用 SC 表示)。处理效果如图 2 所示。

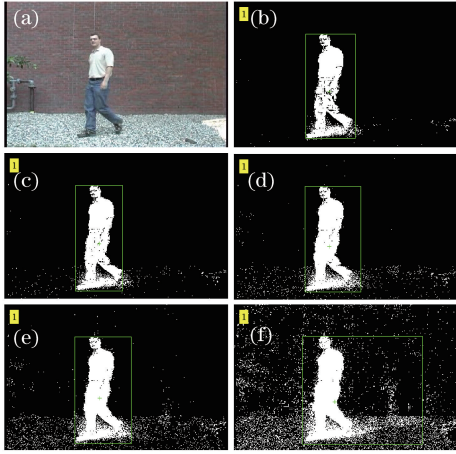


图 2 多阈值运动目标的颜色特征提取。(a)原始图片;
(b) PC; (c) SC1; (d) SC2; (e) SC3; (f) SC4
Fig. 2 Color feature extraction of multi-thresholds moving objects. (a) Original image; (b) PC; (c) SC1; (d) SC2; (e) SC3; (f) SC4

图 2 中用矩形框框出了检测出的运动目标区域,同时在图像左上角标记出了检测到的目标数目,便于后续应用。从图 2 中可以看出,PC 图像检测出了运动目标的主体部分,但较大的阈值导致检测出的目标区域出现了孔洞,部分细节缺失;而 SC 图像则检测出了较多的细节,但同时也更易受噪声影响,含有较多的虚警像素,导致矩形框选出的目标区域较实际的偏大很多。多阈值处理算法的目的就是要充分利用 PC 和 SC 图像的互补性,得到抑制噪声虚警像素的完整前景图像。

对于纹理背景,由于纹理特征描述的有用信息较颜色特征少且易受噪声影响,同时为了减少运算量,对 LBP 建模方法的 T_{dist} 取双阈值进行处理。实验中主阈值 T_{dist} 取为 4,辅阈值取为 2,对应得到主纹理前景图(用 PT 表示)和辅纹理前景图(用 ST 表示)。

PC 与 PT 图像中含有最重要的信息,因此首先将 PC 和 PT 图像结合得到主颜色和纹理结合图像(PCT)。同时 PT 图像几乎不受阴影影响,两者的结合可在一定程度上消除阴影,结果如图 3 右半部分虚线框内所示。

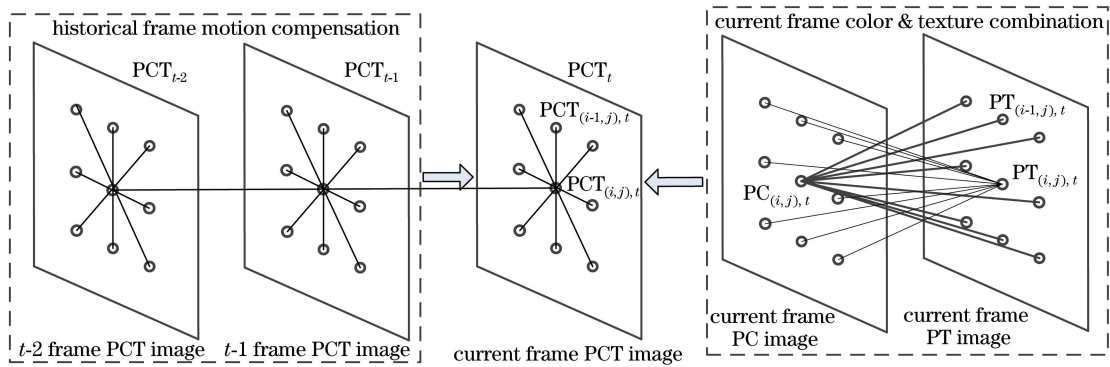


图 3 颜色与纹理特征的多阈值结合与运动补偿示意图

Fig. 3 Schematic of color and texture based multi-thresholds combined with motion compensation

基于邻域信息一致性原则,对于 PC 图像中的任一被判定为前景的像素,如果 PT 中以 (i,j) 处像素为中心、半径为 R (取为 1,即 $3 \text{ pixel} \times 3 \text{ pixel}$)的邻域内存在前景像素(即二值图中存在为 1 的像素),则将 (i,j) 位置的像素点标记为前景 1,否则标记为 0;而后对 PT 图像进行同样的操作,由此得到结合图像 PCT。将 PCT 图像得到的前景像素区域用矩形框框出。将每个 SC 图像所对应的阈值降序排列,利用与之前类似的结合策略,将其依次与 PCT 图像进行结合。为减小运算量,只对 PCT 图像中矩形框标记出的区域内的像素进行操作,得到的结果用主/辅颜色和纹理结合图像(PSCT)表示,

如图 4 所示。

为解决朝向或远离摄像头运动引起的前景目标内部空洞的问题,在上述空间补偿方法的基础上,结合运动线索对前景目标进行补偿,对当前帧 t 的前 m (取为 3)帧历史运动信息加以利用。若一个像素半宽度为 R (取为 1)的邻域的像素在历史帧中属于前景的数目较多,则认为该像素也有较大概率属于前景。运动补偿操作也只在前述步骤标记出的矩形框内进行。如图 3 左半部分所示,定义 $I_{(i,j)}$ 像素属于前景的概率为

$$P_{(i,j)} = \frac{N_{\text{fg}}}{26} = \left(\sum_{k=1}^{26} b_k \right) / 26, \quad (22)$$

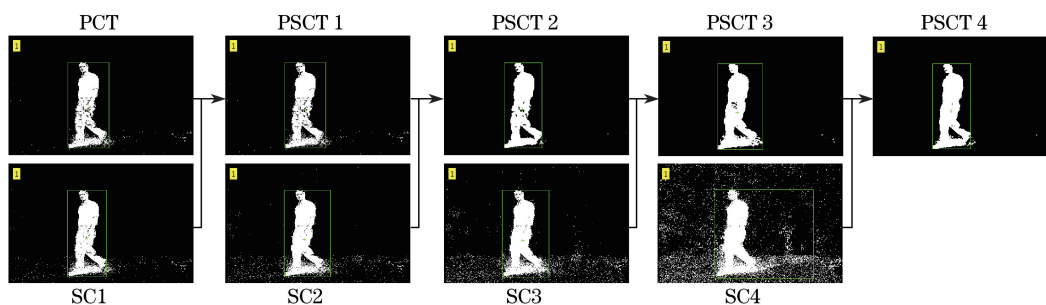


图 4 颜色与纹理特征结合前景目标的补偿

Fig. 4 Compensation of color and texture combined foreground objects

式中 b_k 为前后三帧 R 邻域内每个位置处的像素值 (0 或 1), N_{i_g} 为当前帧及当前帧的前两个相邻帧 R 邻域内属于前景像素的像素个数总和, 26 为除当前 (i, j) 位置的所有的像素的总数。如果计算出的 $P_{(i, j)}$ 大于 0.6, 则将该点标记为 1, 并补充到前景二值图像中。

2.5 结合 Canny 思想的 Kirsch 边缘检测前景优化

基于块区域的纹理特征不能较好地保留边缘信息, 检测出的前景目标通常会产生一些棱角、锯齿形状的边缘, 且可能使邻近的目标连接在一起, 多数情况下不能将处于平坦纹理区域的前景和背景像素分离开。边缘检测能得到运动目标的轮廓信息, 从而能确定运动目标所在的像素区域。相比于前景区域, 轮廓信息不易受光照、目标和相机抖动等的影响, 更具有稳定性, 可用作运动目标检测的辅助手

段。提出了一种结合 Canny 思想的 Kirsch 边缘检测算法, 在消除鬼影和平滑前景目标边缘两个方面对提取的前景图像进行优化。

Canny 算子是一种基于最优化思想的检测算子, 可以得到完整性和连续性较好的边缘。但 Canny 算子对噪声敏感, 提取边缘前需先进行高斯平滑去噪, 这不仅会增加运算量, 还会造成边缘失真。Kirsch 算子是一种带有方向性的边缘检测算子^[10], 可以尽可能多地检测出边缘, 且中心像素点不直接参与计算, 可有效克服传统的基于梯度微分算子的检测算法对噪声敏感的优点。但其得到的图像边缘不是单像素的, 边缘定位不准, 边缘完整性和连续性较差。

借鉴 Canny 算法的最优化思想, 将 Kirsch 算法 8 个方向上的最大值进行非最大值抑制^[11], 得到优化的边缘图像, 过程如图 5 所示。

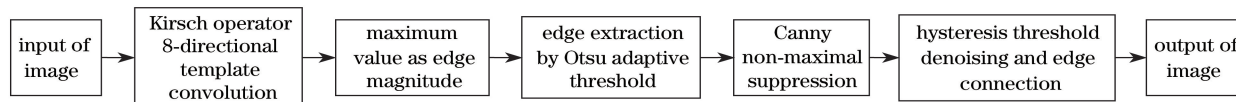


图 5 结合 Canny 思想的 Kirsch 边缘检测算法流程

Fig. 5 Flow chart of Canny edge detection method combined with Canny thoughts

分别对当前输入图像及 PCST 图像进行图 6 所示的边缘检测, 而后用矩形框标记出各自图像中的目标区域, 计算两幅图像中标记区域中重叠部分的面积, 如果该面积小于设定的面积阈值参数 T_{area} , 则认为 PCST 提取的前景图中该区域内原本被标记为前景目标的区域为鬼影, 将其剔除。然后, 将当前帧提取的边缘图像与鬼影抑制后的 PCST 图像进行逻辑或操作, 对前景目标轮廓边缘进行平滑和优化处理, 弥补边缘模糊和缺失。随后进行形态学运算, 去除杂散噪声像素。

鬼影目标形成的主要原因: 场景中原来处于静止状态的目标运动之后, 在其原本占据的像素区域仍然检测出并标记为虚假的运动目标; 或者由于背景模型更新太慢, 运动目标离开之后的相邻若干帧

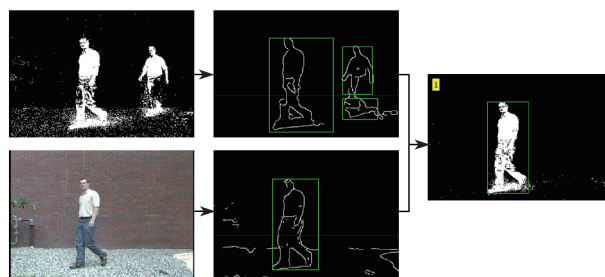


图 6 边缘检测鬼影抑制效果

Fig. 6 Ghost suppression effect in edge detection

中仍然误检出目标区域。鬼影目标的存在对模型建立初始化阶段不含运动目标的背景建模算法的影响尤为严重, 会导致算法收敛速度慢、对场景变化的适应滞后等, 严重影响背景的建模和更新及后续跟踪、分类、识别等过程。

3 实验与结果分析

为了对算法的效果进行验证,分别选取 CDnet2014 数据库^[12]及采用实验室搭建的实验平台获取的视频进行处理,并与效果较好的几种典型的运动目标检测算法进行定性和定量的实验对比分析。这几种算法分别是 Dirichlet 处理高斯混合模型(DPGMM)^[13]、LBP^[14]、马尔可夫随机场算法(MRF)^[15]、ViBe^[16]、IMBS^[17]。

提出的算法参数设置如下: K 最大值设为 3, $\sigma_0=10$, $\alpha_0=0.01$, $T_{\text{GMM}}=0.9$, $T_{\text{MLBP}}=0.75$, $S=15$, $c_T=0.05$, $\alpha_{\text{LBP}}=0.5$ 。对比算法的参数值均按照对应文献中说明的值进行设置。

3.1 CDnet2014 数据库实验

实验选取 CDNet2014 中的 6 类室内外的典型场景,检测对比结果如图 7 所示。理想的目标检测算法应该将每个运动目标都单独检测出来,且目标自身没有断裂,与其他目标之间没有粘连。

图 7 中每一行为选取一个视频数据集中的一帧,图 7(a)~(f)所示依次为车站、羽毛球场、人行道、沙发、PETS2006 及雨雪数据集。

图 7(a)所示数据集的主要挑战在于行人的间歇性运动和长时间停留,以及光线的变化和投射阴影;

图 7(b)、(c)所示数据集的主要挑战在于摄像机的抖动造成背景的不稳定;图 7(d)、(e)所示数据集同为室内场景,主要挑战在于灯光投射造成的阴影,以及目标物的间歇性运动和暂留。同时,图 7(d)所示数据集内的运动目标也存在较明显的大面积阴影;图 7(f)所示数据集则是恶劣天气条件下的场景,空气中飘着较密集的雪花,且摄像机镜头上不断有雨水滑下,导致背景画面中存在很多干扰因素。

从图 7 可以看出,在各类场景下,相较于其他几种算法,所提算法检测得到运动目标的前景图像都更准确和完整,尤其对于摄像机抖动造成的背景扰动具有很好的抑制效果,得到的前景图像误检像素点较少,且前景边界更符合实际情况。传统的 LBP 方法通常会造成检测出的前景目标像素区域比实际的大,轮廓边缘模糊,或在目标离摄像头较远视野内存在多个小目标时目标内部分离的各部分或不同目标之间粘连为一个区域,造成区域误检。从图 7(a)、(d)数据集也可以看出,所提算法对面积较大的阴影区域的抑制效果还有待增强;同时,对于图 7(d)所示暂留时间较长的目标是否被判定为前景或背景的策略还需进一步优化。对于图 7(f)所示恶劣天气场景下,所提算法虽然比其他算法的效果好,但检测结果中还存在较多的误检像素。

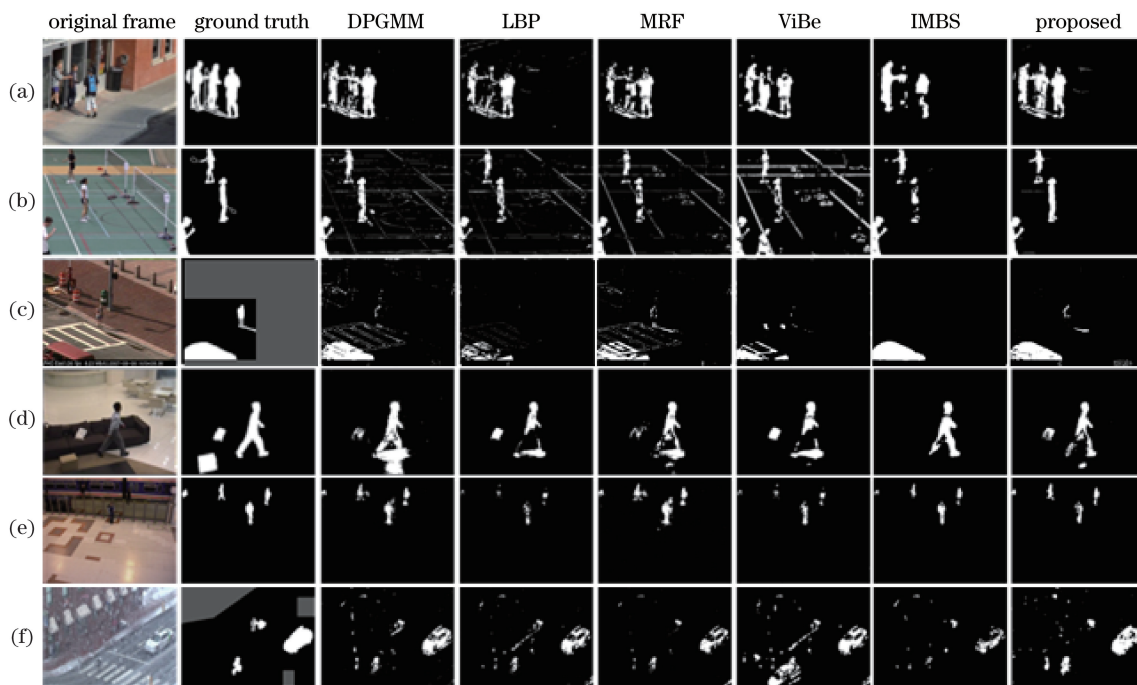


图 7 CDNet2014 数据库不同场景下几种算法的运动目标检测效果

Fig. 7 Moving objects detection results obtained by several algorithms under different scenes from CDNet2014 dataset

为进行定量对比分析,采用 CDNet2014 网站提出的一些评价指标。令 f_{TP} 、 f_{TN} 、 f_{FP} 、 f_{FN} 分别为真

正、真负、误正、误负的像素数目,采用文献^[10]中定义的以下 6 类评价指标:误正率($f_{\text{FPR}}=f_{\text{FP}}/(f_{\text{FP}}+$

f_{TN}); 误负率(f_{FNR})= $f_{FN}/(f_{TP} + f_{FN})$; 误检百分比(f_{PWC})= $100(f_{FN} + f_{FP})/(f_{TP} + f_{FN} + f_{FP} + f_{TN})$; 检出率(R_e)= $f_{TP}/(f_{TP} + f_{FN})$, 衡量结果的完整性; 正确率(P_r)= $f_{TP}/(f_{TP} + f_{FP})$, 衡量检测结果的准确性; F_m 测度(F_m)= $(2 \times R_e \times P_r)/(R_e + P_r)$, 衡量算法的综合性能。

得到 6 种算法的正确率-检出率(P-R)曲线, 如

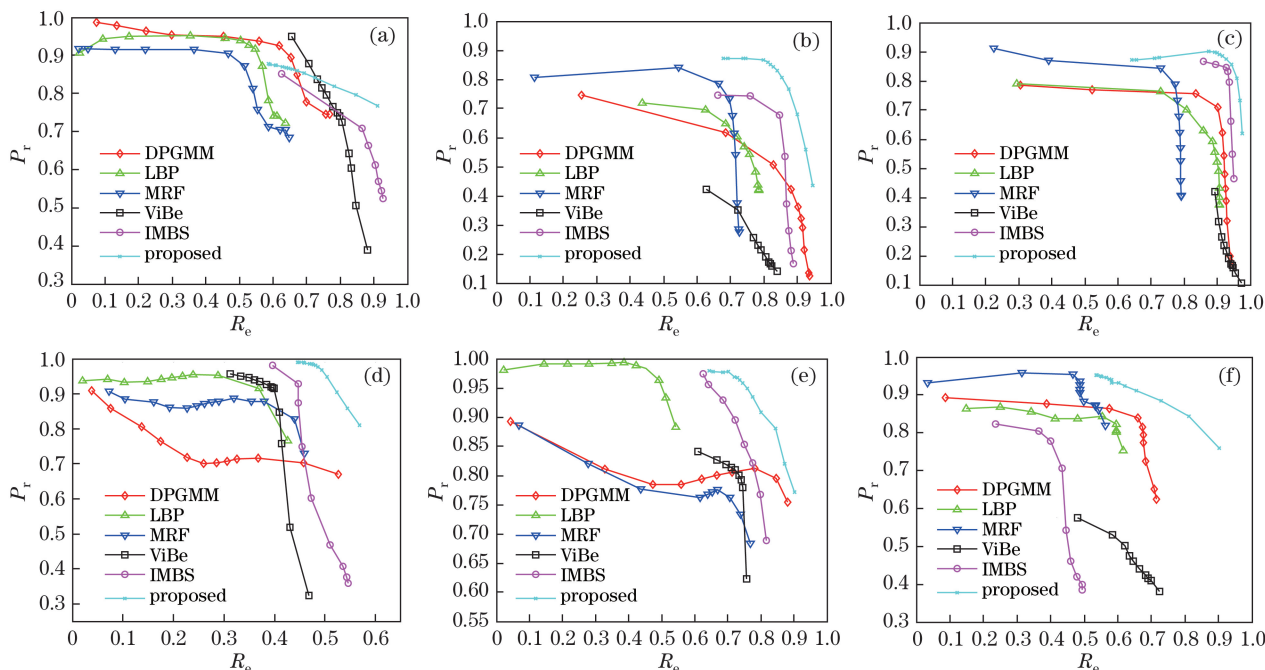


图 8 不同算法在 CDNet2014 数据库 6 个场景下得到的 P-R 曲线。(a) 车站; (b) 羽毛球场; (c) 人行道; (d) 沙发; (e) PETS2006; (f) 雨雪

Fig. 8 P-R curves obtained by different algorithms under six scenes from CDNet2014 dataset.

(a) Bus station; (b) badminton; (c) sidewalk; (d) sofa; (e) PETS2006; (f) wet snow

几种算法在上述 CDNet2014 数据集 6 个场景下得到的 6 类参数的平均值见表 1。

表 1 不同算法在 CDNet2014 数据库 6 个场景下的整体平均性能参数对比

Table 1 Overall average performance comparison for different algorithms under six scenes from CDNet2014 dataset

Algorithm	f_{FPR}	f_{FNR}	f_{PWC}	R_e	P_r	F_m
DPGMM	0.007	0.336	1.944	0.664	0.741	0.686
LBP	0.006	0.447	2.015	0.553	0.821	0.625
MRF	0.003	0.475	2.044	0.525	0.842	0.623
ViBe	0.020	0.298	3.083	0.702	0.627	0.619
IMBS	0.003	0.341	1.431	0.659	0.878	0.730
Proposed	0.004	0.307	1.784	0.694	0.868	0.754

由表 1 可以看出, 由所提算法得到的指标值大都可达最优, 或与最优值接近。同时, 对于实际评测中经常选作综合评价指标的 F_m 值, 所提算法在

图 8 所示。

从图 8 中曲线可以看出, 所提算法在车站数据集下得到的 P-R 曲线效果稍差, 但在其他几类数据集下均更靠近于坐标轴的右上一侧, 即其得到的 P-R 值较大, 表明其算法效果优于其他几种对比算法的。所提算法可以较好地处理恶劣天气、摄像机抖动等实际场景中常见的具有挑战性的复杂背景。

所选的几种对比算法中得到的值最大, 表明所提算法的综合性能最优。

对比结果表明, 在多数测试场景下, 所提算法在检出率、相似度和 F_m 值等指标上都优于其他算法的, 而 P_r 精度指标并不都是最优, 因为补偿过程引入了一些虚警像素。

为了更详细地对比以上 6 种算法的综合性能, 图 9 给出了几种算法在 6 个测试数据集上的 F_m 曲线。

从图 9 中可以看出, 所提算法在车站及人行道数据集上的 F_m 值比 IMBS 算法的略低, 但在其余数据集上均取得了最好的效果。且从曲线的最后一组平均数值来看, 所提算法在 6 种测试数据集上的 F_m 值最优, 这也表明该算法的综合性能优于其他几种对比算法的。

图 10 对所提算法在 6 个数据集上的检测效果进行了综合显示, 以便更直观地看出所提算法的优缺点。

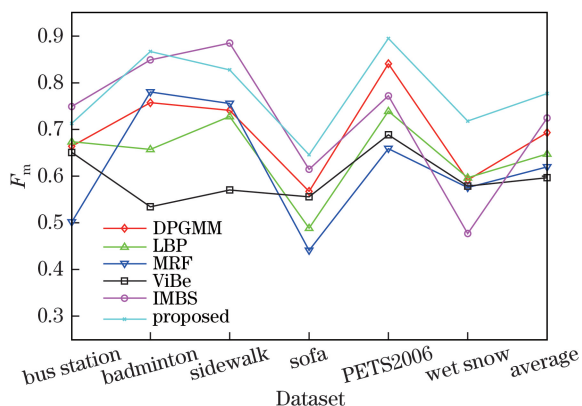


图 9 不同算法在各测试数据集上的 F_m 值

Fig. 9 F_m value in each test dataset for different algorithms

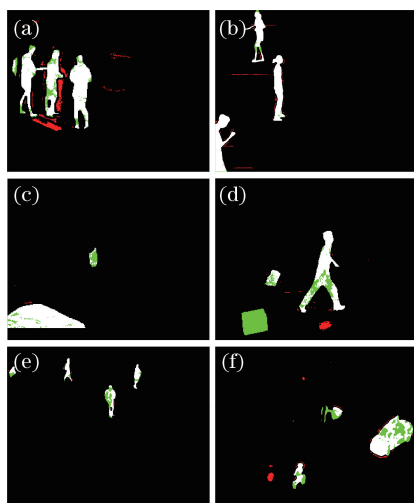


图 10 不同场景下所提算法检测得到的前景及误检像素。(a) 车站; (b) 羽毛球场; (c) 人行道; (d) 沙发; (e) PETS2006; (f) 雨雪

Fig. 10 Foreground and mistakenly detected pixels for the proposed algorithm under different scenes.

(a) Bus station; (b) badminton; (c) sidewalk; (d) sofa; (e) PETS2006; (f) wet snow

在图 10 中, f_{TP} 像素用白色显示, f_{TN} 像素用黑色显示, f_{FP} 像素用红色显示, f_{FN} 像素用绿色显示,

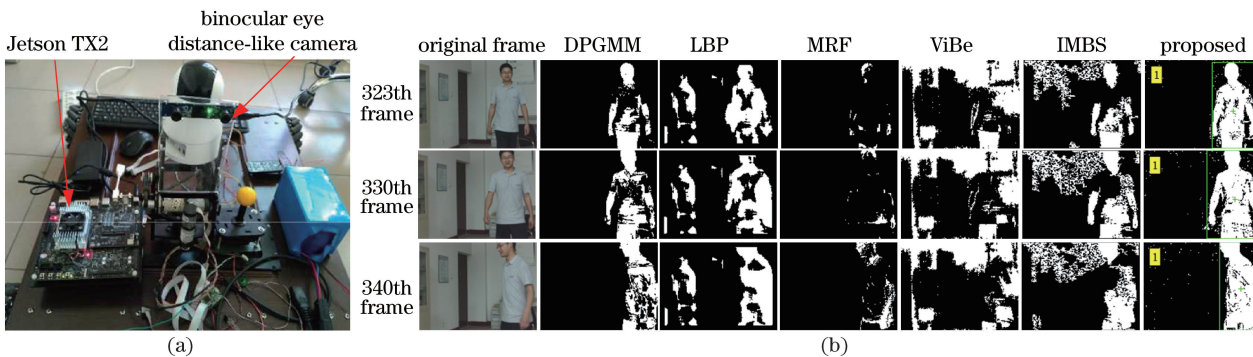


图 11 实验。(a) 实验装置; (b) 各算法目标检测效果

Fig. 11 Experiment. (a) Experimental devices; (b) results of object detection obtained by each algorithm

可以看到,所提算法在各类场景下均能检测出较为完整的前景目标。但对于面积较大的阴影区域,以及长时间滞留的物体会产生误检像素。

3.2 实际捕获视频检测

设计的实验装置如图 11(a)所示,该装置由双目广角摄像头以及底部的机械支撑结构、双自由度电机驱动机构等部分组成,用于后续移动平台下的目标检测、识别及跟踪等研究。这里暂只取其中的单目摄像头获取的视频图像进行处理,验证所提算法的实际应用效果。

测试硬件环境基于 NVIDIA Jetson TX2 平台。软件在 Ubuntu Linux 环境下基于 OpenCV 框架采用 C++ 编程实现算法。实验结果如图 11(b)所示。

采用实验室设计的 1280 pixel×480 pixel 双目同步输出摄像头采集视频,之后裁剪出单幅 640 pixel×480 pixel 的视频帧进行算法处理,帧率为 30 frame/s,检测效果如图 11(b)所示。实验主要考察:1)目标向摄像头方向运动的场景下各种算法的检测效果;2)在视频的第 184 帧关闭了室内的灯光,各算法对光照突变的稳健性及快速适应性;3)视频中运动目标着装的颜色与背景墙壁的颜色十分接近且纹理区域平坦时,基于单特征算法的性能。

为更好地对比各算法的效果,所有算法提取的前景目标均未经过形态学后处理。从图 11(b)可以看出,对于初始帧中含有运动目标的场景,LBP 算法下一直存在鬼影;IMBS 算法大约需要 198 帧才能收敛得到稳定背景,对光照变化的适应性较差,且背景中误正干扰像素较多,稳健性较差;在第 184 帧关闭了灯光,而 ViBe 算法直到第 323 帧仍未适应场景光照变化,因此提取到的前景中的误正像素同样很多;对于初始帧中含有运动目标的场景,ViBe 算法下一直存在鬼影,很难融入背景。

当目标朝向摄像头运动时,由于目标内部的像素相对于摄像头的变化很小,传统的基于单特征、单像素建模的方法将属于前景运动目标内部颜色相近的像素视为静止的,该部分像素随着模型更新逐渐被吸收融入背景之中,从而导致最终检测出的前景目标区域出现内部空洞,或者分裂为多个目标,这会导致虚警,影响高层应用如行为分析、目标识别等的准确性。虽然形态学滤波在有些场景下可以一定程度上弥补这一缺陷,但其会将阴影及噪声像素区域

放大,导致检测出的前景像素区域的准确性变差。

所提算法通过邻域多阈值补偿策略对运动的历史信息加以利用,可以得到更为完整和准确的目标区域,且在开关灯光照变化的情况下也能适应场景变化,保证了其在现实场景中应用的稳健性,为进一步进行目标跟踪、识别等提供了较好的基础。

为加快运算速度,将获取的视频帧缩放为 320 pixel×240 pixel。几种算法对视频的处理速度见表 2。

表 2 不同算法处理速度的对比

Table 2 Processing speed comparison among different algorithms

Algorithm	DPGMM	LBP	MRF	ViBe	IMBS	Proposed
Speed / (frame·s ⁻¹)	29.8	2.69	17.3	23.6	21.7	28.4

从表 2 可以看出,传统的 LBP 算法运算耗时很长。而所提算法通过对邻域二进制纹理描述字串进行汉明距离比较,避免了求取每个像素直方图,增强了算法的实时性。所提算法处理速度能达到 28.4 frame/s,对算法进一步优化之后基本能达到实际应用的实时性要求。

4 结 论

提出了一种将颜色、纹理及轮廓多特征结合的运动目标检测算法。分别通过改进的自适应性高斯混合建模方法和 LBP 纹理描述算子对颜色和纹理特征进行了建模;通过多阈值互补策略将颜色特征和纹理特征建模进行了有机结合,利用多个辅助前景图像对主前景图像的形状及细节进行了优化和修正;通过历史帧信息进行了运动补偿,消除了前景目标内部的空洞和分裂像素块;并结合轮廓特征进行了鬼影消除及轮廓平滑优化处理。实验结果表明,所提算法在多种复杂场景下对运动目标的检测效果优于其他算法的,尤其对于目标朝向摄像头运动的情形。但所提算法对目标间歇性运动及大面积阴影干扰场景的处理效果还有待进一步加强。下一步研究还需要继续优化算法,将其改写为由图形处理器运行的版本,进一步增强算法的实时性。

参 考 文 献

- [1] Cheng F C, Chen B H, Huang S. A hybrid background subtraction method with background and foreground candidates detection[J]. ACM Transactions on Intelligent Systems and Technology, 2015, 1(7): 1-14.
- [2] Chen S Y, Xu T F, Li D Q, *et al.* Moving object detection using scanning camera on a high-precision intelligent holder[J]. Sensors, 2016, 16(10): 1758.
- [3] Chen H Y, Qie L Z, Yang D D, *et al.* Visual background extraction algorithm based on superpixel information feedback[J]. Acta Optica Sinica, 2017, 37(7): 0715001.
陈海永, 鄢丽忠, 杨德东, 等. 基于超像素信息反馈的视觉背景提取算法[J]. 光学学报, 2017, 37(7): 0715001.
- [4] Kumar P, Singhal A, Mehta S, *et al.* Real-time moving object detection algorithm on high-resolution videos using GPUs[J]. Journal of Real-Time Image Processing, 2016, 11(1): 93-109.
- [5] Chua T W, Leman K, Wang Y. Fuzzy rule-based system for dynamic texture and color based background subtraction[C]. IEEE International Conference on Fuzzy Systems, 2012, 7: 10-15.
- [6] Song T, Li O, Cui H L. A moving object detection method based on scene perception[J]. Acta Electronica Sinica, 2016, 44(11): 2625-2632.
宋涛, 李鸥, 崔弘亮. 基于场景感知的运动目标检测方法[J]. 电子学报, 2016, 44(11): 2625-2632.
- [7] Tu L F, Zhong S D, Peng Q. Moving object detection method based on complementary multi resolution background models[J]. Journal of Central South University, 2014, 21(6): 2306-2314.
- [8] Prashant D, Pratik S, Suman K M. Fast and accurate foreground background separation for video surveillance[J]. Lecture Notes in Computer Science, 2015, 9124: 96-103.
- [9] Lee I S, Tsai W H. Data hiding in grayscale images by dynamic programming based on a human visual model[J]. Pattern Recognition, 2009, 42(7): 1604-1611.
- [10] Guo M, Hu L L, Zhao J T. Surface defect detection

- method of ceramic bowl based on Kirsch and Canny operator[J]. *Acta Optica Sinica*, 2016, 36(9): 0904001.
- 郭萌, 胡辽林, 赵江涛. 基于 Kirsch 和 Canny 算子的陶瓷碗表面缺陷检测方法[J]. *光学学报*, 2016, 36(9): 0904001.
- [11] Zou Y B, Zhou W L, Chen X Z. Research of laser vision seam detection and tracking system based on depth hierarchical feature[J]. *Chinese Journal of Lasers*, 2017, 44(4): 0402009.
- 邹焱飏, 周卫林, 陈向志. 基于深度分层特征的激光视觉焊缝检测与跟踪系统研究[J]. *中国激光*, 2017, 44(4): 0402009.
- [12] Wang Y, Yipierre J, Fatih P, *et al.* CDnet 2014: An expanded change detection benchmark dataset[C]. *Computer Vision & Pattern Recognition Workshops*, 2014: 14632506.
- [13] Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction[C]. *International Conference on Pattern Recognition*, 2004, 8(2): 28-31.
- [14] Heikkila M, Pietikainen M. A texture-based method for modeling the background and detecting moving objects[J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2006, 28(4): 657-662.
- [15] Schick A, Auml M B, Stiefelhagen R. Improving foreground segmentations with probabilistic superpixel Markov random fields[C]. *Computer Vision & Pattern Recognition Workshops*, 2012: 12865142.
- [16] Barnich O, van Droogenbroeck M. ViBe: A universal background subtraction algorithm for video sequences[J]. *IEEE Transactions on Image Processing*, 2011, 20(6): 1709-1724.
- [17] Bloisi D D, Pennisi A, Iocchi L. Background modeling in the maritime domain[J]. *Machine Vision and Applications*, 2014, 25(5): 1257-1269.