

监控视频 ROI 完整分割技术研究^①

刘 畅^② * * * * * 王鹏钧 * * * * * 赵 潇 * * 田 霖 * * 周一青 * * * * * 石晶林 * * * * *

(* 移动计算与新型终端北京市重点实验室 北京 100190)

(** 中国科学院计算技术研究所 北京 100190)

(*** 中国科学院大学 北京 100049)

(**** 96901 部队 北京 100094)

摘 要 智能视频监控技术发展迅速,对运动目标进行检测与感兴趣区域(ROI)分割是智能分析的基础。然而,在现有的运动目标检测算法基础上进行区域分割时会出现局部缺失问题,这将导致关键信息(运动目标信息)的损失并影响后续对视频的分析。因此,本文提出了一种基于空间和色彩 2 维特征联合分析的 ROI 完整分割算法,能够基于传统背景差法的检测结果,通过在空间维度进行检测结果的聚类以及误差过滤,并在色彩维度基于超像素聚合算法实现运动目标边界定位,从而实现在保障准确率和召回率的前提下,大幅提高 ROI 区域的完整分割性能。

关键词 智能视频监控;目标检测;感兴趣区域(ROI);超像素;离群检测

0 引 言

随着智慧城市和通信技术的快速发展^[1-3],视频监控的应用领域日趋广泛。如今,视频监控系统逐渐呈现出图像高清化、设备联网化和网络密集化的发展趋势,海量的视频监控数据已经无法通过人工的方式进行分析^[4],智能视频监控技术成为了重要的演进目标。在视频监控数据中,前景目标是智能计算与分析的核心,因此,目标检测技术一直是研究的热点。总体来说,可以将视频监控系统中的目标检测技术分为基于背景建模和基于对象建模的 2 种检测方法。

基于背景建模的运动目标检测算法的目的是检测出画面内的一切运动目标。该类算法通过对图像底层特征进行分析,并对视频监控背景进行建模,从而检测出运动目标信息并分割出对应区域。目前,

已有大量的背景建模算法,包括背景差法、帧差法、极值滤波^[5]、中值滤波^[6]、均值滤波^[7]、非参数模型法^[8]、混合高斯模型法 (generalized method of moments, GMM)^[9]、隐马尔科夫法^[10] 等等。其中, GMM 方法是最为经典的方法之一,视觉背景提取算法 (visual background extractor, ViBe)^[11] 和像素背景自适应分割算法 (pixel-based adaptive segmenter, PBAS)^[12] 是目前综合性能最好的运动目标检测方法^[13]。

基于对象建模的目标检测算法实现了对特定对象的检测,例如行人、车辆等。该类算法受场景的限制较小,可以在场景动态变化时通过对特定对象的学习,进行对应目标特征的检测。然而,对象建模类算法仅能提取出目标所在的规则图像区域,如矩形区域 (bounding box),但无法实现目标精确轮廓的检测。目前,基于对象建模的目标检测方法又可以分为传统学习和深度学习 2 种方法。传统学习方法包

① 北京市自然科学基金 (No. L172049) 资助项目。
② 男,1986 年生,博士生;研究方向:多媒体传输,网络资源分配,计算机视觉;联系人,E-mail: liuchang@ict.ac.cn
(收稿日期:2019-01-02)

括刚性全局检测方法^[14,15]、视觉词典方法^[16,17]、部件检测法^[18,19]等。基于深度学习的目标检测方法近年来发展迅速^[20-22],该类方法将特征的学习与分类融合在一起,通过复杂的神经网络来模拟人类大脑的判决机制。该类方法一般会从输入的图像中提取出一个区域块,并用卷积神经网络对其中的图像特征进行分析和计算,随后进行检测与区域分割。

在视频监控系统中,考虑到基于对象建模的检测算法存在只能对特定目标进行检测、处理速度慢、计算开销较大等问题^[23],因此主要通过基于背景建模的目标检测算法进行运动目标的检测与感兴趣区域(region of interest, ROI)分割。基于背景建模的运动目标检测算法一般将检测结果按照背景或前景的分类输出二值图像,图像中的信息是以像素为单位进行统计的。然而,现有的背景建模类检测算法受光线、阴影、噪声等问题干扰较大,二值图像中的误检与漏检问题严重,直接进行区域分割会导致严重的目标局部缺失问题,不利于智能分析等后续应用。因此,本文基于传统的背景差分算法,进一步研究了结合空间聚类与过滤、色彩超像素分析的运动目标区域 ROI 完整分割方法。一方面,在空间维度上,充分考虑了检测结果中像素的空间分布特点,采用 K-means 聚类^[24]的方法对相近的同类信息进行聚类,并用 T 离群检测^[25]方法对噪声信息进行过滤,从而聚合出初步的目标图像区域;另一方面,在色彩分析时,基于目标色彩特征的连续性,通过超像素分割^[26]方法对目标区域进行精确定位。最后,将空间和色彩维度的分析结果进行联合分析,从而实现完整的运动目标 ROI 分割。接下来,将对该算法进行详细介绍。

1 运动目标 ROI 完整分割算法

基于前文分析可知,一方面,在监控视频中由于背景相对固定,前景信息多为运动经过的人、车等目标,如果将监控背景进行建模,采用背景差分法对画面中的运动目标进行检测,并基于聚类算法等过滤掉检测结果中的错误信息,可以初步定位出视频场景中的运动目标 ROI 位置。另一方面,针对 ROI 的完

整分割提取问题,在空间初步判定结果的基础上,可以通过超像素分割技术,进一步通过分析目标色彩特征的连续性,从而得到完整的运动目标 ROI 区域。

因此,基于以上思路,本文设计了一种基于空间与色彩 2 维特征联合分析的运动目标 ROI 完整分割方法。算法流程如图 1 所述,可分为预处理、空间分析和色彩分析共 3 个步骤。

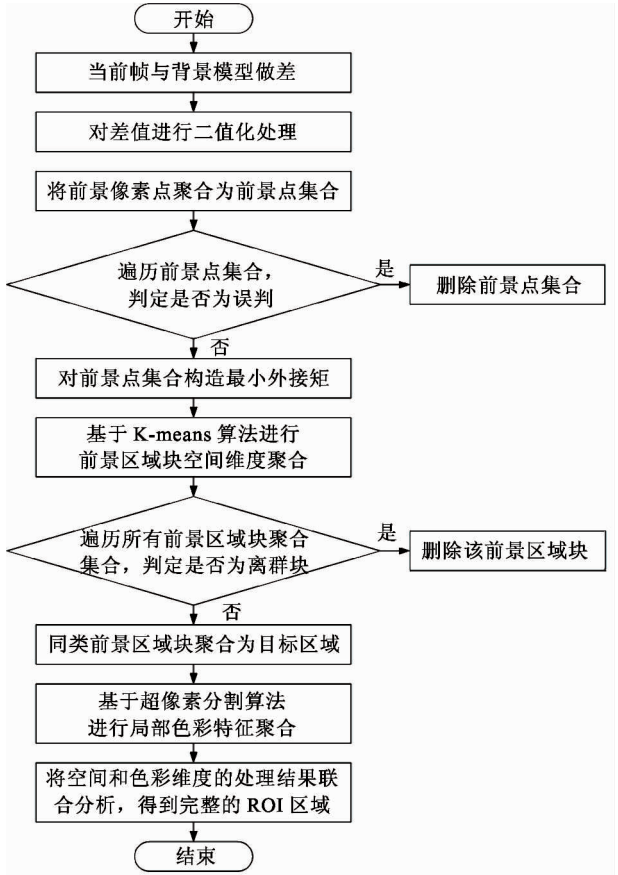


图 1 基于空间与色彩特征分析的 ROI 分割方法处理流程

第 1 步,预处理。将当前帧与背景模型进行差值运算,得到二者的差异信息(像素级),其中,与背景模型相同的定义为背景信息,其他信息定义为前景信息。随后,对分布较为集中的前景信息进行初步聚合,形成前景点集合,并删除无法聚类的前景像素点。

第 2 步,空间分析。对每个前景点集合构造最小外接矩形,形成前景区域块,并基于区域块的中心位置信息,执行 K-means 聚类操作,可将区域块按照目标进行聚合,同时,采用离群检测算法,进一步过

滤掉光照、阴影等造成的错误前景区域块。

第 3 步,色彩分析。充分考虑目标色彩的连续性,基于超像素分割算法在色彩空间将具有相似纹理、颜色、亮度等特征的相邻像素聚合,构造出目标的精确边缘。最后,将空间维度聚合结果与色彩维度超像素分割结果的重叠区域取并集运算,得到完整的 ROI。

接下来,将通过示例对算法流程进行介绍。如图 2 所示,首先在空间维度上,通过将实时输入的视视频帧与背景模型进行差值运算,得到差异的二值图像,对二值图像的像素点聚类,可以得到初步的聚合结果,即前景区域块。然而,在这些前景区域块中,有很多是由错误的检测结果导致的,其面积较小、分布较为稀疏。因此,需要进一步对这些前景区域块

进行聚类,并过滤掉错误的前景区域块,从而得到准确的运动目标区域。

同时,在色彩维度上,可以通过超像素分割的方法,打破像素级的图像描述形式,将色彩、位置、纹理具有较强相关性的像素聚合在一起,成为粒度更大的超像素。超像素在生成时由于充分考虑了色彩的相似性,因此其边界可以反映出目标内部色彩特征的连续性。因此,可以进一步基于超像素的边界,对空间维度的聚合结果进行分析,提取出与二者的并集,从而分割出完整的运动目标区域。根据以上分析可知,空间维度的检测结果聚类与误差过滤、色彩维度的超像素聚合与分析是本文所提方法的重点,接下来,将进一步对空间检测结果聚合与过滤方法、超像素聚合与分析方法进行详细分析。

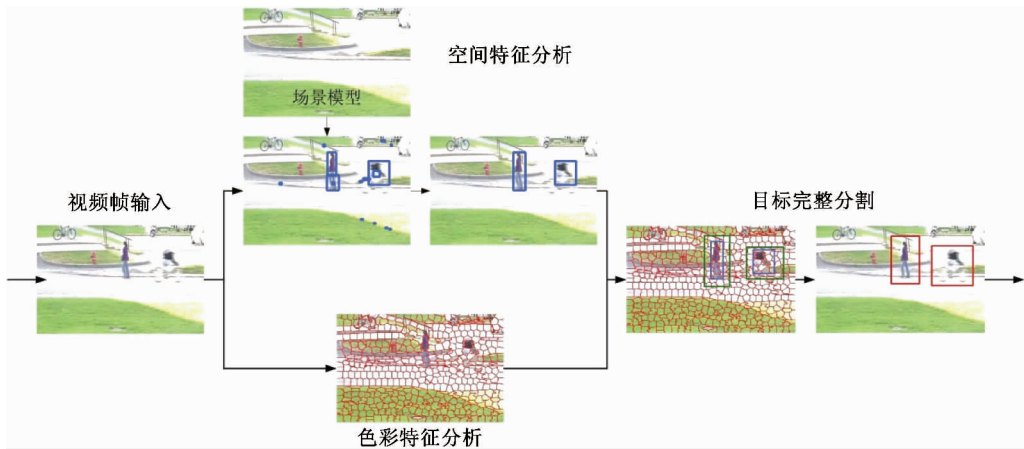


图 2 基于空间与色彩二维特征分析的运动目标 ROI 分割方法示例

1.1 空间维度检测结果聚类与过滤

监控视频中,由于光照和阴影等外界环境的影响,在进行与背景模型的差值运算时,会产生大量的误差信息。如图 3 所示,这些错误的差异信息是将背景信息误检为前景信息导致的,在空间上体现出密度较大、分布较为稀疏等特点。

空间维度的前景点聚合与过滤方法主要包括基于 K-means 的检测结果聚类和基于 T 离群分析的误差过滤 2 部分。接下来,将对这 2 部分进行详细介绍。

1.1.1 基于 K-means 的检测结果聚类

K-means 算法^[24]属于经典的无监督学习算法,聚类的样本中没有给定数据的类别标签,只有数据



图 3 背景差方法的检测结果示例

的特征,聚类的目的是找到每个数据样本所属的类别,并将同类别的样本放在一起,达到同一类间样本相似度最高,而不同类之间样本相似度最低的目的。因此,通过背景差法得到目标区域的二值图像之后,

搜索到每一个前景点集合的最小外接矩形,并得到外接矩形的位置信息,包括位置和尺寸信息。在得到目标区域位置信息之后,可以看出目标区域可能被分为多个部分。随后,可以通过聚类算法实现目标区域的定位。

算法基本流程如下所述:

第 1 步,初始化。将每一个图像块的中心点作为该图像块的表征,定义为 (x_c, y_c) ,并初始化 K 个点 $\mu_1, \mu_2, \dots, \mu_K \in R^n$ 作为聚合的种子点。

第 2 步,距离计算。计算每个区域中心点 (x_c, y_c) 与初始化 K 个种子点之间的距离。针对每一个聚类中心,计算剩余的所有目标区域中心点到该聚类中心的距离,选出距离该聚类中心最近的几个样本作为一类,如式(1)所示:

$$c^{(i)} = \operatorname{argmin} \|x^{(i)} - \mu_j\|^2 \tag{1}$$

其中, $x^{(i)}$ 代表当前目标区域的中心点集。

第 3 步,更新聚类中心点。针对第 2 步得到的类,根据每一类的目标位置,重新估计该类的中心点位置,如式(2)所示:

$$\mu_j = \frac{\sum_{i=1}^m 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}} \tag{2}$$

其中, m 为初步得到的目标数目,1 为真值判断。

第 4 步,迭代计算。如图 4 所示,重复上述过程直至收敛,即目标区域的中心点位置不再变化,实现将目标区域划分为 L 簇,则 L 为当前场景中的目标区域数目,完成目标区域数目和位置的定位。



图 4 基于 K-means 的目标定位示例

1.1.2 基于 T 离群分析的误差过滤

在目标检测算法输出的二值图像中,包含了大量的误差区域,该类区域基本为光影、噪声等干扰所

导致,而非场景中的真实目标,需要进行过滤。通过分析初步分割结果可知,在图像的分割结果中,面积最大的区域一般为真实目标区域的位置所在,因此,以最大面积区域为基准,计算不同量纲的数据,然后通过标准化处理,基于离群检测对目标离群点进行过滤。从图 5 中可以看出,由于光影等的影响,会出现离群区域,不属于目标,因此需要过滤删除。



图 5 聚合结果的光影误差示例

在光影过滤过程中,首先计算不同量纲数据,分别为边距、圆心距和重叠面积,计算方法如下。

(1) 边距

以检测结果中最大面积的矩形为标准,假设当前目标区域为第 i 个区域,则边距为

$$d_L(i) = d_{x, \min}(i) + d_{y, \min}(i) \tag{3}$$

其中, $d_{y, \min}(i)$ 为当前目标区域 i 与最大面积目标区域在 x 方向的最小距离,同样地, $d_{y, \min}(i)$ 为当前目标区域 i 与最大面积目标区域在 y 方向的最小距离。然后对得到的边距进行归一化处理,归一化后的边距为

$$\theta_L(i) = \frac{d_L(i) - d_{L, \text{avg}}}{d_{L, \text{max}}} \tag{4}$$

其中, $d_{L, \text{avg}}$ 表示所有前景图像块和最大目标块之间距离的平均值, $d_{L, \text{max}}$ 表示这些距离中的最大值。令前景目标块的数量为 N ,则 $d_{L, \text{avg}}$ 可以表示为

$$d_{L, \text{avg}} = \frac{1}{N-1} \sum_{i=1}^{N-1} d_L(i) \tag{5}$$

(2) 圆心距

以面积最大的目标区域为标准,计算检测结果中其他区域到最大面积区域中心点之间的距离,并进行归一化处理,假设当前矩形中心点坐标为 (x_i, y_i) ,最大面积矩形的中心点为 (x_o, y_o) ,则区域的圆心距为

$$d_o(i) = \sqrt{(x_o - x_i)^2 + (y_o - y_i)^2} \quad (6)$$

对得到的圆心距进行归一化处理,即:

$$\theta_o(i) = \frac{d_o(i) - d_{o,avg}}{d_{o,max}} \quad (7)$$

其中, $d_{o,avg}$ 表示所有前景图像块和最大目标块之间圆心距的平均值, $d_{o,max}$ 表示圆心距最大值。

(3) 重叠面积

分割结果中目标区域面积最大的区域为目标所处区域之内,因此,其他目标区域与最大区域的重叠面积可以作为是否检测准确的衡量标准之一。归一化之后的重叠区域面积为

$$\theta_s(i) = \frac{S_{OL}(i)}{S_{OL,max}} \quad (8)$$

其中, $S_{OL}(i)$ 为前景块 i 与最大前景块之间的重合面积, $S_{OL,max}$ 为所有块中得到的最大重叠面积。

接下来,对上述得到的 3 种数据进行标准化,数据的标准化是将数据按比例缩放,使之落入一个小的特定区间,可去除数据的单位限制,将其转化为无量纲的数值,便于将不同单位或量级的指标进行比较和加权。数据的归一化是典型的标准化处理,即将数据统一映射到 $[0,1]$ 区间上。目前数据标准化方法有多种,归结起来可以分为直线型方法(如极值法、标准差法)、折线型方法(如三折线法)、曲线型方法(如半正态分布)。不同的标准化方法,对系统的评价结果会产生不同的影响。本文对所得到的数据进行加权平均处理,如式(9)所示。

$$\theta_i = \frac{1}{1 + \exp\left(-\frac{1}{3}(\theta_L(i) + \theta_o(i) + \theta_s(i))\right)} \quad (9)$$

最后,对得到的标准化后的值进行 T 离群检测,分析出一组数据中的离群点。T 离群检测方法^[25]主要用于样本含量较小(例如 $n < 30$)、总体标准差 σ 未知的分布。T 检测用于检验 2 个小样本的平均值差异程度,然后用 T 分布理论来推断差异发生的概率,从而判断 2 个小样本的差异是否显著,是一种检验随机变量的数学期望是否等于某个已知值的假设检验方法。T 检验法在处理数据前,预先剔除了被检验的值,保证了计算标准偏差的正确性和独立性,在理论上得到了较严格的结果,提高了检验

的精度和灵敏度。

单总体 T 检验是检验一个样本与一个已知的总体平均数的差异是否显著。单总体 T 检验统计量为

$$t = \frac{\bar{X} - \mu}{\sigma_X / \sqrt{n-1}} \quad (10)$$

令 $\bar{\theta}$ 为样本平均数, n 表示当前图像中包含的目标数目。因此,样本标准偏差可以表示为

$$s = \sqrt{\sum_{i=1}^n \frac{(\theta_i - \bar{\theta})^2}{n-1}} \quad (11)$$

该统计量在零假说 $\mu = \mu_o$ 为真的条件下服从自由度为 $n-1$ 的 T 分布。使用 T 检验法检验离群点,统计量可以表示为

$$K = \frac{|\theta_d - \bar{\theta}|}{S} \quad (12)$$

其中 θ_d 指当前被检验的点, $\bar{\theta}$ 和 S 是由不包括当前点在外的 $n-1$ 个数据计算得到的算术平均值和标准偏差。根据事先确定的置信度和测定次数查表,如果 K 大于所查数值,则 θ_d 离群点为异常值,予以舍弃,否则应予以保留。

离群检测结果示例如图 6 所示,第 2 行矩形为离群检测之后的结果,可以看出,经过检测之后,错误的前景块被删除,检测结果更加准确。接下来,将所有正确的矩形进行取并集运算,可以得到空间维度的检测结果,如第 3 行矩形所示。可见,经过离群检测之后可以使目标提取的结果更准确,包含较少冗余信息,减少伪目标区域对后续处理的影响。



图 6 离群检测结果示例

1.2 色彩维度的目标精确分割

在完成空间维度的聚合和过滤之后,为了更精确地提取目标区域,避免色彩相似对于目标边缘的影响,需采取色彩维度的特征聚合方法,如超像素分割算法等。色彩维度的特征分析主要解决目标区域的精确分割问题,图像的色彩特征可以直观反映出所包含的信息,而且无需复杂的计算,只需要将图像中的像素值进行相应的转换,因此,色彩特征成为表达图像信息的一个较好的方式。基于此,本文结合超像素分割算法,实现 ROI 的完整分割。

超像素是指具有相似纹理、颜色、亮度等特征的相邻像素所构成的具有一定视觉意义的不规则像素块。该理念利用像素之间特征的相似性,将图像中的像素点分组,用少量的超像素代替大量的像素来表达图像信息,该分割方法可以实现目标区域边界的完整提取。线性迭代聚合 (simple linear iterative clustering, SLIC) 算法^[26] 是简单的线性迭代聚类算法,思想简单、易于实现,该算法首先将彩色图像转换为 CIELAB 颜色空间中的 5 维特征向量,然后对 5 维特征向量构造距离度量标准,以此对图像像素进行聚类,该算法可以生成紧凑、近似均匀的分割效果,在运算速度、物体轮廓保持、形状均匀性等方面有较高的性能,符合期望的分割效果。

在 SLIC 算法中,用于图像分割的色彩特征为 CIELAB 颜色空间,其中,L 表示亮度,值域为 0(黑色)到 100(白色);A 表示从品红色至绿色的范围(A 为负值指示绿色而正值指示品红色);B 表示从黄色至蓝色的范围(B 为负值指示蓝色而正值指示黄色)。SLIC 分割算法的具体实现步骤如下所述:

第 1 步,初始化种子点(聚类中心)。按照初始化设定的超像素个数,在图像内均匀分配种子点。假设图片总共有 N 个像素点,预分割为 K 个相同尺寸的超像素,那么每个超像素的大小为 N/K ,则相邻种子点的距离(步长)近似为 $S = \sqrt{N/K}$ 。

第 2 步,在种子点邻域内重新选取种子点。计算周围小邻域内所有像素点的梯度值,将种子点移到该邻域内梯度最小的地方,这样做的目的是为了

避免种子点落在梯度较大的轮廓边界上或者是图像中的噪点上而影响后续聚类效果。

第 3 步,为每个种子点的 $S \times S$ 邻域内像素点分配标签(即属于哪个聚类中心)。与标准的 K-means 算法搜索整幅图像不同,SLIC 的搜索范围限制为 $2S \times 2S$,可以加速算法的收敛速度。在给定的搜索范围内,度量每个像素点到种子点之间的距离,包括颜色距离和空间距离,限制搜索区域的大小显著地减少了需要计算距离的数目,将算法控制为线性复杂度。

第 4 步,距离测量。对于搜索范围内的每个像素点,根据欧氏距离公式,计算该像素点和种子点之间的距离,具体计算方式如下:

$$d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \tag{13}$$

$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{14}$$

$$D' = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{m}\right)^2} \tag{15}$$

其中, d_c 代表 Lab 空间的色度距离, d_s 代表空间距离, m 代表紧密度参数,通常取一个固定常数,适用于每个聚类, $N_s = S$ 是类内的最大空间距离。

第 5 步,迭代优化。由于搜索范围为 $2S \times 2S$,所以每个像素点会被周围多个种子点搜索到,取多个距离中的最小值对应的种子点作为该像素点所属的聚类中心。理论上,上述步骤可以不断迭代直至收敛(即每个像素点所属的种子点不再发生变化为止),实验中发现经过 10 次迭代后,大部分图片可以得到较理想的效果,所以一般迭代次数取 10。

第 6 步,增强连通性。经过上述迭代操作后可能会出现以下瑕疵:出现多连通情况、超像素尺寸过小,单个超像素被分割成多个不连续超像素等,这些情况可以通过增强连通性解决。主要思路是新建一张标记表,对表内元素标记初始值,按照 Z 字走向(从左到右,从上到下顺序)将不连续的超像素、尺寸过小超像素重新分配给临近的超像素,遍历过的像素点分配给相应的标签,直到所有点遍历完毕为止。

如图 7 所示,网络形状即为超像素视图,矩形区域为结合色彩维度分析出的连续性信息和空间维度分析结果取并集得到的完整 ROI 区域。



图7 色彩维度分析结果示例

2 实验方法及结果分析

本节将对本文所提方法进行性能测试与评估。本文的实验环境为 Intel Core i7, 内存为 8G。实验数据选自 CDnet2014^[27] 数据集, 包括该数据集中 baseline 分类里的 office 测试例和 highway 测试例, shadow 分类里的 bus station 测试集, 覆盖了室内、室外、静止、运动等多个典型视频监控场景。

实验对比的算法是目前检测算法中性能较好的 GMM 算法^[9]、ViBe 算法^[11] 和 PBAS 算法^[12]。其中, 本章对 GMM、ViBe 和 PBAS 算法采用了传统的形态学处理, 包括基于膨胀、腐蚀效果的开、闭运算, 从而消除了部分检测结果中的漏检和误检信息。

此外, 对比实验中选择了目标区域的交并比指标(intersection over union, IoU)和完整率指标(intact rate, IR)。其中, IoU 指标表示由算法检测出的 ROI 区域与真实的 ROI 区域的交集与并集的比值。可以看出, 无论是交集区域过大或并集区域过大都无法获得较好的 IoU 评分, 只有当交集和并集相似, 即检测区域和真实区域相近时, 才能获得较高的 IoU 评分。同时, 在 IoU 指标的测量中, 会出现小于真实区域的检测结果和大于真实区域的检测结果的 IoU 分数相同的情况。在这 2 种情况中, 检测区域大于真实区域意味着保留了更完整的目标信息。因此, 本文进一步定义了完整性指标 IR, 用于表示在用于检测的视频序列中, 检测算法得到的检测区域大于真实目标区域的视频帧在全部视频帧中所占的比例, 即检测的完整性。因此, 当 2 种算法的 IoU 指标相近时, IR 指标越大, 证明该算法分割出的目标区域越完整。

首先, 本文对 GMM、PBAS、ViBe 算法和本文方法的运动目标分割结果进行了主观对比分析。实验结果示例如图 8 所示。highway 场景和 office 场景中, GMM 算法和 ViBe 算法尽管采用了形态学的处理, 但由于误检情况比较严重, 因此分割的 ROI 远大于基于 groundtruth 提取的标准 ROI。这也证明, 在传统的背景建模类算法中, 形态学处理仅能起到辅助作用, 无法从根本上解决目标检测结果中的误判现象。此外, PBAS 算法在 3 个场景下的分割结果尽管优于 GMM 和 ViBe 算法, 但局部缺失现象严重。相比上述对比算法, 本文方法分割出的 ROI 分割结果, 不仅提高了运动目标 ROI 分割的准确率, 还大幅提高了完整性。

4 种算法在上述测试场景中的平均 IoU 和 IR 指标如图 9 所示。从实验结果可以看出, 与 GMM 算法相比, 本文所提出的目标区域分割算法在 IoU 指标小幅下降的前提下 ($< 3\%$), 能够提高 25% 左右的检测完整性, 优势非常明显。此外, 在本文算法与 ViBe 算法和 PBAS 算法的对比中, 本文算法无论在 IoU 指标和 IR 指标上都要优于其他 2 种算法, 其中, 本文算法的 IoU 指标比 ViBe 算法和 PBAS 算法分别提高了约 7% 和 12%, IR 指标分别提高了约 17% 和 55%。实验结果表明, 本文算法不仅在目标区域分割的准确率上明显优于现有背景建模算法, 而且分割结果的完整性优势非常明显。

3 结论

本文提出了一种面向监控视频的 ROI 完整分割算法。首先, 通过对比视频监控的背景模型, 基于帧差法计算得到前景信息, 并通过空间的前景信息聚类与误差删除操作, 过滤掉了大部分误差信息; 随后, 对当前视频帧采取了超像素分割处理, 从而得到接近目标实际轮廓的超像素边界集合; 最后, 通过对空间和色彩 2 个维度下得到的结果求并集, 实现了 ROI 的完整分割。仿真结果表明, 本文提出的目标区域完整提取算法与现有的 GMM 算法、ViBe 算法和 PBAS 算法相比, 能够在目标区域的准确率不下降甚至小幅度提高的前提下, 大幅增加完整分割的

概率,从而避免目标区域局部漏检导致的关键信息丢失问题。



图 8 不同算法的检测结果示例

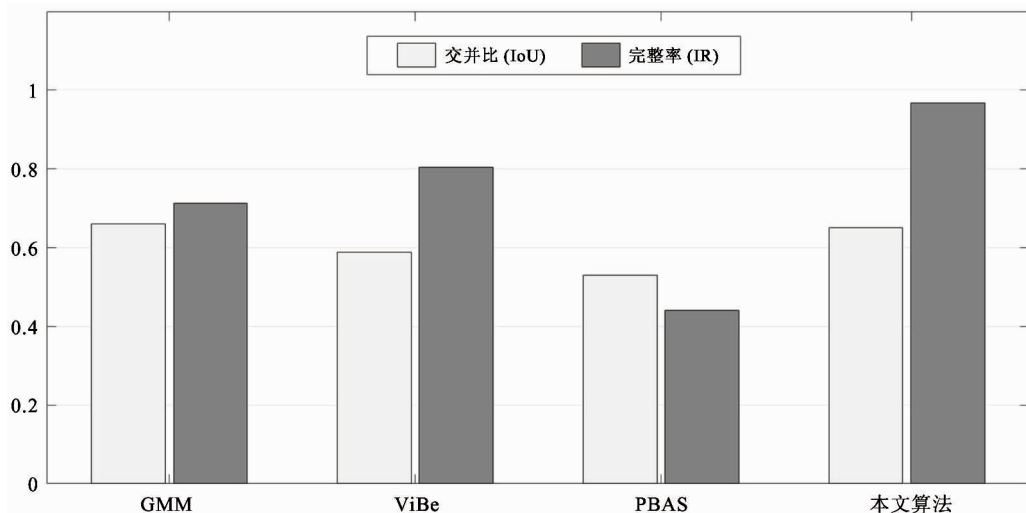


图 9 不同算法的检测结果对比

参考文献

- [1] Liu L, Zhou Y Q, Garcia V, et al. Load aware joint CoMP clustering and Inter-cell resource scheduling in heterogeneous ultra dense cellular networks [J]. *IEEE Transaction on Vehicular Technology*, 2018, 67 (3) : 2741-2755
- [2] Zhou Y Q, Liu H, Pan Z D, et al. Energy efficient two-stage cooperative multicast based on device to device transmissions: effect of user density [J]. *IEEE Transaction on Vehicular Technology*, 2016, 65 (9) : 7297-7307
- [3] Liu L, Zhou Y Q, Tian L, et al. CPC-based backward compatible network access for LTE cognitive radio cellular networks [J]. *IEEE Communications Magazine*, 2015, 53 (7) : 93-99
- [4] 黄铁军, 张贤国, 田永鸿, 等. 支持监控视频高效压缩与识别的 IEEE 1857 标准 [J]. *电子产品世界*, 2013, 20 (7) : 22-26
- [5] Liu Z, Huang K, Tan T. Foreground object detection using top-down information based on EM framework [J]. *IEEE Transactions on Image Processing*, 2012, 21 (9) : 4204-4217
- [6] 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
- [7] David C, Gui V. Automatic background subtraction in a sparse representation framework [C] // 2013 20th International Conference on Systems, Signals and Image Processing, Bucharest, Romania, 2013 : 63-66
- [8] Bayestehtashk A, Shafran I. Parsimonious multivariate copula model for density estimation [C] // 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, Canada, 2013 : 5750-5754
- [9] Zhu Q, Song Z, Xie Y. An efficient r-KDE model for the segmentation of dynamic scenes [C] // Proceedings of the 21st International Conference on Pattern Recognition, Tsukuba, Japan, 2012 : 198-201
- [10] Liu Z, Huang K, Tan T. Cast shadow removal in a hierarchical manner using MRF [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2012, 22 (1) : 56-66
- [11] Barnich O, Van D M. ViBe: a powerful random technique to estimate the background in video sequences [C] // 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, China, 2009 : 945-948
- [12] Hofmann M, Tiefenbacher P, Rigoll G. Background segmentation with feedback: The pixel-based adaptive segmenter [C] // 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, USA, 2012 : 38-43
- [13] Xu Y, Dong J, Zhang B, et al. Background modeling methods in video analysis: a review and comparative evaluation [J]. *CAAI Transactions on Intelligence Technology*, 2016, 1 (1) : 43-60
- [14] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C] // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005 : 886-893
- [15] Viola P, Jones M J. Robust real-time face detection [J]. *International Journal of Computer Vision*, 2004, 57 (2) : 137-154
- [16] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*,

- 2004, 60(2): 91-110
- [17] Bay H, Tuytelaars T, Gool L V. SURF: speeded up robust features [C] // European Conference on Computer Vision, Graz, Austria, 2006: 404-417
- [18] Fergus R, Perona P, Zisserman A. Object class recognition by unsupervised scale-invariant learning [C] // Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, USA, 2003: 264-271
- [19] Girshick R B, Felzenszwalb P F, Mcallester D A. Object detection with grammar models [C] // Proceedings of the 24th International Conference on Neural Information Processing Systems, Granada, Spain, 2011: 442-450
- [20] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Washington DC, USA, 2014: 580-587
- [21] Redmon J, Farhadi A. YOLOv3: an incremental improvement [J]. *arXiv*:1804.02767, 2018
- [22] Veeramani B, Raymond J W, Chanda P. DeepSort: deep convolutional networks for sorting haploid maize seeds [J]. *BMC Bioinformatics*, 2018, 19(9): 85-93
- [23] 黄凯奇, 陈晓棠, 康运锋, 等. 智能视频监控技术综述 [J]. 计算机学报, 2015, 20(3): 1093-1118
- [24] Ren X, Malik J. Learning a classification model for segmentation [C] // Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France, 2003: 13-16
- [25] Williamson D F, Parket P A, Kendrick J S. The box plot: A simple visual method to interpret data [J]. *Annals of Internal Medicine*, 1989, 110(11): 916-921
- [26] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274-2282
- [27] Wang Y, Jodoin P M, Porikli F, et al. CDnet 2014: an expanded change detection benchmark dataset [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, USA, 2014: 387-394

A novel intact ROI segmentation method for surveillance video

Liu Chang^{* ** ** *}, Wang Pengjun^{***}, Zhao Xiao^{**}, Tian Lin^{* **}, Zhou Yiqing^{* ** ** *}, Shi Jinglin^{* ** ** *}

(^{*} Beijing Key Laboratory of Mobile Computing and Pervasive Device, Beijing 100190)

(^{**} Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

(^{***} University of Chinese Academy of Sciences, Beijing 100049)

(^{****} Unit 96901, Beijing 100094)

Abstract

Intelligent video surveillance technology has been developing rapidly, in which region of interest (ROI) is the key building block of various subsequent intelligent applications. However, based on the existing moving target detection algorithm, the local missing problem will occur when the region is segmented, which will lead to the loss of key information (moving target information) and affect the subsequent analysis of video. To address this problem, a novel intact ROI segmentation method for surveillance video is proposed in this paper. Both spatial and color features are taken into consideration in the proposed method. For spatial feature, K-means and T outliers detection algorithms are used to filter errors caused by light and shadow. For color feature, superpixel algorithm is used to guarantee continuity of segmentation. Experimental results show that the proposed method can not only achieve better precision and recall ratio, but also improve the intact segmentation ratio significantly.

Key words: intelligent video surveillance, object detection, region of interest (ROI), superpixel, outliers detection