DOI:10.16136/j.joel.2023.04.0339

联合二分思想和时空管道的监控视频运动分割

张云佐*,郭凯娜

(石家庄铁道大学 信息科学与技术学院,河北 石家庄 050043)

摘要:监控视频运动分割是视频浓缩、行为识别等视频智能处理的基础和前提,是计算机视觉领域的研究热点。现存运动分割方法大多步骤繁琐、计算量大,难以应用于计算能力有限的领域。为此,提出了一种联合二分思想和时空管道的监控视频运动分割方法。该方法首先使用嵌套椭圆时空管道模型计算初始累计时空流量来判断目标轨迹完整性(completeness of target trajectory,CTT);然后结合二分思想动态地调节椭圆采样线,自适应地捕捉采样区域的运动目标;最后提取采样线上的全部像素点形成自适应时空管道进行运动分割。实验结果表明,所提方法在保证精度的同时计算速度明显优于对比方法,且所提方法鲁棒性强,对运动情况多变的监控场景同样适用。

关键词:目标轨迹完整性(CTT);二分法;自适应时空管道;运动分割;时空流量

中图分类号:TP391 **文献标识码**:A **文章编号**:1005-0086(2023)04-0371-07

Motion segmentation of surveillance video by combining dichotomous and spatio-temporal tube

ZHANG Yunzuo*, GUO Kaina

(School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, Hebei 050043, China)

Abstract: Motion segmentation of surveillance video is the basis and premise of intelligent video processing such as video synopsis and behavior recognition, and is a research hotspot in the field of computer vision. Nevertheless, existing motion segmentation methods usually suffer from cumbersome steps and mass calculation, and their application is restricted in the field of limited computing power. To address these issues, we propose a method called motion segmentation of surveillance video by combining dichotomous and spatio-temporal tube. Firstly, the initial spatio-temporal flow is calculated using nested elliptical spatio-temporal tube model to judge the completeness of target trajectory (CTT). Secondly, we adjust dynamically the elliptical sampling line by combining the bisection to capture adaptively the moving target in the sampling area. Finally, the pixels on the sampling line are extracted to form the adaptive spatio-temporal tube for motion segmentation. Experimental results demonstrate that the proposed method outperforms the state-of-the-art methods in terms of both computing speed and accuracy, which has strong robustness and is also suitable for surveillance scenarios with changing motion.

Key words: completeness of target trajectory (CTT); bisection; adaptive spatio-temporal tube; motion segmentation; spatio-temporal flow

0 引 言

随着智能媒体的发展,视频监控系统在日常生活和安保安防中都得到了广泛的应用,与此同时也产生了大量的监控视频数据,为人们的快速浏览和检索带来了困难[1,2]。运动分割技术可以

* E-mail:zhangyunzuo888@sina.com

收稿日期:2022-05-09 修订日期:2022-08-12

基金项目:国家自然科学基金(61702347,62027801)、中央引导地方科技发展资金项目(226Z0501G)、河北省自然科学基金(F2022210007,F2017210161)和河北省高等学校科学技术研究项目(ZD2022100,QN2017132)资助项目

从海量的、存在大量静止片段的监控视频中提取出运动片段，是实现冗长视频快速获取信息的有效手段[3,4]。

无监督的运动分割方法大多以目标的外观、运动特征等底层视觉信息为依据检测运动目标，再提取存在运动目标的帧形成运动片段。文献[5]通过计算历史运动图像的能量反映视频中运动目标的时空信息，实现运动分割；文献[6]使用混合滤波器改进了 Canny 边缘算子，并结合高斯混合模型检测运动目标；文献[7]改进了传统光流法，通过额外计算光流场 4 个边界的平均运动向量来提高运动分割的精度；文献[8]提出了基于自适应混合高斯的改进三帧差分算法，通过基于边缘提取的三帧差分改进算法对视频图像的目标轮廓进行提取，保证目标信息的完整。使用上述方法进行运动分割需要逐帧逐像素处理数据，计算量极大，耗费时间长。

随着人工智能的兴起，深度学习方法被广泛应用于目标检测和运动分割领域[9,10]。文献[11]提出了一个对象引导的外部存储网络，存储效率由对象引导的硬注意力处理，以选择性地存储有价值的特征；文献[12]提出了一个集成检测、时间传播和跨尺度细化的统一框架，以便寻求平衡性能和成本的策略；文献[13]提出了一套从零开始学习的目标检测器设计原则，其关键是深度监督，由骨干网络和预测层中的分层密集连接实现，在从头开始学习的检测器中起着关键作用；文献[14]提出了一种基于空间－通道注意力机制的 SSD（single shot multibox detector）目标检测算法来增强高层特征图语义信息，提高了目标的检测精度。基于深度学习的方法已被证明是有效的，但仍存在许多缺陷，一方面，基于深度学习的方法需要大量的数据预先训练模型，在数据量有限的应用场景下，无法实现对数据规律的无偏差估计；另一方面，基于深度学习的模型计算复杂度高，在计算能力有限的领域难以应用。

针对现存方法计算量大、计算复杂度高的问题，文献[15]提出了一种嵌套椭圆时空管道方法进行运动分割，该方法只处理采样线上的像素而无需处理视频空间域的全量数据，减小了计算量。但在真实的监控视频场景中，目标的运动方式具有不确定性和随机性，嵌套椭圆时空管道方法需要人工确定采样线的数量和位置，虽然在一些视频上取得了较为理想的效果，但对运动情况多变的监控视频无法兼顾运算速度和精确度，鲁棒性差。为此，本文提出一种联合时空管道和二分思想的监控视频运动分割方法，该方法定义了用于判断监控视频中目标运动情况的目标轨迹完整性（completeness of target trajectory，CTT），动态地调节采样线来构建自适应时空管道，最后根据累计时空流量进行运动分割。

1　嵌套椭圆时空管道模型

时空管道是通过对视频序列采样并提取采样线上的像素点形成的，是一种结合空间维度与时间维度的视频分析方法[15]。嵌套椭圆时空管道根据不同宽高比的视频序列，由内切的椭圆采样线递进采样产生。在如图 1 所示监控视频中，利用 3 条椭圆采样线将视频分割为 4 个采样区域并在每一帧的相同位置上采样，提取采样线上的像素并按照时间顺序排列。一条采样线上的像素产生一个时空管道，递进椭圆采样即形成了嵌套椭圆时空管道。

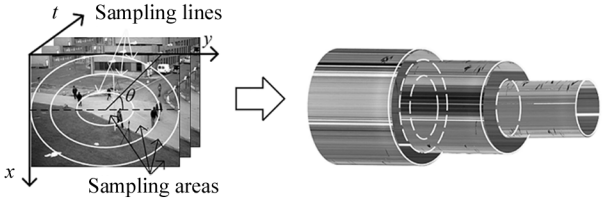

图 1　嵌套椭圆时空管道的形成

Fig. 1　Formation of nested elliptical spatio-temporal tube

该模型提取每个时空管道上的像素以便进行运动分析，为了计算每条采样线上的像素点坐标，首先假设某一采样线的周长为 L，定义图 1 中 θ 为：

$$\theta=\frac{2\pi S}{L},(S=1,2,\cdots,L)\ , \tag{1}$$

式中，S 是 θ 的对应弧长，根据式(2)—(5)计算采样线上的像素点坐标 (x,y)：

$0\leqslant\theta<\frac{\pi}{2}$ 或 $\frac{3\pi}{2}<\theta\leqslant 2\pi$ 时：

$$\begin{cases}x=\dfrac{ab}{\sqrt{b^2+a^2\tan^2\theta}}+b\\ y=\dfrac{ab\tan\theta}{\sqrt{b^2+a^2\tan^2\theta}}+a\end{cases}\ , \tag{2}$$

$\frac{\pi}{2}<\theta<\frac{3\pi}{2}$ 时，

$$\begin{cases}x=-\dfrac{ab}{\sqrt{b^2+a^2\tan^2\theta}}+b\\ y=-\dfrac{ab\tan\theta}{\sqrt{b^2+a^2\tan^2\theta}}+a\end{cases}\ , \tag{3}$$

$\theta=\frac{\pi}{2}$ 时，

$$x=b,y=a+b\ , \tag{4}$$

$\theta=\frac{3\pi}{2}$ 时，

$$x = b, y = a - b, \tag{5}$$

式中,a、b 分别表示椭圆采样线长轴和短轴的长度。

将提取出的像素自顶向下排列形成长度为 L 的列向量,将每一帧提取的列向量依次排列生成时空平面图 $\boldsymbol{S}$:

$$\boldsymbol{S} = \begin{bmatrix} (x_{i_1}^1, y_{j_1}^1) & (x_{i_1}^2, y_{j_1}^2) & \cdots & (x_{i_1}^T, y_{j_1}^T) \\ (x_{i_2}^1, y_{j_2}^1) & (x_{i_2}^2, y_{j_2}^2) & \cdots & (x_{i_2}^T, y_{j_2}^T) \\ \vdots & \vdots & \ddots & \vdots \\ (x_{i_L}^1, y_{j_L}^1) & (x_{i_L}^2, y_{j_L}^2) & \cdots & (x_{i_L}^T, y_{j_L}^T) \end{bmatrix}, \tag{6}$$

式中,(x_i^k, y_j^k) 表示视频序列第 k 帧上第 i 行第 j 列的像素。为了分析监控视频中目标的运动方向,对时空平面图进行混合高斯建模提取运动前景,通过次采样线[15]确定目标在采样区域的进入和退出。将进入采样区域的运动目标像素值赋为 1、退出采样区域的运动目标像素值赋为 −1 计算累计时空流量[16]。将视频帧中采样线上的像素值求和可以得到视频帧的瞬时时空流量,累加第 1 帧到第 k 帧的瞬时时空流量即为第 k 帧的累计时空流量 $AF(f_k)$,其计算式为:

$$AF(f_k) = \sum_{i=1}^{k} \sum_{j=1}^{L} P_{j,i}^k, \tag{7}$$

式中,$P_{j,i}^k$ 表示像素点 (x_j^k, y_i^k) 对应的像素值,f_k 表示输入视频的第 k 帧,L 为采样线的周长。不同于传统的对监控视频逐帧逐像素分析的前景提取方法,嵌套椭圆时空管道模型通过提取采样像素的方式只分析视频帧的小部分像素,大大减少了计算量。

2　结合二分思想的自适应时空管道

本文提出了一种联合时空管道和二分思想的监控视频运动分割方法,该方法的整体流程图如图 2 所示,主要包括 3 个部分:1)判断输入视频的 CTT;2)二分法动态调节采样线;3)生成自适应时空管道分割运动片段。以下将对其具体过程进行阐述。

2.1　CTT

在监控视频中,运动目标可能只跨越采样线一次,如图 3 所示,目标由采样区域右侧边界进入,沿图中虚线运动至采样区域中的门内并退出采样区域,此时嵌套椭圆时空管道模型的内侧采样线只能捕捉到目标进入采样区域,无法捕捉到目标退出采样区域,因为该模型需人工确定采样线,难以应对运动情况多变的监控场景,鲁棒性差。类似的情况还包括目标由门后进入采样区域再运动至采样区域边界并退出,本文将此类情形定义为目标的运动轨迹不完整,即不满足 CTT。

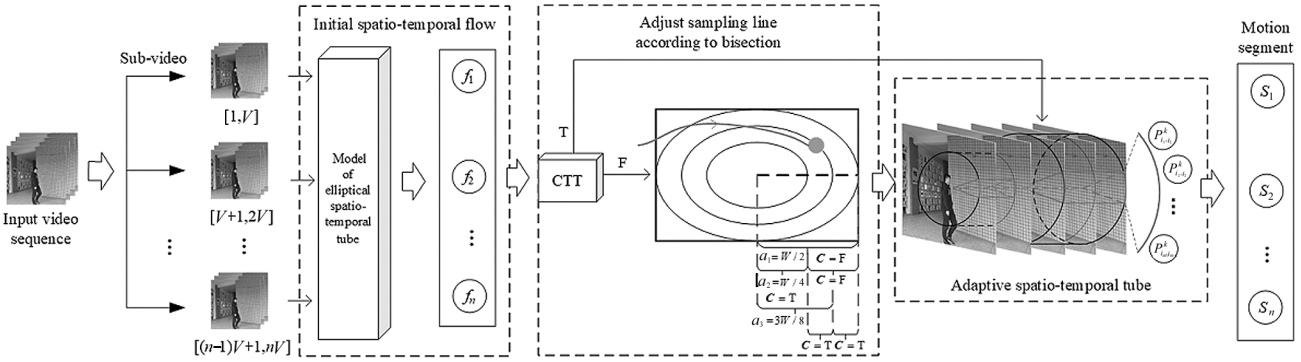

图 2　本文所提方法流程图

Fig. 2　The flow chart of the proposed method in this paper

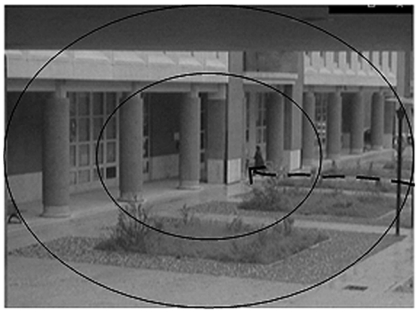

图 3　目标轨迹不完整

Fig. 3　Incomplete target trajectory

在现实需求中,检索监控视频中的运动目标时,针对的多为人、车、宠物等具有一定体积的目标,占据监控视频中的多个像素而不是一个[17],因此对于大小为 $W \times H \times T$ 的视频序列,定义运动分割的容差值 δ:

$$\delta = \frac{3c}{p}, \tag{8}$$

式中,$c = \min(W, H)$,p 为视频序列的帧率,W、H、T 分别为视频帧的宽、高和视频序列的长度。

设置 CTT 向量 $\boldsymbol{C}$,将不满足 CTT 的视频定义为

$\boldsymbol{C}$=F，满足 CTT 的视频定义为 $\boldsymbol{C}$=T。计算视频的初始累计时空流量，利用 $\boldsymbol{C}$ 判断视频的 CTT：

$$\boldsymbol{C}=\begin{cases}\mathrm{F}, & |AF(f_{\mathrm{end}})|>\delta\\ \mathrm{T}, & \text{else}\end{cases}, \tag{9}$$

式中，f_{end} 为当前子视频段的结束帧，AF 为初始累计时空流量，δ 为容差值。

2.2 二分思想划分采样线

二分查找是一种快速检索方法，其原理是将待查找的元素与有序数列的中间元素进行比较，分析比较结果舍弃不符合条件的一半元素，再将待查找元素与保留部分的中间元素进行比较，反复循环搜索直到找到目标元素的位置。二分思想将一个大问题一分为二，每次只需要处理其中的一个小问题。本文将监控视频等分成小的子视频段，每次只处理一个子视频段，结合二分思想自适应地划分采样线，动态地寻找当前子视频段的最优采样位置。

本文利用二分思想调整采样线，对 $\boldsymbol{C}$=T 的采样区域直接根据累计时空流量分割运动片段，对 $\boldsymbol{C}$=T 的子视频段自适应地调整采样线位置直至采样区域的 CTT 满足 $\boldsymbol{C}$=T。假设子视频段长度为 N，视频帧的宽和高分别为 W 和 H，对输入的子视频段采用内切于视频帧的椭圆采样线作为初始采样线，初始采样参数为 $q_1=\{a_1=W/2,b_1=H/2\}$，$l=1$，其中 a、b 分别为采样椭圆长轴和短轴的长度，l 为椭圆采样线的数量，根据式(7)—(9)判断子视频段的 CTT。

对于 CTT 向量为 $\boldsymbol{C}$=T 的子视频段，说明一条采样线即可完整捕捉采样区域的运动目标并完成运动分割；对于 CTT 向量为 $\boldsymbol{C}$=F 的子视频段添加采样线进一步采样。设置参数 $i=1$，并计算第一个折半采样参数 $q_2=\{a_2=W/4,b_2=H/4\}$，$l=2$。

分析由采样参数 q_l 形成的内外两个采样区域 Ω_{l-1} 和 Ω_l 的 CTT，计算后的结果存在 4 种可能的组合，采样区域 Ω_{l-1} 和 Ω_l 的 CTT 向量分别为 $\boldsymbol{C}$=F 和 $\boldsymbol{C}$=F、$\boldsymbol{C}$=F 和 $\boldsymbol{C}$=T、$\boldsymbol{C}$=T 和 $\boldsymbol{C}$=F、$\boldsymbol{C}$=T 和 $\boldsymbol{C}$=T。对 $\boldsymbol{C}$=T 的采样区域无需进一步处理，对 $\boldsymbol{C}$=F 的采样区域，按照式(10)—(14)自适应地调整采样线。

$$i=\begin{cases}2i, & \boldsymbol{C}(\Omega_{l-1})=\mathrm{F}\\ 2i+1, & \boldsymbol{C}(\Omega_l)=\mathrm{F}\end{cases}, \tag{10}$$

$$l=l+1, \tag{11}$$

$$d=\lfloor\log_2 i\rfloor+1, \tag{12}$$

$$j=i-(2^{d-1}-1), \tag{13}$$

$$q_l=\frac{2j-1}{2^{d-1}}\times q_{l-1}。 \tag{14}$$

对新产生的采样区域计算累计时空流量并判断 CTT，直到子视频段所有采样区域的 CTT 向量都满足 $\boldsymbol{C}$=T，表明当前子视频段的计算已经完成。

2.3 运动分割

完成自适应采样的子视频段 $V=\{x_k\}_{k=1}^{T/n}$，帧号为 $[1,T/n]$，提取采样线上的像素，构建自适应时空管道，根据式(7)计算累计时空流量并记录：

$$Y=\{y_k\}_{k=1}^{T/n}, \tag{15}$$

其中，

$$y_k=\begin{cases}1, & AF(f_k)>0\\ 0, & \text{else}\end{cases}, \tag{16}$$

子视频段分割出的运动片段 S_v 为：

$$S_v=\{x_k\mid k\in E\}, \tag{17}$$

式中，$E=\{k\in[1,T/n]\mid y_k=1\}$。将每个计算完成的子视频段的运动片段拼接并输出，生成监控视频的运动片段。

3 算法流程

本文算法核心是对不同的子视频段动态采样以形成自适应时空管道，最大限度地兼顾计算量和精确度，图 4 清晰直观地展示了采样动态线调整方法，图中实心圆形表示运动目标，其连接曲线表示目标的运动轨迹。具体计算步骤如下：

步骤 1：输入大小为 $W\times H\times T$ 的视频序列，将监控视频分割成 n 个子视频段，每个子视频段的长度为 T/n，确定初始采样参数为 $q_1=\{a_1=W/2,b_1=H/2\}$，该采样线在图 4 中标号为①；

步骤 2：计算子视频段的初始累计时空流量，得到采样区域的 CTT 向量为 $\boldsymbol{C}$=F；

步骤 3：开始折半查找，确定第一个折半采样参数为 $q_2=\{a_2=W/4,b_2=H/4\}$，生成的采样线在图 4 中标号为②，设置参数 $i=1$，获得新的采样区域 Ω_1 和 Ω_2；

步骤 4：计算累计时空流量得到 $\boldsymbol{C}(\Omega_1)$=T、$\boldsymbol{C}(\Omega_2)$=F，采样区域 Ω_1 无需再处理。

步骤 5：对采样区域 Ω_2 继续折半查找，根据式(10)令 $i=3$，调整采样线的数量 $l=3$，根据式(12)—(14)计算得出 $q_3=\{a_3=3W/8,b_3=3H/8\}$，生成的采样线在图中标号为③；

步骤 6：根据采样参数 q_3 更新采样区域 Ω_2 和 Ω_3，分析计算得到 $\boldsymbol{C}(\Omega_2)$=T、$\boldsymbol{C}(\Omega_3)$=T，该子视频段的所有采样区域均满足 $\boldsymbol{C}$=T，停止当前子视频段的采样线划分，根据式(15)—(17)完成运动分割，跳转下一个子视频段并重复步骤 1。

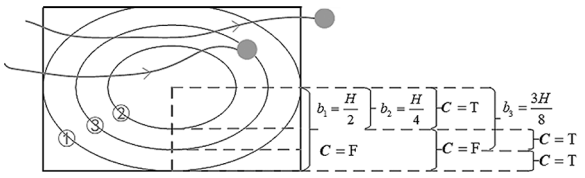

图 4　二分思想调整采样线

Fig. 4　Adjusting the sampling line according to bisection

4　实验结果与分析

本文所有实验均在 Windows10、Intel(R) Core(TM)i5-8265 U CPU、NVIDIA GeForce MX 250 显卡、16G运行内存、64位操作系统的计算机环境下进行,实验用软件为 Matlab R2018b,测试数据集为 10 段真实环境拍摄的监控视频。

4.1　实验对比

为了准确客观地评价运动分割效果,使用计算时间、准确率、召回率和 $F1$ 分数作为评估指标衡量运动分割方法的性能,并与当前主流的文献[5]、文献[7]和文献[15]中的运动分割方法进行实验对比,其中文献[15]中的嵌套椭圆时空管道方法分别选取 1 条椭圆采样线(1E)、2 条椭圆采样线(2E)和 3 条椭圆采样线(3E)进行实验,表 1 展示了对比方法和本文方法在 10 段实验视频上的平均计算时间、平均准确率、平均召回率和平均 $F1$ 分数的值。

表 1　实验结果对比

Tab. 1　Comparison of experimental results

Methods	Average calculating time/s	Average precision/%	Average recall/%	Average $F1$ score/%
Method in [5]	922.09	80.39	87.46	83.78
Method in [7]	714.88	91.78	93.37	92.57
Method in [15] (1E)	33.82	74.73	70.82	72.72
Method in [15] (2E)	75.61	80.26	74.67	77.36
Method in [15] (3E)	139.04	79.10	77.16	78.12
Proposed method	59.35	95.49	91.13	93.25

从表 1 的平均计算时间可以看出,本文方法的平均计算时间远低于文献[5]、文献[7]和文献[15](2E、3E)的计算时间,虽然较高于文献[15] (1E),但本文方法的平均准确率、平均召回率和平均 $F1$ 分数均存在明显优势。假设输入的视频大小为 $W\times H\times T$,本文方法的计算复杂度为 $O(\frac{T}{n}\times(n(4a+2(\pi-2)b)+4\sum_{i=1}^{n}a_i+2(\pi-2)\sum_{i=1}^{n}b_i))$,文献[15]的计算复杂度为 $O(T\times(4\times(a_1+a_2+\cdots)+2\times(\pi-2)\times(b_1+b_2+\cdots)))$,文献[5]和文献[7]的计算复杂度为 $O(W\times H\times T)$,可以看出,本文方法的计算复杂度明显小于对比方法,有效降低了计算时间。

从表 1 的平均准确率可以看出,本文方法平均准确率高于对比方法,原因是对于 CTT 向量为 $\boldsymbol{C}=$T 的监控视频,本文方法和文献[15]运动分割的结果类似,但对于 CTT 向量为 $\boldsymbol{C}=$F 的监控视频,文献[15]无法完整捕捉目标的运动轨迹,本文方法通过动态的调节采样线,构建自适应时空管道,可以适用于运动情况多变的监控场景。从平均召回率可以看出,本文方法略低于改进光流法,所提方法仅计算采样线上的像素点,监控视频场景过于复杂时可能遗漏小部分运动目标,但从表 1 可以看出,本文方法的平均 $F1$ 分数高于对比方法,说明综合考虑准确率和召回率时,本文方法精度更高,同时本文方法有效降低了计算时间,提高了计算速度,证明了本文方法具有良好的监控视频运动分割性能。

4.2　实验结果分析

为了验证本文方法的正确性和鲁棒性,选取具有代表性的监控视频 Video1、Video6 和 Video8 进行实验分析。由于文献[15] (2E)的方法与所提方法的平均计算时间最接近,进一步对比分析本文方法和文献[15] (2E)方法分割结果的精确度,图 5 展示了两种方法在三段视频上分割的运动片段结果,并与实验视频中实际存在的运动片段相比较,图中横轴代表时间轴,黑色块状图代表监控视频中运动片段的真实情况,条纹填充块状图和方格填充块状图分别代表使用本文算法和文献[15](2E)分割的运动片段。

在图 5(a)所示的 Video1 中,使用本文方法和嵌套椭圆时空管道方法提取的运动片段结果类似,第 31、571、1 132 和 1 506 帧为运动目标进入采样区域,第 434、967、1 296 和 1 793 帧为运动目标退出采样区域,两种方法都可以精准地分割运动片段。从图 5(b)中可以看出,对于不满足 CTT 的监控视频 Video6,文献[15] (2E)分割出的运动片段误差较大,如在第 839 和 994 帧分别为运动目标进入和退出采样

区域，但文献[15]（2E）无法捕捉目标退出采样区域的位置，产生了许多错误帧，如分割的 1 140 帧中不存在运动目标；使用本文方法对于此种情况可以完整捕捉到运动目标进入和退出采样区域，继尔进行运动分割。

从图 5(c)中可以看出，对于第 1 段运动片段，使用本文方法可以检测到运动目标并分割出运动片段，而文献[15]（2E）无法分割出运动片段，因为在视频开始时目标就已经存在于监控视频中，图中第 1 帧展示了其位置，文献[15]（2E）不能捕捉到该目标的完整运动轨迹，导致无法分割运动片段；对于分割出的第 2 和第 4 段运动片段，本文方法的分割结果更精确，如图中第 867、3 551 帧，目标进入采样区域后在门口处有短暂停留，文献[15]（2E）无法捕捉到此时的运动目标，目标继续运动后才被采样，但本文方法可以完整捕捉此类目标；使用本文方法和文献[15]（2E）分割出的第 5、6 段运动片段结果类似，因为第5、6段监控视频中的运动目标都是由采样区域

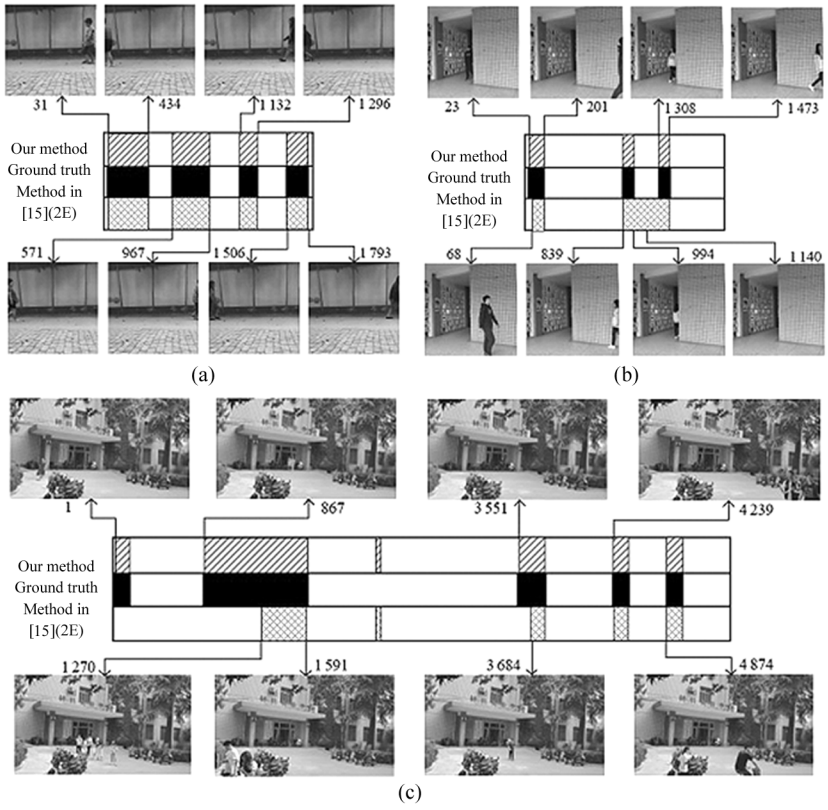

图 5　两种方法在 3 段视频上分割的运动片段和真实值的对比：(a) Video1；(b) Video6；(c)Video8

Fig. 5　Comparison of motion segments segmented by two methods and ground-truths on three videos: (a) Video1; (b) Video6; (c)Video8

的一侧进入，另一侧退出，满足 CTT 向量 $\boldsymbol{C}=\mathrm{T}$。

综上所述，本文方法在保证精度的同时进一步降低了计算量，通过自适应采样捕捉运动目标的采样方式提高了运动分割的鲁棒性，实现了运动片段的快速、精准分割。

5　结　论

本文针对运动分割问题提出了一种联合时空管道和二分思想的运动分割算法，该方法将 CTT 向量作为监控视频中目标运动情况的判定依据，并通过嵌套椭圆时空管道模型计算初始累计时空流量，结合二分思想动态地调节采样线，避免人工确定采样线的局限性，最后提取采样线上的像素构建自适应时空管道以分割运动片段。实验结果表明，所提方法在保证精确度的前提下，计算速度明显优于对比方法，从实验对比和实验结果分析可得，所提方法解决了现有的运动分割算法计算复杂度高的问题，且所提方法鲁棒性高，也适用于运动情况复杂多变的监控场景。

参考文献：

[1] PENG J L，ZHAO Y L，WANG L M. Research on video abnormal behavior detection based on deep learning[J]. Laser & Optoelectronics Progress，2021，58（6）：

0600004.
彭嘉丽,赵英亮,王黎明.基于深度学习的视频异常行为检测研究[J]. 激光与光电子学进展,2021,58(6):0600004.

[2] ZHANG Z,NIE Y,SUN H,et al. Multi-view video synopsis via simultaneous object-shifting and view-switching optimization[J]. IEEE Transactions on Image Processing,2020,29(1):971-985.

[3] LI T Y,BING B,WU X X. Boundary discrimination and proposal evaluation for temporal action proposal generation[J]. Multimedia Tools and Applications,2021,80(2):2123-2139.

[4] AN P,LIANG J X,MA J. LiDAR-camera-system-based 3D object detection with proposal selection and grid attention pooling[J]. Applied Optics,2022,61(11):2998-3007.

[5] MURTAZA F,YOUSAF M H,VELASTIN S A. PMHI:proposals from motion history images for temporal segmentation of long uncut videos[J]. IEEE Signal Processing Letters,2018,25(2):179-183.

[6] LU H C,HE H Z,HUANG Y Q,et al. Improved Canny edge operator and Gaussian mixture model for moving target detection[J]. Journal of Electronic Measurement and Instrumentation,2019,33(10):142-147.
陆华才,贺华展,黄宜庆,等. 改进 Canny 边缘算子和高斯混合模型的运动目标检测[J]. 电子测量与仪器学报,2019,33(10):142-147.

[7] GUO F,WANG W G,SHEN Z Y,et al. Motion-aware rapid video saliency detection[J]. IEEE Transactions on Circuits and Systems for Video Technology,2020,30(12):4887-4898.

[8] YANG J Q,HAN X H. Improved three-frame difference algorithm based on adaptive mixture gaussian[J]. Computer Engineering and Design,2021,42(6):1699-1705.
杨嘉琪,韩晓红. 基于自适应混合高斯的改进三帧差分算法[J]. 计算机工程与设计,2021,42(6):1699-1705.

[9] SUN P,LV L,QIN J. Moving object extraction based on saliency detection and adaptive background model[J]. Optoelectronics Letters,2020,16(1):59-64.

[10] PENG W,SHI J,ZHAO G. Spatial temporal graph deconvolutional network for skeleton-based human action recognition[J]. IEEE Signal Processing Letters,2021,28(1):244-248.

[11] DENG H,HUA Y,SONG T,et al. Object guided external memory network for video object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE,2019:6677-6686.

[12] CHEN K,WANG J,YANG S,et al. Optimizing video object detection via a dcale-time lattice[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 18-21, 2018, Salt Lake City, UT, USA. New York: IEEE,2018:7814-7823.

[13] SHEN Z,LIU Z,LI J,et al. Object detection from scratch with deep supervision[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(2):398-412.

[14] XU G Y,YIN M Y. Improved SSD object detection algorithm based on space-channel attention[J]. Journal of Optoelectronics · Laser,2021,32(9):970-978.
许光宇,尹孟园. 基于空间一通道注意力的改进 SSD 目标检测算法[J]. 光电子·激光,2021,32(9):970-978.

[15] ZHANG Y Z,GUO K N,CAI Z Q,et al. Nested elliptical spatio-temporal tubes for fast motion segment extraction in surveillance videos[J]. Acta Photonica Sinica,2022,51(3):0310005.
张云佐,郭凯娜,蔡昭权,等. 嵌套椭圆时空管道的监控视频运动片段快速提取[J]. 光子学报,2022,51(3):0310005.

[16] ZHANG Y Z,LI W X,YANG P L. Surveillance video motion segments segmentation based on spatio-temporal flow[J]. Laser & Optoelectronics Progress,2022,59(8):0810012.
张云佐,李汶轩,杨攀亮. 基于时空流量的监控视频运动片段分割[J]. 激光与光电子学进展,2022,59(8):0810012.

[17] ZHUANG X T. Research on deep learning networks for small object detection based on multi-level feature fusion [D]. Nanjing: Nanjing University of Posts and Telecommunications,2021:1-13.
庄幸涛. 基于多级特征融合的小目标深度检测网络研究[D]. 南京:南京邮电大学,2021:1-13.

作者简介:

张云佐 (1984—),男,博士,副教授,博士生导师,主要从事智能视频分析、图像处理、大数据方面的研究.