

引用格式: ZHANG Yunzuo, GUO Kaina, CAI Zhaoquan, 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

# 嵌套椭圆时空管道的监控视频运动片段快速提取

张云佐<sup>1,2</sup>, 郭凯娜<sup>1</sup>, 蔡昭权<sup>3</sup>, 李文轩<sup>1</sup>

(1 石家庄铁道大学 信息科学与技术学院, 石家庄 050043)

(2 石家庄铁道大学 河北省电磁环境效应与信息处理重点实验室, 石家庄 050043)

(3 汕尾职业技术学院, 广东 汕尾 516600)

**摘要:**针对现存运动片段提取方法大多步骤繁琐、计算量大,无法满足快速提取运动片段的现实需求的问题,提出一种嵌套椭圆时空管道的监控视频运动片段快速提取算法,以最大限度地减小计算量。该算法首先对监控视频进行椭圆时空采样形成时空管道,其次递进椭圆时空采样组成嵌套椭圆时空管道,然后展开嵌套椭圆时空管道生成时空平面图,最后对时空平面图进行背景去除,构建时空流量模型以提取运动片段。在 SISO 等公开数据集上与当前主流的运动片段提取方法进行了实验对比,结果表明,本文算法在计算时间上具有明显优势,可以实现监控视频运动片段的全面、快速提取。

**关键词:**图像处理;运动片段提取;椭圆时空采样;嵌套椭圆时空管道;时空平面图

中图分类号: TP391

文献标识码: A

doi:10.3788/gzxb20225103.0310005

## 0 引言

随着人们安全意识的提升,视频监控系统得到了广泛的应用,由此产生的监控视频数据量日益剧增。监控摄像机通常固定安装且长时间持续拍摄,导致监控视频背景单一、运动片段和静止片段交叉存在,而人们通常只关注其中的运动片段。如何从海量的监控视频中快速提取运动片段有着迫切的现实需求,这也引起了研究人员的广泛关注<sup>[1-3]</sup>,并涌现出了大量的研究成果。

现有的运动片段提取方法多是基于运动目标检测取得的,在传统的运动目标检测方法中,文献[4]融入了改进的关键帧提取方法,以单个运动对象为目标进行视频空间上的压缩;文献[5]对背景图像和使用滤波处理后的连续帧间差分图像进行差分运算,在该方法中运动速度较慢的物体也能得到较好的检测;文献[6]通过计算历史运动帧的能量来反映视频中运动目标的时空信息,实现运动片段提取;文献[7]将改进的局部纹理特征二值模式和颜色特征相结合用于背景建模,并结合像素的邻域相关性抑制虚假前景,可以很好地应对多模态背景场景;文献[8]将背景差分法和改进顺序轮廓分离策略相结合,在背景差分法检测前景运动过程中,将改进的顺序轮廓分离策略用于运动目标发现,可以区分复杂情况下大小和数量不断变化的运动目标;文献[9]在传统光流法的基础上计算了光流场四个边界的平均运动向量来提取前景目标,检测效果有所提升但运算量明显增大。基于上述目标检测方法进行运动片段提取可以取得较为理想的效果,但由于这些方法需要处理视频空域全量数据,逐帧检测运动目标,计算量大、耗费时间长。

随着人工智能的快速发展,深度学习方法被广泛应用于目标检测领域,文献[10]设计了一种端到端的用于间歇性目标检测的深度卷积神经网络,该网络以视频中的相邻两帧作为输入,根据像素之间的对应关

**基金项目:**广东省重点领域研发计划(No.2019B010137006),国家自然科学基金(Nos. 61702347, 62027801, 61972267),河北省省自然科学基金(No. F2017210161),河北省教育厅科学研究项目(No.ZD2016052)

**第一作者(通讯作者):**张云佐(1984—),男,副教授,博士,主要研究方向为图像处理、视频智能分析、大数据处理。Email: zhangyunzuo888@sina.com

**收稿日期:**2021-09-27; **录用日期:**2021-11-02

<http://www.photon.ac.cn>

系检测出运动目标;文献[11]提出了金字塔注意力模块来获取关键目标信息,并将扩张卷积块专用于多尺度特征提供语义信息和几何细节,来获得运动前景。此类算法需要大量的样本数据预先训练模型,算法复杂度且对硬件设备有较高要求。

随着移动设备的广泛应用,在线运动片段提取的需求日益迫切,现有方法通常运算量大、算法复杂度高,难以应用到计算资源受限的移动设备上。文献[12]提出了边界检测方法降低运算量,很好地解决了跨越监视区域边界的运动目标检测问题,但对监视区域内部的目标检测效果并不理想。为此,本文提出了一种嵌套椭圆时空管道的监控视频运动片段快速提取算法,该算法采用椭圆时空采样来兼顾运动目标检测的准确率和待处理的视频数据量;采用时空平面图全面分析监视区域内的运动状况;实现全面、快速地提取监控视频运动片段。

## 1 本文算法

本文提出了一种嵌套椭圆时空管道的监控视频运动片段快速提取算法,首先对输入视频序列进行椭圆时空采样以检测跨越视频边界的运动目标;其次递进椭圆时空采样生成嵌套椭圆时空管道;然后展开嵌套椭圆时空管道形成时空平面图;最后基于时空平面图构建时空流量模型进而提取监控视频运动片段。本文所提算法的流程如图1。

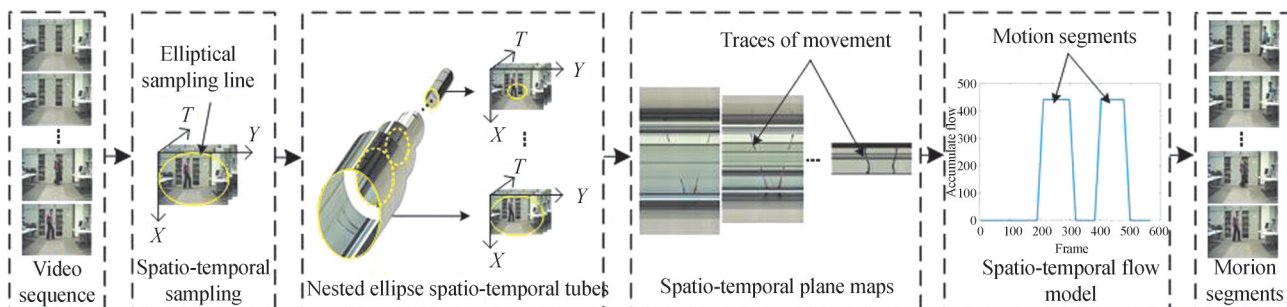


图1 本文算法流程

Fig. 1 Flowchart of proposed algorithm

### 1.1 椭圆时空采样

时空切片是一种能够高效表达视频序列时空信息的方式,被广泛应用于视频分析领域<sup>[13]</sup>。根据采样线的选择方式不同,时空切片具有不同的表现形式。在现实监控场景下,目标在监视区域内通常沿直线运动,如果运动目标在水平或垂直时空切片中一直以平行于采样线的方向运动,目标将不会跨越采样线,此时就会造成目标的丢失,导致漏检。边界时空采样<sup>[12]</sup>可以检测到来自各个方向的运动目标,但不能检测到在监视区域内部运动的目标。

本文提出椭圆时空采样,如图1时空采样模块所示,根据不同宽高比的视频序列,自适应地产生内切于视频帧的椭圆采样线,当运动目标进入或退出监视区域时,会跨越椭圆采样线,通过计算图中椭圆采样线的周长可以得出,其周长明显小于边界采样线的周长,因此该采样方式可以兼顾来自各个方向的运动目标并且需要处理的数据量更小。对于现实监控场景来说,运动目标的轨迹多抽象为直线、拐角及其组合,因此不会导致漏检。提取视频序列每一帧的采样像素,按照时间轴 $T$ 依次向后排,形成椭圆时空管道,有目标跨越椭圆采样线时,会在椭圆时空管道上留下运动痕迹。

### 1.2 嵌套椭圆时空管道

为了全面检测监视区域内的运动目标,采用递进椭圆时空采样的方式来划分监视区域,图2(a)给出了3条椭圆采样线的划分示例,对应形成了3个子监视区域。同理,采用 $n$ 条椭圆采样线时,视频将被划分为 $n$ 个子监视区域,椭圆采样线条数越多,运动目标的漏检率越低。

不同的子监视区域对应不同周长的椭圆时空管道,将各椭圆时空管道按时间轴 $T$ 排列,周长越短的排列越靠后,形成如图2(b)所示的嵌套椭圆时空管道。

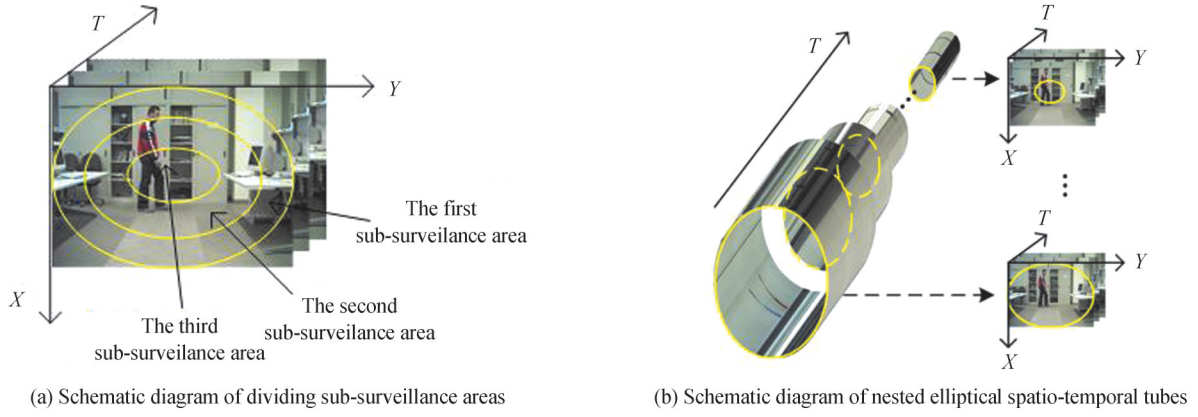


图2 嵌套椭圆时空管道的形成

Fig. 2 The formation of nested elliptical spatio-temporal tubes

### 1.3 时空平面图

沿嵌套椭圆时空管道的时间轴  $T$  方向将其展开可得到时空平面图。对于一个宽高为  $M \times N$  共有  $T$  帧的监控视频,由外向内椭圆采样线周长分别为  $L_1, L_2, \dots, L_n$ ,计算周长为  $L$  的椭圆采样线上某一像素在视频帧中的坐标  $(x, y)$ ,首先定义  $\theta$  为

$$\theta = \frac{2\pi S}{L} \quad (S = 1, 2, \dots, L) \quad (1)$$

式中,  $S$  是弧长,  $L$  是对应椭圆采样线的周长,则  $\theta$  是椭圆采样线上坐标点与椭圆中心的连线和  $x$  轴正方向的夹角。

当  $0 \leq \theta < \frac{\pi}{2}$  或  $\frac{3\pi}{2} < \theta \leq 2\pi$  时,有

$$\begin{cases} x = \frac{ab}{\sqrt{b^2 + a^2 \tan^2 \theta}} + b \\ y = \frac{ab \tan \theta}{\sqrt{b^2 + a^2 \tan^2 \theta}} + a \end{cases} \quad (2)$$

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

$$\begin{cases} x = -\frac{ab}{\sqrt{b^2 + a^2 \tan^2 \theta}} + b \\ y = -\frac{ab \tan \theta}{\sqrt{b^2 + a^2 \tan^2 \theta}} + a \end{cases} \quad (3)$$

当  $\theta = \frac{\pi}{2}$  时,有

$$\begin{cases} x = b \\ y = a + b \end{cases} \quad (4)$$

当  $\theta = \frac{3\pi}{2}$  时,有

$$\begin{cases} x = b \\ y = a - b \end{cases} \quad (5)$$

式中,  $a$  是椭圆采样线长轴的长度,  $b$  是椭圆采样线短轴的长度,  $a, b$  分别为监控视频帧宽度的  $1/2$  和高度的  $1/2$ 。将第  $k$  帧视频中椭圆采样线上各个像素的坐标自上向下排列,可以形成长度为  $L$  的列向量  $E^k$ ,可表示为

$$E^k = [p_{1,1}^k, p_{2,1}^k, p_{3,1}^k, \dots, p_{L,1}^k]^T \quad (6)$$

对视频序列中的每一帧进行相同的椭圆采样,提取每个椭圆上的像素,用  $p_{i,j}^k$  表示第  $k$  帧上第  $i$  行第  $j$  列

的像素,得到每一帧的 $E^k$ ,将这些向量按时间顺序从左到右依次排列,可以得到椭圆时空管道的平面图 $S_E$ ,可表示为

$$S_E = \begin{bmatrix} p_{i_1, j_1}^1 & p_{i_1, j_1}^2 & \cdots & p_{i_1, j_1}^l & \cdots & p_{i_1, j_1}^T \\ p_{i_2, j_2}^1 & p_{i_2, j_2}^2 & \cdots & p_{i_2, j_2}^l & \cdots & p_{i_2, j_2}^T \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{i_r, j_r}^1 & p_{i_r, j_r}^2 & \cdots & p_{i_r, j_r}^l & \cdots & p_{i_r, j_r}^T \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{i_l, j_l}^1 & p_{i_l, j_l}^2 & \cdots & p_{i_l, j_l}^l & \cdots & p_{i_l, j_l}^T \end{bmatrix} \quad (7)$$

将嵌套椭圆时空管道沿时间轴 $T$ 方向展开,可以得到如图3所示的嵌套椭圆时空管道的时空平面图,从图中可以看到运动目标跨越椭圆采样线时在时空平面图上留下的痕迹。

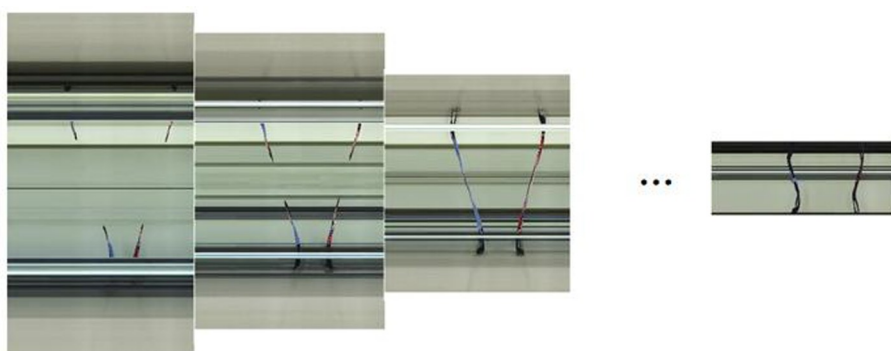


图3 嵌套椭圆时空管道的时空平面图

Fig. 3 Spatio-temporal plane maps of nested elliptical spatio-temporal tubes

#### 1.4 时空流量模型

运动片段由目标进入监视区域开始,到目标退出监视区域截止,目标进入监视区域赋予正流量,退出监视区域赋予负流量。为了确定子监视区域中运动目标的进入和退出,引入了次采样线,如图4所示,若运动目标先跨越子监视区域的采样线后跨越次采样线,则为进入子监视区域;若运动目标先跨越次采样线后跨越采样线,则为退出子监视区域。

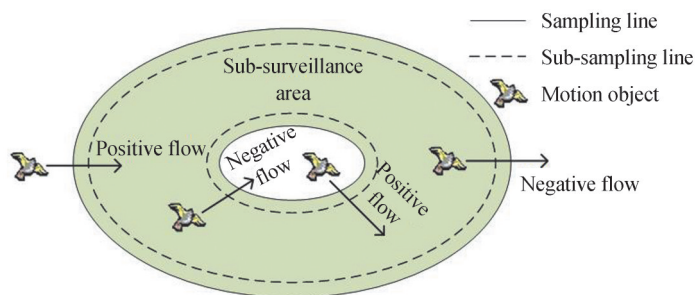


图4 次采样线及时空流量示意图

Fig. 4 Schematic diagram of sub-sampling lines and spatio-temporal flow

将时空平面图中运动目标的质心归一化,如果运动目标进入子监视区域,将运动目标像素的值赋为1,如果运动目标退出子监视区域,将运动目标像素的值赋为-1。则第 $n$ 个子监视区域的椭圆时空切片展开图中第 $k$ 帧的瞬时时空流量 $F_n(f_k)$ 为

$$F_n(f_k) = \sum_{i=1}^L p_{i,k} \quad (8)$$



$$p_{i,k} = \begin{cases} 1 & \text{entry} \\ -1 & \text{exit} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

式中, $p_{i,k}$ 表示展开图中第*i*行第*k*列的像素的值,即第*k*帧的像素, $L$ 表示对应椭圆采样线的周长。设*N*为子监视区域的总数目,那么第*n*个子监视区域的瞬时时空流量定义为

$$F^n(f_k) = \begin{cases} F_n(f_k) + F_{n+1}(f_k) & n < N \\ F_n(f_k) & n = N \end{cases} \quad (10)$$

第*n*个子监视区域的累计时空流量定义为

$$AF^n(f_k) = \sum_{i=1}^k F^n(f_i) \quad (11)$$

根据式(8)~(11)可计算出椭圆时空管道的瞬时时空流量和累计时空流量,构建时空流量模型。在该模型中,瞬时时空流量大小反映了相应帧中运动目标是进入还是退出子监视区域,累计时空流量大小反映了相应帧子监视区域中运动目标的数量。

获得每个子监视区域的累计时空流量 $AF^i$ 后,依据监控视频中运动目标的稠密程度,将监控视频划分为*N*个子监视区域,将每个子监视区域的累计时空流量取并集,可以得到整个监视区域的累计时空流量,可表示为

$$AF = AF^1 \cup AF^2 \cup \dots \cup AF^i \cup \dots \cup AF^N \quad (12)$$

根据式(12)获得的累计时空流量可以提取出整个监控视频的运动片段。

## 2 实验设置与分析

### 2.1 实验设置

实验环境为 Windows10 系统、Intel(R) Core(TM) i5-8265U CPU、NVIDIA GeForce MX250 显卡,内存 16 G。为了验证本文算法的正确性和有效性,在 20 段监控视频上进行了实验,实验视频涉及多种场景,Video1-Video13 来自公共数据集 SISR, Video14、Video15 来自公共数据集 KTH, Video16-Video18 来自公共数据集 CAVIAR, Video19、Video20 为真实场景采集的监控视频,视频的详细信息如表 1。

表 1 实验视频信息  
Table 1 Information of experimental videos

Video sequence	Frame rate	Resolution ratio	Frame number
Video1	25	384×288	566
Video2	25	720×576	543
Video3	25	320×240	944
Video4	10	384×288	3 795
Video5	25	360×288	369
Viede6	25	360×288	336
Video7	25	360×288	289
Video8	25	320×256	137
Video9	10	384×288	1 591
Video10	20	352×288	2 308
Video11	25	352×288	4 561
Video12	10	384×288	2 567
Video13	25	352×288	1 013
Video14	25	720×480	892
Video15	25	160×120	741
Video16	25	384×288	293
Video17	25	384×288	381
Video18	25	384×288	2 358
Video19	30	720×480	1 332
Video20	30	720×480	744

## 2.2 实验对比

为了分析本文算法的有效性,使用计算时间(Calculating time,  $T_c$ )、准确率(Precision,  $P$ )、召回率(Recall,  $R$ )和  $F_1$ -score( $F_1$ )来衡量算法的性能,统计实验结果中所有被识别为运动片段的视频帧,若帧中包含运动目标,则该帧为真正例( $N_{TP}$ ),若帧中不包含运动目标,则该帧为假正例( $N_{FP}$ ),对于包含运动目标但没有被识别为运动片段的帧,称为假负例( $N_{FN}$ )。准确率、召回率和  $F_1$ -score 的计算表达式分别为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (13)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (14)$$

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (15)$$

分别选取1条椭圆采样线(1E)、2条椭圆采样线(2E)、3条椭圆采样线(3E)和4条椭圆采样线(4E)对表1中的20段视频进行实验分析,并与当前主流的文献[4]、文献[6]和文献[9]中的运动片段提取算法进行实验对比,表2为对比算法和本文算法在20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  $F_1$  的值。

表2 实验结果对比表  
Table 2 Table for comparing experimental results

Algorithm	Average calculating time/s	Average precision /%	Average recall/%	Average $F_1$ /%
Pixel difference <sup>[4]</sup>	614.09	92.64	82.11	87.06
PMHI <sup>[6]</sup>	1 167.82	85.39	87.46	86.41
Improved optical flow <sup>[9]</sup>	1 184.88	91.70	94.37	93.02
Proposed method(1E)	33.82	99.73	85.82	92.10
Proposed method (2E)	81.41	97.26	91.48	94.28
Proposed method (3E)	145.42	96.45	93.82	95.12
Proposed method (4E)	199.53	96.01	94.75	95.38

从表2的平均计算时间可以看出,本文算法的计算时间明显低于文献[4]、文献[6]和文献[9]中算法的计算时间,主要原因是本文算法仅对时空平面图进行处理,即每帧只处理椭圆采样线上的像素,而不需逐帧逐像素处理,大大减少了数据量。假设视频的大小为  $M \times N \times T$ ,  $M$  表示视频帧的高,  $N$  表示视频帧的宽,  $T$  表示视频序列的长度,则本文算法的计算复杂度为  $O(2 \times N + (\pi - 2) \times M)$ , 而对比算法的计算复杂度为  $O(M \times N \times T)$ ,  $O(2 \times N + (\pi - 2) \times M) \ll O(M \times N \times T)$ , 可以看出本文算法的计算复杂度远小于对比算法,因此大大降低了计算时间。

从表2的平均准确率可以看出,本文算法的平均准确率高于文献[4]、文献[6]和文献[9]中算法,随着椭圆采样线数量的增多,平均准确率会略微降低,这是因为椭圆采样线的增多可能会将小目标误认为运动目标,降低准确率,但本文算法的平均准确率依然高于对比算法。从平均召回率可以看出,当椭圆采样线的数量小于4时召回率略低于改进光流法,这是因为所提算法仅计算椭圆采样线上的像素点,当椭圆采样线数量过少时可能遗漏部分运动目标,但随着椭圆采样线数量的增多,平均召回率逐渐提高,采用4条椭圆采样线时,本文算法的平均召回率已经高于对比文献的算法。从平均  $F_1$  可以看出,采样线数量增多时,  $F_1$  逐渐提高,说明本文算法整体的性能相较于对比算法有所提升,证明了本文算法具有良好的监控视频运动片段提取性能。

## 2.3 实验结果分析

为了验证本文算法的正确性,选用具有代表性的视频 Video1 和 Video19 进行分析,其中 Video1 为行人在室内运动的场景,两个行人分别进入监视区域并退出, Video19 为多个行人在室外运动的场景。

Video1 的瞬时时空流量、累计时空流量如图5。当瞬时时空流量大于0时,应为目标进入监视区域,相应的累计时空流量应逐渐上升;当瞬时时空流量小于0时,应为目标退出监视区域,相应的累计时空流量应

逐渐下降,Video1的时空流量曲线跳变处的视频帧如图6。

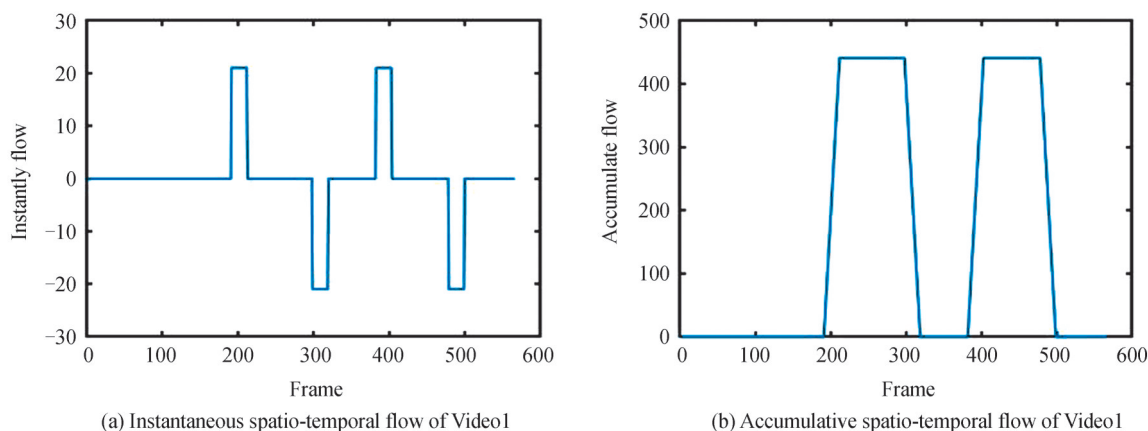


图5 Video1的时空流量  
Fig. 5 Spatio-temporal flow of video1



图6 Video1的时空流量曲线跳变处的部分视频帧  
Fig. 6 The part of the video frames at spatio-temporal flow curve changed of video1

Video1的第181、385帧为运动目标进入监视区域,对应的瞬时时空流量大于0,累计时空流量呈上升趋势;第321、505帧为运动目标退出监视区域,对应的瞬时时空流量小于0,累计时空流量呈下降趋势。通过图6可以看出,181~321帧为运动目标从监视区域右侧进入,从左侧退出,是一段完整的运动片段,对应的累计时空流量大于0;385~505帧为运动目标从监视区域左侧进入,从右侧退出,是一段完整的运动片段,对应的累计时空流量大于0;其余帧累计时空流量等于0,为静止片段,印证了本文算法的正确性。

对Video19使用2条椭圆采样线进行实验分析,使用本文算法,2条椭圆采样线可以将视频序列分割为两个子监视区域,以第一个子监视区域为例进行分析,Video19中目标的运动状态、对应的瞬时时空流量和累计时空流量如图7所示。

从图7中可以看出,第75、341、852、931帧为运动目标跨越椭圆采样线进入子监视区域,子监视区域相应的瞬时时空流量大于0,累计时空流量呈上升趋势;第172、761、1235帧为运动目标跨越椭圆采样线退出子监视区域,对应的瞬时时空流量小于0,累计时空流量呈下降趋势,印证了本文算法的正确性。

当累计时空流量大于0时,表明进入子监视区域的运动目标数量大于退出子监视区域的运动目标数量,该部分视频为运动片段;当累计时空流量等于0时,表明子监视区域没有运动目标进出,该部分视频为静止片段。

特别地,当累计时空流量小于0时,表示退出子监视区域的运动目标数量大于进入子监视区域的运动目标数量,说明有目标由静止转为运动。

综上所述,本文算法在保证精确度的同时,有效降低了计算复杂度,实现了监控视频运动片段的快速提取。实际应用中,子监视区域划分的越多,运动片段提取的精确度也就越高,但计算量也会随之增大。

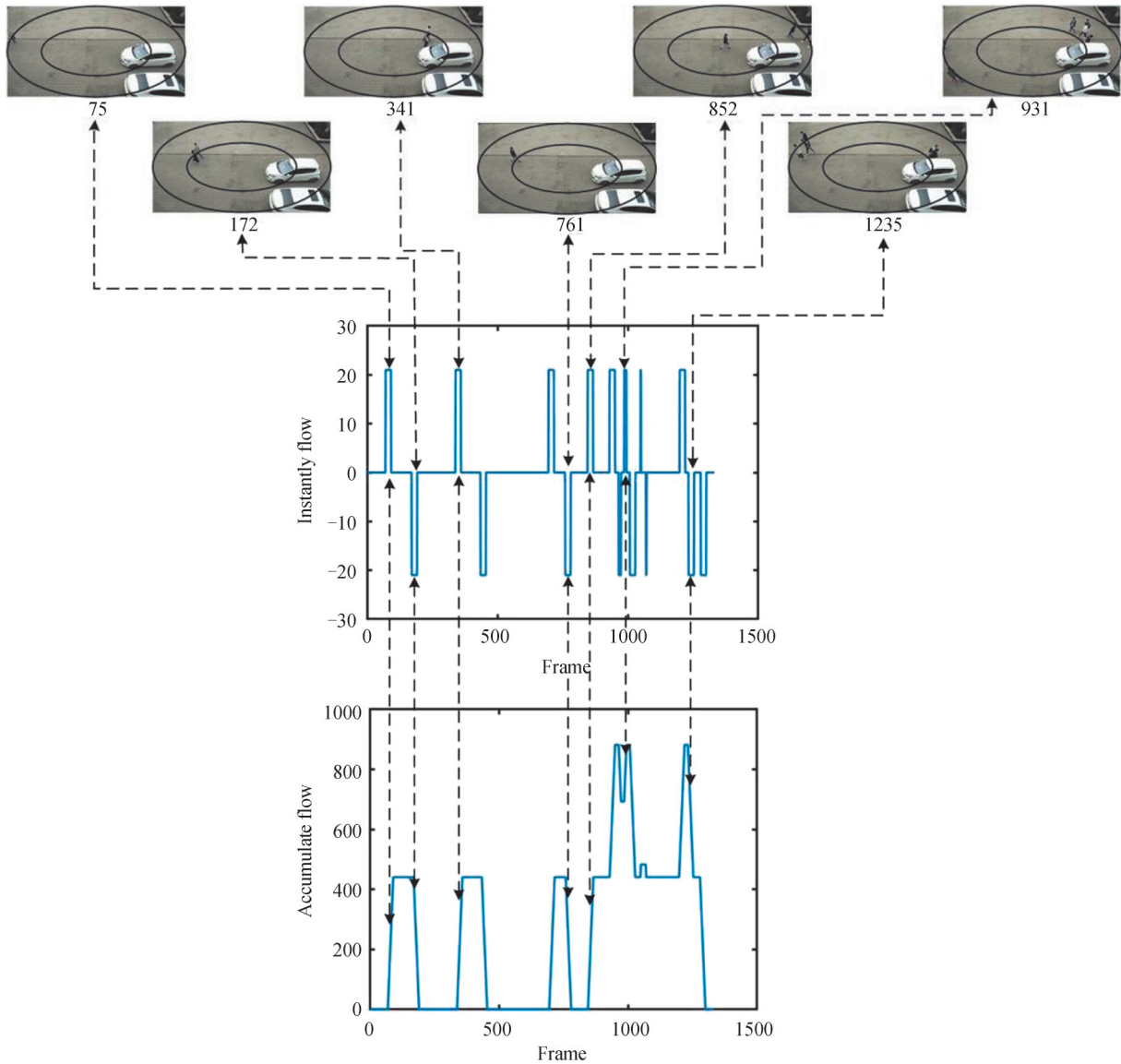


图7 Video19的时空流量分析图

Fig. 7 Spatio-temporal flow analysis chart of video19

### 3 结论

本文提出了一种嵌套椭圆时空管道的运动片段快速提取算法,利用递进的椭圆采样线对监控视频序列进行采样,将监控视频划分为多个子监视区域并检测运动目标,通过构建时空流量模型提取运动片段,在保证检测精度的前提下极大地减小了计算量,可以实现监控视频运动片段的快速提取。下一步工作,将研究运动片段中运动模式分类与目标精准识别。

#### 参考文献

- [1] 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: 971-985.
- [2] WANG Yibin, ZHENG Jia, LV Zhuowen, et al. Traffic monitoring image dehazing algorithm based on wavelength-related physical imaging model[J]. Acta Photonica Sinica, 2019, 48(9): 0910004.  
王一斌, 郑佳, 吕卓纹, 等. 波长相关物理成像模型的交通监控图像去雾算法[J]. 光子学报, 2019, 48(9): 0910004.
- [3] NIE Y, LI Z, ZHANG Z, et al. Collision-free video synopsis incorporating object speed and size changes[J]. IEEE Transactions on Image Processing, 2020, 29: 1465-1478.
- [4] WANG Yapei, LI Renwang, LIU Xiang. Extraction method of surveillance video synopsis combines objects and keyframes[J]. Industrial Control Computer, 2015, 28(3): 11-13.



- 王亚沛,李仁旺,刘翔.对象和关键帧相结合的监控视频摘要提取方法[J].工业控制计算机,2015,28(3):11-13.
- [5] QU Jingjing, XIN Yunhong. Combined continuous frame difference with background difference method for moving object detection[J]. Acta Photonica Sinica, 43(7): 219-226.  
屈晶晶,辛云宏.连续帧间差分与背景差分相融合的运动目标检测方法[J].光子学报,2014,43(7):219-226.
- [6] 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.
- [7] JIN Jing, DANG Jianwu, ZHAI Fengwen, et al. Moving target detection based on KDE combining local texture feature [J]. Journal of Jilin University(Engineering and Technology Edition), 2019, 49(2): 647-655.  
金静,党建武,翟凤文,等.融合局部纹理特征的核密度估计运动目标检测[J].吉林大学学报(工学版),2019,49(2):647-655.
- [8] THENMOZHI T, KALPANA A M. Adaptive motion estimation and sequential outline separation based moving object detection in video surveillance system[J]. Microprocessors and Microsystems, 2020, 76: 103084.
- [9] 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.
- [10] LIU Junming, MENG Weihua. Infrared small target detection based on fully convolutional neural network and visual saliency[J]. Acta Photonica Sinica, 2020, 49(7): 0710003.  
刘俊明,孟卫华.融合全卷积神经网络和视觉显著性的红外小目标检测[J].光子学报,2020,49(7):0710003.
- [11] QU S, ZHANG H, WU W, et al. Symmetric pyramid attention convolutional neural network for moving object detection [J]. Signal, Image and Video Processing, 2021, 15(8): 1747-1755.
- [12] ZHANG Yunzuo, LI Wenxuan, YANG Panliang. 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.
- [13] SOUZA M R E, MAIA H D, VIEIRA M B, et al. Survey on visual rhythms: a spatio-temporal representation for video sequences[J]. Neurocomputing, 2020, 402: 409-422.

## Nested Elliptical Spatio-temporal Tubes for Fast Motion Segment Extraction in Surveillance Videos

ZHANG Yunzuo<sup>1,2</sup>, GUO Kaina<sup>1</sup>, CAI Zhaoquan<sup>3</sup>, LI Wenxuan<sup>1</sup>

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

(2 Hebei Key Laboratory of Electromagnetic Environmental Effects and Information Processing, Shijiazhuang Tiedao University, Shijiazhuang 050043, China)

(3 Shanwei Institute of Technology, Shanwei, Guangdong 516600, China)

**Abstract:** With the development of people's security awareness, video surveillance systems have been widely used, resulting in an increasing amount of surveillance video data. Surveillance videos are usually fixed and continuous for a long time, which leads to a single surveillance video background, motion segments and static segments cross-exist. However, people usually only focus on the motion segments. There is an urgent need to quickly extract motion segments from massive surveillance videos, which has received a lot of attention from researchers. Motion segment extraction is the basis and prerequisite of behavior recognition, surveillance video synopsis and other subsequent processing, and it is also a research hotspot in the field of computer vision. The existing motion segments extraction algorithms are mainly divided into traditional methods and deep learning-based methods. The former needs to process the full amount of data in the spatial domain of the video and detect motion targets frame by frame, which is computationally intensive and time-consuming and can't meet the real-time needs. The latter requires a massive amount of sample data to pre-train the model, which has high algorithm complexity and high requirement for hardware devices. Addressing this problem, this paper proposed a fast motion segment extraction method via nested ellipse spatio-temporal tubes in surveillance video, which can extremely save the amount of calculation. Firstly, surveillance video is elliptically spatio-temporal sampled. The elliptical sampling lines are adaptively generated according to the video sequences with different aspect ratios and pixel on the sampling lines of each frame in the video sequence are extracted to form an elliptical spatio-temporal tube. Secondly, multiple elliptical spatio-temporal tubes sampled progressively according to

surveillance scene are integrated to nested elliptical spatio-temporal tubes. Then, nested elliptical spatio-temporal tubes are expanded to generate spatio-temporal plane maps. Finally, the background of spatio-temporal plane maps is removed and the spatio-temporal flow model is constructed to extract motion segments. In this model, the instantaneous spatio-temporal flow curve reflects whether moving targets enter or exit the sub-surveillance area in the corresponding frame, and the accumulative spatio-temporal flow curve reflects the number of moving targets in the sub-surveillance area in the corresponding frame. Flowchart of proposed algorithm as shown in the figure. The proposed algorithm utilizes elliptical spatio-temporal sampling instead of traditional full spatial data processing and requires no pre-training, which can greatly reduce the amount of surveillance video data to be processed. The algorithm form nested elliptical spatio-temporal tubes by progressive sampling, which reduces the computational effectively and can also take into account the targets moving inside the surveillance areas. The experimental on SISOR, KTH and CAVIAR public data sets comparison with mainstream motion segments extraction algorithms. Experimental results show that the proposed algorithm has obvious advantage in calculating time, greatly reduces the amount of calculation under the premise of ensuring detection accuracy, and it can realize fast motion segments extraction in surveillance videos.

**Key words:** Image processing; Motion segment extraction; Elliptical spatio-temporal sampling; Nested elliptical spatio-temporal tubes; Spatio-temporal plane maps

**OCIS Codes:** 100.2000; 100.3008; 150.1135