DOI:10.16136/j.joel.2022.09.0890

顾及方向信息的时空联合监控视频摘要方法

张云佐[1*]，郭亚宁[1]，蔡昭权[2]，张嘉煜[1]

(1.石家庄铁道大学 信息科学与技术学院，河北 石家庄 050043；2.汕尾职业技术学院 广东 汕尾 516600)

摘要：视频摘要是快速获取视频关键信息的有效途径，现有的视频摘要方法通常计算复杂度高，在计算资源受限的场景下难以实际应用。为此，提出了一种高效的顾及方向信息的时空联合监控视频摘要方法。该方法首先采用水平切片获得目标时空运动轨迹；其次去除时空轨迹背景并计算直线轨迹斜率，依据目标时空轨迹斜率完成目标运动方向判定；然后检测采样域运动片段以确定目标在视频中的时序位置；最后依据目标时序位置及其运动方向自适应构建视频摘要。实验结果表明，所提方法的帧平均处理时间(average frame processing time，*AFPT*)达到了0.374 s，明显优于对比方法，且所生成的视频摘要简洁高效、用户体验好。

关键词：算法；时空联合域；时空切片；自适应；目标运动方向；视频摘要

中图分类号：TP391 **文献标识码**：A **文章编号**：1005-0086(2022)09-0992-09

A spatiotemporal joint method for surveillance video summarization via the direction information

ZHANG Yunzuo[1*], GUO Yaning[1], CAI Zhaoquan[2], ZHANG Jiayu[1]

(1. School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, Hebei 050043, China; 2. Shanwei Institute of Technology, Shanwei, Guangdong 516600, China)

Abstract: Video summarization is an effective way to quickly obtain the key information of video. Existing video summarization methods usually have high computational complexity and are difficult to be applied in the scene with limited computing resources. Therefore, a spatiotemporal joint method for surveillance video summarization via the direction information is proposed. This method first uses the horizontal slice to obtain the object spatiotemporal motion trajectory. Secondly, the spatiotemporal trajectory background is eliminated and the linear trajectory slope is calculated, and the object motion direction is determined according to the object spatiotemporal trajectory slope. Thirdly, the motion segment in the sampling domain is detected to determine the timing position of the object in the video. Finally, the video summarization is constructed adaptively according to the object timing position and motion direction. Experimental results show that the average frame processing time (*AFPT*) of the proposed method reaches 0.374 s, which has obvious advantages over those of the comparison methods. In addition, the generated video summarization is concise and efficient, and the user experience is good.

Key words: algorithms; spatiotemporal joint domain; spatiotemporal slice; self-adaption; object motion direction; video summarization

1 引 言

随着监控系统的普及与广泛应用，监控视频数据量呈现爆炸式增长，如何从海量监控视频中快速获取有效信息是当前亟待解决的难题，有着迫切的现实需求[1,2]。视频摘要[3,4]能以简洁的形式呈现视频重要内容，是实现冗长视频快速浏览、检索的有效解决方案。

* **E-mail**：zhangyunzuo888@sina.com

收稿日期：2021-12-30 **修订日期**：2021-01-27

基金项目：国家自然科学基金(61702347，62027801)、广东省重点领域研发计划项目(2019B010137002)、河北省自然科学基金(F2022210007，F2017210161)、河北省高等学校科学技术研究项目(ZD2022100，QN2017132)和中央引导地方科技发展资金项目(226Z0501G)资助项目

传统的视频摘要技术大多基于聚类、图模型、稀疏编码等无监督学习方法,以外观、运动特征等底层视觉信息为依据提取视频摘要。无监督方法长期以来一直作为视频摘要领域的主导方法,该方法无需大量数据作为训练基础,其简洁性、有效性促使其进一步地发展,现如今仍有大批研究者在此基础上提出了许多新方法[5-10]。葛钊等[6]将聚类和图模型相结合并优化,提出了一种最短路径的视频摘要方法,该方法将视频摘要问题转化为最短路径求解问题;冀中等[7]通过构建视频超图模型,提出了基于超图随机游走的视频摘要算法,并通过求取关键帧之间的曼哈顿距离进行了去冗余操作以避免重复关键帧的出现;ANURADHA 等[8]提出了一种高压缩率的视频摘要方法,该方法通过对初始帧的前景信息不断更新进而获得具有高压缩率的视频摘要;BAGHERI 等[9]提出了将监控视频映射到时空剖面的视频摘要方法,该方法提供了紧凑图像中运动目标的直观摘要;THOMAS 等[10]提出了感知视觉摘要,该方法首次在摘要框架中引入人类视觉系统的特征,以允许强调具有感知意义的事件。现有的无监督视频摘要方法大多以视频帧为单位完成处理过程,对于较长的视频序列而言,此类方法的计算复杂度很高,在对时间要求较高的场景下较难实时应用。

随着深度学习的兴起,许多网络模型被用于视频摘要任务中[11-15],最具代表性的是 ZHANG 等[13]将长短期记忆网络(long short-term memory,LSTM)应用到视频摘要任务,并提出了 VSLSTM(video summary long-short term memory)模型,该方法通过双向长短期记忆网络(bi-directional long-short term memory,BiLSTM)预测视频帧的重要性得分,同时提出了 DPPLSTM(determinantal 码器-解码器框架将输入序列编码为固定长度的中间向量,然后将其解码为满足任务要求的输出序列,大批研究者基于此模型提出了改进[14,15]。例如,JI 等[15]将注意力机制与编码解码器相结合,其中编码部分由 BiLSTM 网络构成,解码器部分通过基于注意力机制的 LSTM 网络完成;武光利等[16]提出了一种基于改进的 BiLSTM 的视频摘要生成方法,该方法将卷积神经网络(convolutional neural networks,CNN)作为编码器以提取视频帧的深度特征,以 BiLSTM 作为解码器获得时序特征,并通过最大池化以完成特征优化,改进后的视频摘要生成模型提升了生成视频摘要的准确性。基于深度学习的视频摘要方法已被证明是有效的,但仍然存在许多缺陷,一方面,编码器-解码器框架的计算复杂度非常高,若使用 BiLSTM 更是加剧了计算复杂度,另一方面,基于深度学习的模型在较长的视频序列上的计算复杂度更高。

为了解决视频摘要生成过程中计算复杂度高的问题,本文以时空切片为单位完成视频摘要处理,考虑到时空切片存在目标运动方向不能确定的问题,提出了顾及方向信息的时空联合监控视频摘要方法,该方法以水平切片目标轨迹(target trajectory on horizontal slice,TTHS)斜率作为运动方向的判定标准,通过检测采样域内运动片段序列帧以确定目标在视频中的时序位置,最后将监控视频映射至时空剖面并自适应构建视频摘要。

2　所提方法

本文提出了一种顾及方向信息的时空联合监控视频摘要方法,该方法的整体流程如图 1 所示,主要包括 4 个部分:(1) 水平切片的生成与预处理;(2) 目标运动方向判定;(3) 运动片段检测;(4) 自适应构建视频摘要。以下将对其具体过程进行阐述。

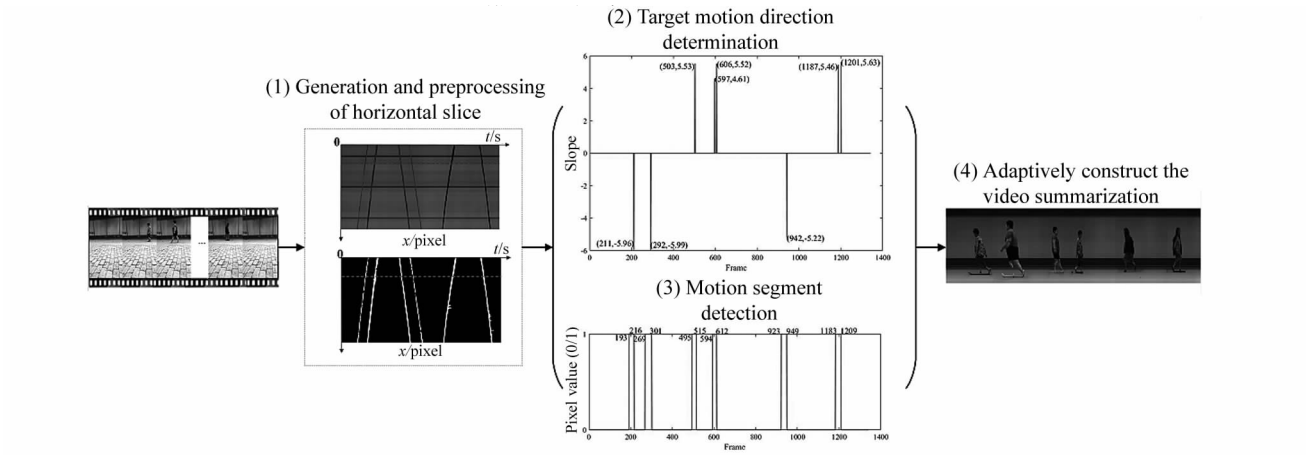

图 1　本文所提方法整体流程图

Fig. 1　The whole flow chart of the proposed method in this paper

2.1　水平切片的生成与预处理

时空切片是由时间维与空间维或组成的二维图像，其记录了完整的目标历史信息[17]，是一种高效的视频时空分析方法。定义高为 H 宽为 W 的视频 $V=[F_1,F_2,\cdots,F_i,\cdots,F_t]$，其中，$F_i$ 代表视频的第 i 帧，t 代表视频帧数，设 $T(F_i)=\boldsymbol{S}_i$ 是将 F_i 映射为 $W\times1$ 的列向量的操作[18]，则水平切片 $\boldsymbol{Sli\_H}$ 是一幅大小为 $W\times1$ 的二维图像，其表达式为：

$$\boldsymbol{Sli\_H}=[T(F_1)\quad T(F_2)\quad\cdots\quad T(F_i)\quad\cdots\quad T(F_t)]=[\boldsymbol{S}_1\quad\boldsymbol{S}_2\quad\cdots\quad\boldsymbol{S}_i\quad\cdots\quad\boldsymbol{S}_t],\tag{1}$$

式中，$\boldsymbol{S}_i$ 为：

$$\boldsymbol{S}_i=(p_h^1\quad p_h^2\quad p_h^3\quad\cdots\quad p_h^j\quad\cdots\quad p_h^w)'_{t=i},\tag{2}$$

式中，p_h^j 表示(h,j)位置的像素值，其中 $h\in[1,H]$，$j\in[1,W]$。

视频水平切片除目标轨迹外，还包含冗余的背景信息，混合高斯模型作为对单高斯背景模型的一种扩展和改进，对抖动、光线变化等具有较好的适应性，因此本文选用混合高斯模型以去除背景。将 $\boldsymbol{S}_i$ 作为混合高斯模型的一个输入并将其用个单高斯模型描述，将新输入的 $\boldsymbol{S}_{i+1}$ 与 λ 个高斯分布的均值依次进行比较，依照匹配条件找到匹配的高斯分布模型即可结束匹配过程。匹配条件为：

$$|\boldsymbol{S}_{i+1}-\mu_{\lambda,i}|<2.5\sigma_{\lambda,i},\tag{3}$$

式中，$\mu_{\lambda,i}$ 和 $\sigma_{\lambda,i}$ 分别代表第 i 列第 λ 个单高斯模型的均值和标准差。若符合匹配条件的高斯模型存在，则仅对第一个匹配的高斯模型的所有模型参数进行更新，而对于其他高斯分布模型，只更新权值 $w_{\lambda,i+1}$，模型的均值和标准差保持不变；若满足条件的高斯模型不存在，则建立一个具有较高方差、较低权重、均值为 $\boldsymbol{S}_{i+1}$ 的高斯分布模型来替换第 λ 个高斯模型。图 2 展示了去除背景前后的 TTHS 对比图。

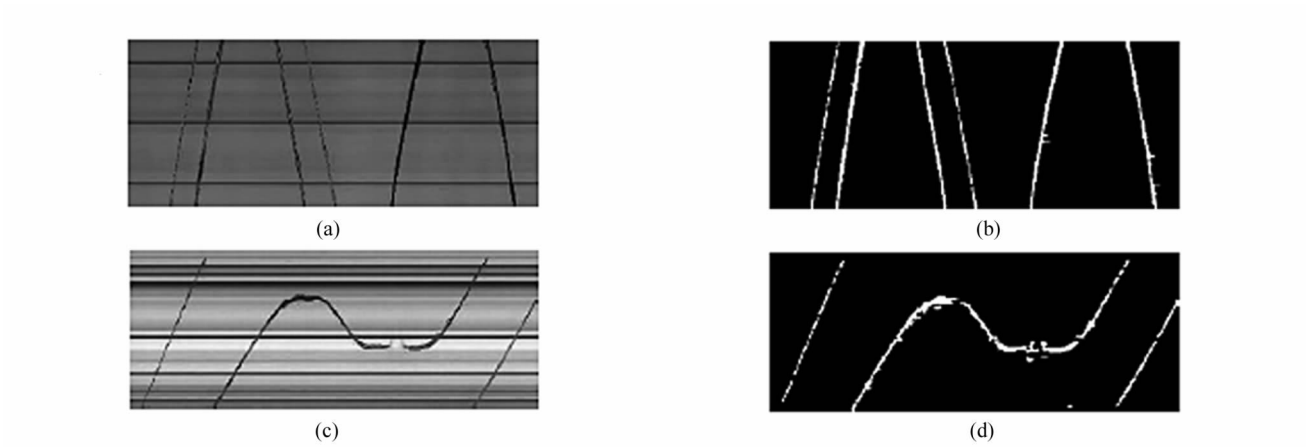

图 2　去除背景前后的 TTHS 对比图：(a) 去除背景前的直线 TTHS 图；(b) 去除背景后的直线 TTHS 图；(c) 去除背景前的曲线 TTHS 图；(d) 去除背景后的曲线 TTHS 图

Fig. 2　Comparison of TTHS before and after removing the background: (a) Straight line TTHS diagram before removing the background; (b) Straight line TTHS diagram after removing the background; (c) Curve TTHS diagram before removing the background; (d) Curve TTHS diagram after removing the background

2.2　目标运动方向判定

所有向前移动的目标在时空切片中都面朝左侧，导致其真实运动方向难以获知[9]，因此本文提出利用 TTHS 辅助判定目标运动方向的方法。运动目标会在水平切片留下明显的 TTHS，TTHS 可具体分为直线类型和不规则曲线类型两类，直线 TTHS 的检测比较容易，对于曲线 TTHS 的检测，本文利用数学和物理学中的以直代曲思想实现不规则曲线向直线的转化，作为后续处理的基础。

本文提出的目标运动方向确定方法需依据 TTHS 斜率参数完成判定，因此精准高效的 TTHS 直线段检测是必要的，Hough 检测是最常用的直线段检测方法，但该方法检测精度不够且对不规则曲线型 TTHS 很难有较好的检测结果，为此，本文采用 LSD(line segment detector)算法[19]实现 TTHS 直线段检测，其原理是通过计算图像中所有点的梯度大小和方向并将梯度变换小且相邻的点作为一个连通域，对连通域做改善和筛选以保留满足条件的域，进而得到最后的直线段检测结果。但由于 LSD 算法缺少筛选合并机制，使得图像局部密集区域内大量

相似的线段被分割为众多小线段,因此引入线段分组策略[20]以合并潜在的同源线段,通过计算基线 l_i 的相邻线段到 l_i 的距离 d,并设定线段近似标准 $\delta_l=\mu\cdot|l_i|$,其中,μ 为比例因子,$|\cdot|$ 为线段 · 的长度,若 $d>\delta_l$,则取消线段分组,反之,减小两线段的角度差 θ,如果 θ 小于角度近似标准 δ_θ,则对线段进行分组。角度的近似标准 δ_θ 为:

$$\delta_\theta=\varphi\left|\frac{1}{\exp[X(\lambda+Y)]}-1\right|, \tag{4}$$

式中,X、Y 和 φ 为设定参数,λ 为归一化系数。最终检测到的 TTHS 可以表示为 $\phi=\{l_1,l_2,l_3,\cdots,l_n\}$,每条线段 l_i 被定义为 $l_i=\{x_{1i},t_{1i},x_{2i},t_{2i},k_i\}$,$(i=1,2,\cdots,n)$,其中$(x_{1i},t_{1i})$和$(x_{2i},t_{2i})$分别为线段的起点和终点坐标,$k_i$ 为斜率:

$$k_i=\frac{t_{2i}-t_{1i}}{x_{2i}-x_{1i}}(t_{2i}>t_{1i}\text{ 恒成立})。 \tag{5}$$

目标轨迹直线段可分为两种情况:1) $x_{2i}>x_{1i}$,对应空间域中的 left→right,此时 $k_i>0$;2) $x_{2i}<x_{1i}$,对应空间域中的 right→left,此时 $k_i<0$;即:

$$\begin{cases}k>0,\text{left}\rightarrow\text{right}\\k<0,\text{right}\rightarrow\text{left}\end{cases}。 \tag{6}$$

依据目标在标识线上的 TTHS 斜率值的正负判定该目标的运动方向,并以视频帧号为横坐标,斜率值为纵坐标,生成斜率统计图。对测试视频作目标运动方向判定,生成的斜率统计图及其对应的视频帧如图 3 所示。

由图 3 可知,第 211、292、942 帧中目标运动方向为 right→left,其对应的 TTHS 斜率值为负;第 503、597、606、1 187、1 201 帧中目标运动方向为 left→right,其对应的 TTHS 斜率值为正,说明真实场景中目标的运动情况与理论保持一致,印证了本文所提方法的正确性。

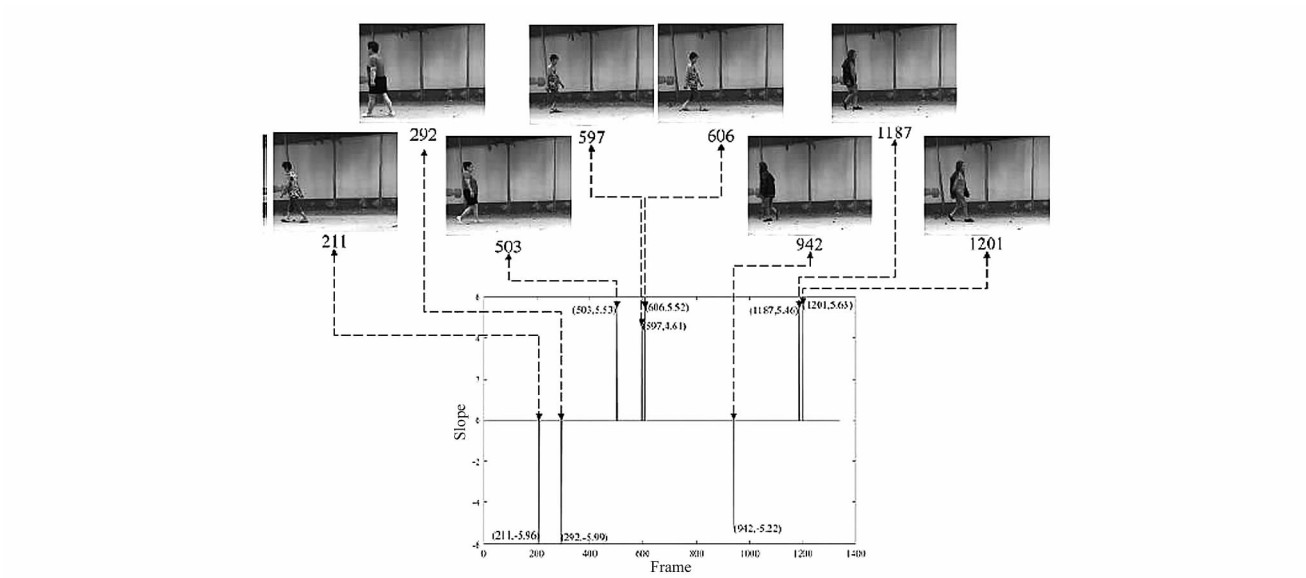

图 3　斜率统计图及其对应视频帧

Fig. 3　Slope statistical graph and its corresponding video frames

2.3　运动片段检测

本文将采样域内不含运动目标的片段称为静止片段,含有运动目标的片段称为运动片段,运动片段重要内容多、信息量丰富,针对其作特殊处理可保留视频的主要信息。现有的运动片段检测方法大多通过目标检测[21]等处理视频空域全量数据、逐帧检测以达到目的,计算量较大且耗时长,为此,本文将空域的视频帧检测问题转化为时空域的 TTHS 像素点分析问题,并依此完成采样域运动片段检测。水平切片去除背景后为二值图像,其中,白色为目标时空轨迹,黑色为背景,标识线上连续的白色像素点代表运动片段序列帧,采用形态学处理以准确检测运动片段,算法具体步骤如下:

步骤 1 输入大小为 $H_{Img}\times W_{Img}$ 的水平切片二值化图像 Img,确定标识线为 $L_{\text{ide}}=(p_j^1\quad p_j^2\quad p_j^3\quad\cdots\quad p_j^{W_{bg}})$,令计数值 $i=1$,阈值 $\zeta=4$,修正值 $\gamma=5$。

步骤 2 迭代停止条件:$i>W_{Img}$,跳转步骤 6,否则继续步骤 3。

步骤 3 如果 $p_j^i=0$,输出 p_j^i 值,令 $i=i+1$,迭代步骤 2。

步骤 4 如果 $p_j^i=1$ 且 $\exists(p_j^{i+1},\cdots,p_j^{i+\zeta-1})=0$,则令 $p_j^i=0$,$i=i+1$,,返回步骤 2。

步骤 5 如果 $p_j^i=1$ 且 $\forall(p_j^{i+1},\cdots,p_j^{i+\zeta-1})=1$，令 $num=i$，$i=i+\zeta-1$，求解子问题：

①令 $i=i+1$，若 $p_j^i=1$，输出 p_j^i 值，返回步骤①；

②若 $p_j^i=0$ 且 $\forall(p_j^{i+1},\cdots,p_j^{i+\zeta-1})=1$，则令 $p_j^i=1$，返回步骤①，否则继续步骤③；

③令 $\forall(p_j^{num-\gamma},p_j^{num-\gamma+1},\cdots,p_j^{i+\gamma-1})=1$，$i=i+\gamma$，完成像素值修正，返回步骤 2。

步骤 6 “frame→x 轴，pixel value→y 轴”生成运动片段序列图，算法终止。

对测试视频作运动片段检测，所生成的运动片段序列图及其对应视频帧如图 4 所示。

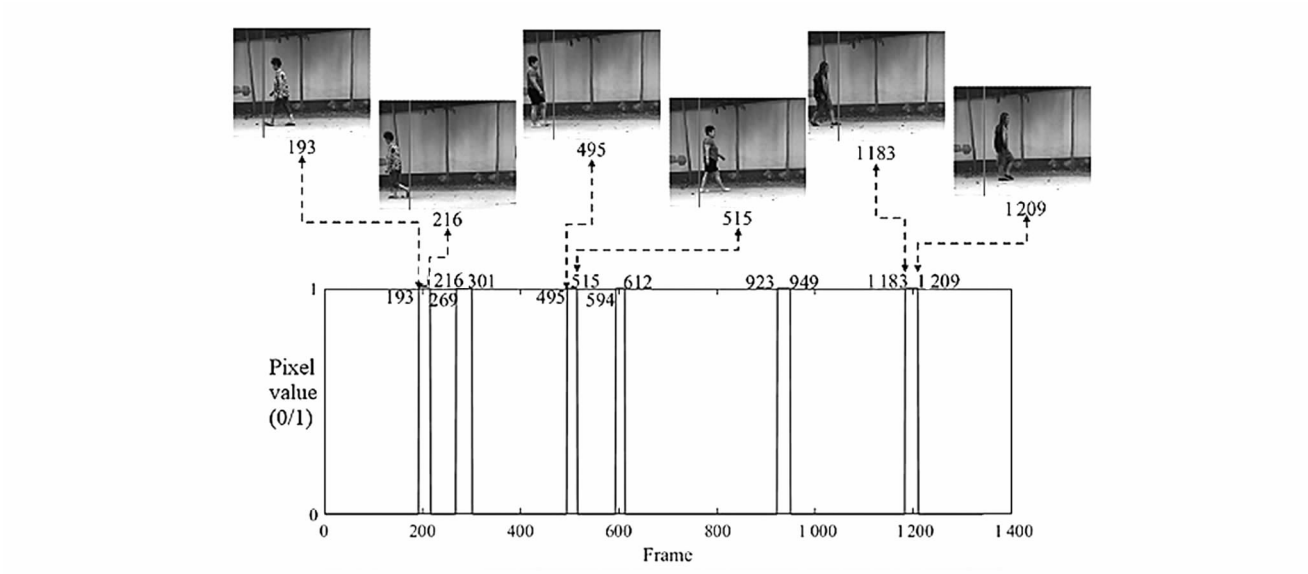

图 4　运动片段序列图及其对应视频帧

Fig. 4　Sequence diagram of motion segment and its corresponding video frames

由图 4 可知，第 193、269、495、594、923、1 183 帧为目标进入采样域的帧，第 216、301、515、612、949、1 209 帧为目标退出采样域的帧，目标时序图中像素值为 1 所对应的 193～216、269～301、495～515、594～612、923～949、183～1 209 帧为运动片段序列帧，即目标跨越采样域的帧序列。所提方法生成的运动片段序列图与目标实际运动情况一致，印证了本文算法的正确性。

2.4　自适应构建视频摘要

静止片段属于输入视频的冗余片段，包含较少的信息量，运动片段包含正反向运动目标，为进一步降低数据处理计算复杂度，对静止片段与运动片段做针对性处理。对于静止片段，采用单列采样顺序拼接，以较低计算成本保证摘要时间连续性；对于运动片段，依据斜率统计图寻找运动片段序列帧范围对应的斜率值，并依据斜率值作对应拼接处理。此外，为保证摘要中目标的外观与空域保持一致，在摘要生成过程中加入自适应目标形变算法[22]以确保目标的外观特征保持不变，算法具体步骤如下：

步骤 1 输入大小为 $H_{Img}\times W_{Img}$ 的水平切片二值化图像 Img，令计数值 $f=1$。

步骤 2 迭代停止条件：$f>W_{Img}$，跳转步骤 6，否则继续步骤 3。

步骤 3 若 $p_j^f=0$，则属于静止片段，取 $\Delta x=1$ 个采样单位，令 $f=f+1$，返回步骤 2。

步骤 4 若 $p_j^f=1$，则属于运动片段，假设该运动片段起始帧为 $f_{\min}$，终止帧为 $f_{\max}$。

步骤 5 遍历 $f_{\min}\rightarrow f_{\max}$，寻找满足条件 $k\neq 0$ 且 $f_{\min}\leq f_k\leq f_{\max}$ 的帧号 f_k；

①f_k 不存在，则作修正：令 $\forall(p_j^{f_{\min}},\cdots,p_j^{f_{\max}})=0$，$f=f_{\max}+1$，返回步骤 2；

②f_k 不存在：$k>0$ 时，依照 $f_{\max}\rightarrow f_{\min}$，取 $\Delta x=|k|$ 个采样单位；$k<0$ 时，依照 $f_{\min}\rightarrow f_{\max}$，取 $\Delta x=|k|$ 个采样单位；$k=0$，目标与背景相融合，将前景作为背景处理。

令 $f=f_{\max}+1$，返回步骤 2。

步骤 6 输出自适应生成的监控视频摘要，算法终止。

经过以上步骤，可自适应生成顾及方向信息的时空联合监控视频摘要。

3　实验及结果分析

3.1　实验环境及数据

本文实验的硬件环境为处理器 Intel(R) Core (TM) i5-4210U CPU、显卡 AMD Radeon HD 8500M、内存 8 G,软件平台为 MATLAB R2016a。测试数据集为 10 段真实环境拍摄的监控视频和 2 段来自公开数据集 KTH 和 VISOR 的视频,其详细信息如表 1 所示。

表 1　实验视频基本信息

Tab. 1　Basic information of experimental videos

Video sequence	Number of objects	Frame rate	Resolution ratio	Frame number
Video 1	1	25	320×256	138
Video 2	1	30	240×480	187
Video 3	2	30	960×544	255
Video 4	5	30	1 920×1 080	361
Video 5	1	25	160×120	362
Video 6	3	60	1 920×1 080	448
Video 7	1	25	160×120	565
Video 8	4	25	1 920×1 080	1 004
Video 9	3	30	544×960	1 342
Video 10	6	30	544×960	1 423
Video 11	7	25	360×288	1 496
Video 12	4	30	606×1 080	1 573

3.2　实验环境及数据

本文选用 Video1(单目标场景)和 Video9(多目标运动场景)两段具有代表性的视频进行实验结果展示与分析,并将本文方法与当前主流的方法在主观和客观上进行评价,以下将进行具体说明。

3.2.1　主观评价

图 5 是对 Video1 处理后的结果,图 5(a)中目标的运动方向与空间域中目标的运动方向不一致,且目标轮廓参差不齐,不能很好地展示运动目标的完整面貌;而图 5(b)中的目标与其实际运动方向保持一致,且目标外观轮廓流畅饱满,与空间域中目标外观基本一致。从视觉效果上来看,本文方法很好地顾及了时空域目标的方向信息。

图 6 分别展示了文献[5]、文献[8]、文献[9]、文献[10]和本文方法对 Video9 的处理结果。由图可见,文献[9]采用多列采样得到视频摘要,以透明度表征方向信息,但由于图像分辨率的限制在一定程度上影响了用户对目标运动方向的判断,此外,还出现了目标形变问题;文献[5]和文献[8]形成的视频摘要出现了目标的“伪碰撞”,且视频摘要帧数较多,不利于结果存储和用户浏览;文献[10]方法相较于文献[5]和文献[8],生成摘要的帧数更少,也更加便于用户观看。本文所提方法生成了一个具有连续性视频时序信息的摘要,摘要中的目标运动方向与其实际运动方向保持一致,且摘要中的目标分辨率高,易于辨认。

图 5　视频摘要(Video1):
(a) 不加入顾及方向信息算法结果图;
(b) 加入顾及方向信息结果图;(c) 空间域中目标图像

Fig. 5　Video summarization (Video1):
(a) Result diagram of algorithm without considering direction information; (b) Result diagram of algorithm considering direction information;
(c) Object image in the spatial domain

3.2.2　客观评价

为了全面评价本文所提方法的客观性能,选取了帧平均处理时间(average frame processing time, $AFPT$)、帧浓缩率(frame condensation rate, FR)和碰撞率(overlap rate, OR)3 种典型的视频摘要指标进行评价,并与文献[5]、文献[8]、文献[9]、文献[10]结果进行了对比。

1) $AFPT$:计算处理每一张视频帧所耗费的平均时间,定义为:

$$AFPT = \frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{F_i}, \tag{7}$$

式中,T_i 代表运行第 i 段视频所花费的时间,F_i 代表第 i 段视频的帧数,N 代表总测试视频段数量。视频的 $AFPT$ 越小,算法复杂度越低。

2) FR[23]:计算处理后的视频帧数与原视频帧数的比值,定义为:

$$FR = |S| / |V|, \tag{8}$$

式中,$|S|$ 为处理后的视频帧数目,$|V|$ 为处理前的视

频帧数目。FR越小，说明得到的摘要越简洁，视频处理效果越好。

3) OR[23]：统计视频摘要中运动目标的碰撞情况，即前景重合像素数占处理后视频总像素数的比例，定义为：

$$OR = \frac{1}{w \cdot h \cdot |S|} \sum_{t=1}^{w \cdot h \cdot |S|} \{1 \mid \text{if } p(t) \in \text{the collision foreground}\}, \tag{9}$$

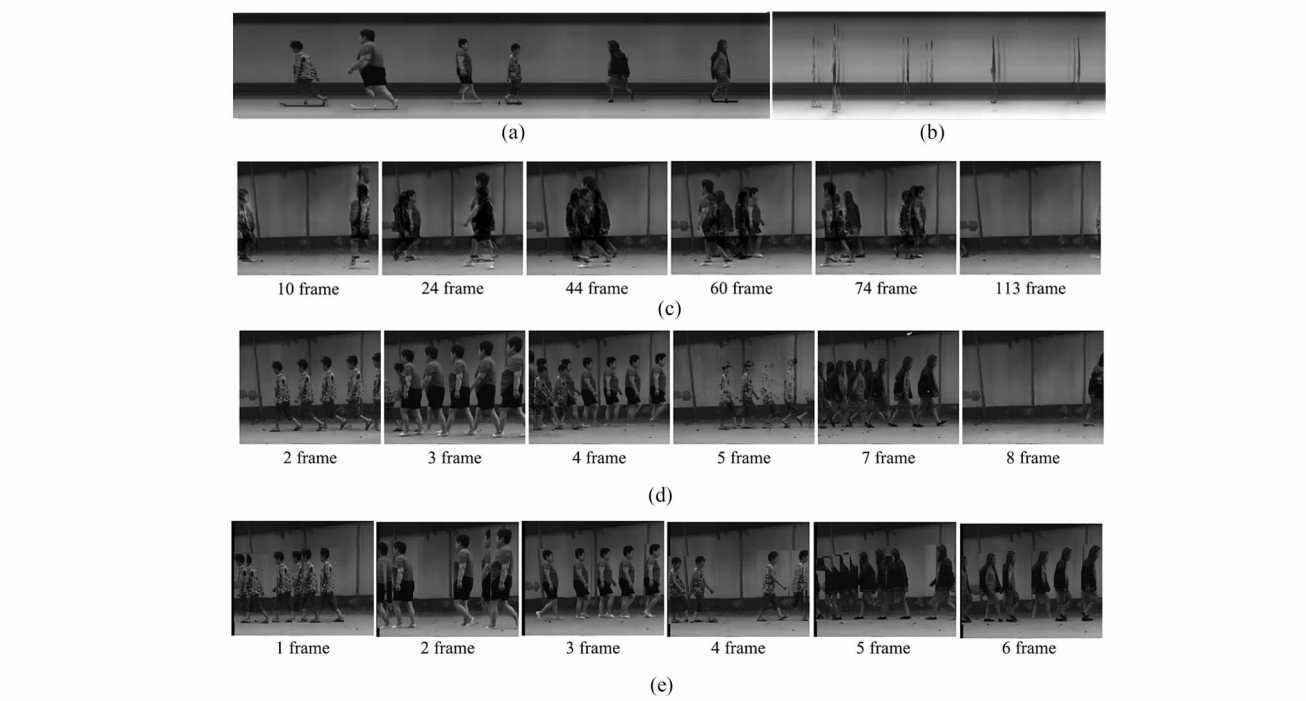

图 6　视频摘要 (Video9)：(a)本文方法；(b)文献[9]；(c)文献[5]；(d)文献[8]；(e) 文献[10]

Fig. 6　Video summarization (Video9): (a)Ours; (b) Ref.[9]; (c) Ref.[5]; (d) Ref.[8]; (e) Ref.[10]

式中，w、h 和 $|S|$ 分别代表处理后视频帧的宽度、高度和数量，$p(t)$代表视频摘要中的某一像素。OR 越小，说明处理效果越好。

本文采用 5 种不同方法对表 1 中的测试视频进行处理，并计算了生成的视频摘要的评估结果，实验结果对比如表 2 所示。

表 2　实验结果对比

Tab. 2　Comparisons of experimental results

Methods	$AFPT$/(s · frame^{-1})	FR	OR
Ref.[5]	1.270	0.084 2	0.137 2
Ref.[8]	1.916	0.027 9	0.006 0
Ref.[9]	0.489	**0.002 5**	0.002 9
Ref.[10]	1.749	0.005 0	0.007 5
Ours	**0.374**	**0.003 9**	**0.000 2**

由表 2 可知：1) 所提方法在 $AFPT$ 及 OR 上的评估结果明显优于其他对比方法；2) 所提方法的 FR 与文献[9]方法估评结果相当，但优于其他对比方法。

综上所述，相比于对比方法，所提方法很好地解决了视频摘要方法计算复杂度高的问题，且能够顾及方向信息生成简洁紧凑的视频摘要，具有良好的用户体验。

4　结　论

本文提出了一种顾及方向信息的时空联合监控视频摘要方法，该方法将 TTHS 斜率作为时空域目标的运动方向确定依据，并通过统计采样域运动片段序列帧以确定目标在视频中的时序位置，依据运动片段序列帧及其运动方向自适应构建视频摘要。实验结果表明，所提方法的 $AFPT$ 达到了 0.374 s，从主观和客观分析可得，所提方法解决了现有的视频摘要方法计算复杂度高的问题，且生成了清晰紧凑、易于浏览和检索的视频摘要。下一步将开展所提方法在移动平台上的部署和移植工作。

参考文献:

[1] LIU B. Survey of video summary[J]. Journal of Nanjing University of Information Science and Technology (Natural Science Edition),2020,12(3):274-278.
刘波.视频摘要研究综述[J].南京信息工程大学学报(自然科学版),2020,12(3):274-278.

[2] MAO Z Y,WANG Y C,WANG X,et al. Vehicle video surveillance and analysis system for the expressway[J]. Journal of Xidian University,2021,48(5):178-189.
毛昭勇,王亦晨,王鑫,等.面向高速公路的车辆视频监控分析系统[J].西安电子科技大学学报,2021,48(5):178-189.

[3] BASKURT K B,SAMET R. Video synopsis:a survey[J]. Computer Vision and Image Understanding,2019,181:26-38.

[4] TIWARI V,BHATNAGAR C. A survey of recent work on video summarization: approaches and techniques[J]. Multimedia Tools and Applications,2021,80(6):27187-27221.

[5] MOUSSA M M,SHOITAN R. Object-based video synopsis approach using particle swarm optimization[J]. Signal, Image and Video Processing,2021,15(4):761-768.

[6] GE Z,ZHAO Y. A video summary method based on shortest path algorithm[J]. Journal of Hefei University of Technology (Natural Science),2021,44(2):193-198.
葛钊,赵烨.一种基于最短路径的视频摘要方法[J].合肥工业大学学报(自然科学版),2021,44(2):193-198.

[7] JI Z,FAN S F. Video summarization with random walk on hypergraph[J]. Journal of Chinese Computer Systems,2017,38(11):2535-2540.
冀中,樊帅飞.利用超图随机游走的视频摘要生成方法[J].小型微型计算机系统,2017,38(11):2535-2540.

[8] ANURADHA K, ANAND V, RAAJAN N R. An effective technique for the creation of a video synopsis[J]. Journal of Ambient Intelligence and Humanized Computing,2020(7):1-6.

[9] BAGHERI S,ZHENG J Y,SINHA S. Temporal mapping of surveillance video for indexing and summarization[J]. Computer Vision and Image Understanding:CVIU,2016,144:237-257.

[10] THOMAS S S,GUPTA S,SUBRAMANIAN V K. Perceptual video summarization—a new framework for video summarization[J]. IEEE Transactions on Circuits and Systems for Video Technology,2017,27(8):1790-1802.

[11] HUA R,WU X X,ZHAO W T. Video summarization by learning semantic information[J]. Journal of Beijing University of Aeronautics and Astronautics,2021,47(3):650-657.
滑蕊,吴心筱,赵文天.融合语义信息的视频摘要生成[J].北京航空航天大学学报,2021,47(3):650-657.

[12] LIU Y J,TANG S J,GAO Y B,et al. Label distribution learning for video summarization[J]. Journal of Computer-Aided Design and Computer Graphics,2019,31(1):104-110.
刘玉杰,唐顺静,高永标,等.基于标签分布学习的视频摘要算法[J].计算机辅助设计与图形学学报,2019,31(1):104-110.

[13] ZHANG K,CHAO W L,SHA F,et al. Video summarization with long short-term memory[C]//The 14th European Conference on Computer Vision,October 8-16,2016,Amsterolam,The Netherlands. Berlin:Springer International Publishing,2016:766-782.

[14] JI Z,JIANG J J. Video summarization based on decoder attention mechanism[J]. Journal of Tianjin University (Science and Technology),2018,51(10):1023-1030.
冀中,江俊杰.基于解码器注意力机制的视频摘要[J].天津大学学报(自然科学与工程技术版),2018,51(10):1023-1030.

[15] JI Z,XIONG K,PANG Y,et al. Video summarization with attention-based encoder-decoder networks[J]. IEEE Transactions on Circuits and Systems for Video Technology,2020,30(6):1709-1717.

[16] WU G L,LI L T,GUO Z Z,et al. Video summarization generation model based on improved bi-directional long short-term memory network[J]. Journal of Computer Applications,2021,41(7):1908-1914.
武光利,李雷霆,郭振洲,等.基于改进的双向长短期记忆网络的视频摘要生成模型[J].计算机应用,2021,41(7):1908-1914.

[17] ZHANG Y Z,TAO R,WANG Y. Motion-state-adaptive video summarization via spatiotemporal analysis[J]. IEEE Transactions on Circuits and Systems for Video Technology,2017,27(6):1340-1352.

[18] SOUZA M,MAIA H,VIEIRA M B,et al. Survey on visual

rhythms: a spatio-temporal representation for video sequences[J]. Neurocomputing, 2020, 402: 409-422.

[19] GIOI R, JAKUBOWICZ J, MOREL J M, et al. LSD: a fast line segment detector with a false detection control[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(4): 722-732.

[20] FANG Q, WANG X H, SU J. Simultaneous localization and mapping algorithm with point and line features based on grouping strategy[J]. Laser and Optoelectronics Progress, 2021, 58(14): 1415003.
方琪, 王晓华, 苏杰. 基于分组策略的点线特征融合同步定位与地图构建算法[J]. 激光与光电子学进展, 2021, 58(14): 1415003.

[21] XU G Y, YI 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-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.

[22] ZHANG Y, GUO Y, ZHANG J. Adaptive restoration of pedestrian appearance in the spatiotemporal domain[J]. The Journal of Engineering, 2022, 2022(1): 10-15.

[23] NAMITHA K, NARAYANAN A, GEETH M. Interactive visualization-based surveillance video synopsis[J]. Applied Intelligence, 2021, 52: 3954-3975.

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.