MBO:基于多目标平衡优化的监控视频浓缩

张云佐 朱鹏飞

(石家庄铁道大学信息科学与技术学院 石家庄 050043)

摘 要 视频浓缩可在极大压缩视频长度的同时完整保留目标运动信息,在学术界和工业界受到广泛关注.然而现有浓缩方法无法精准保留目标之间的交互行为,且难以平衡压缩和碰撞,严重阻碍了视频浓缩的性能提升和实际应用.为此,本文提出了一种基于多目标平衡优化的监控视频浓缩方法(Multi-Objective Balance Optimization,MBO).首先,提出了一种基于目标交互帧数量和动态阈值对比的交互行为判断方法,用以组建多目标单元,结合目标在每帧的移动方向并采用动态阈值提升交互行为判断的准确性;其次,定义了碰撞矩阵和插入位置占比率,分别记录目标碰撞和插入位置深浅;然后,提出了一种压缩与碰撞的动态平衡方法,以优化重排目标,能在极大程度缩短视频长度的同时减少产生的目标碰撞;最后,融合视频背景和重排后的目标生成浓缩视频.VISOR、CAVIAR和KTH等多个数据集上的实验结果表明,相较于当前主流方法,本文所提方法保留交互行为的F-score提升高达0.472,并且能够有效平衡压缩和目标碰撞.

关键词 视频浓缩;多目标;平衡优化;交互行为

中图法分类号 TP391 **DOI号** 10.11897/SP.J.1016.2024.02104

MBO: Surveillance Video Synopsis Based on Multi-Objective Balance Optimization

ZHANG Yun-Zuo ZHU Peng-Fei

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

Abstract Video synopsis, which can greatly compress video length while preserving complete object motion information, has received widespread attention in both academic and industrial circles. However, existing synopsis methods cannot accurately preserve the interactive behaviors between objects and have difficulties in balancing compression and collisions, which seriously hinders the performance improvement and practical application of video synopsis. To address this issue, this paper proposes a surveillance video synopsis method based on multi-objective balance optimization (MBO). Firstly, a method for judging interactive behaviors based on the number of interactive frames and dynamic threshold comparison is proposed to form multi-objective units, combining the movement direction of the object in each frame and using dynamic thresholds to improve the accuracy of interaction behavior judgment. Secondly, the collision matrix and insertion position ratio are defined to record target collisions and the depth of insertion positions, respectively. Then, a dynamic balancing method between compression and collisions is proposed to optimize the rearrangement of objects, can greatly compress video length while reducing object collisions. Finally, the video background and rearranged objects are fused to generate the synopsis video. Experimental results on multiple datasets such as VISOR, CAVIAR, and KTH show

收稿日期:2023-05-15;在线发布日期:2024-05-31.本课题得到国家自然科学基金(No.61702347,NO.62027801)、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)、中央引导地方科技发展资金项目(226Z0501G)、研究生创新项目(YC2023081)资助.**张云佐**(通信作者),博士,副教授,中国计算机学会(CCF)高级会员,主要研究领域为计算机视觉、人工智能、大数据.E-mail:zhangyunzuo888@sina.com.**朱鹏飞**,硕士研究生,主要研究领域为计算机视觉、图像处理.

that compared with current mainstream methods, our method improves the F-score of preserving interactivity by up to 0.472 and can effectively balance compression and object collisions.

Keywords　video synopsis; multi-objective; balance optimization; interactive behavior

1　引　言

视频监控系统以其直观、便捷、高效等特点在智慧安防、交通管控、国防军事、文化教育等诸多领域发挥着越来越重要的作用,由此产生的监控视频数据量呈井喷式增长,给视频存储和检索带来了极大困难[1].随着5G技术的快速发展与普及,视频监控系统逐步实现网络化、多域化、高清化,致使此问题变得愈发凸显.监控视频呈现海量、冗余等特点,直接浏览散落其中的零星运动目标十分困难,以致大部分监控视频直接被舍弃[2-4],如何高效利用冗长视频中的有效信息是当前亟待解决的难题.

为此,研究人员提出了多种视频压缩方法缩短视频长度.常见的有视频快进[5]、视频描述[6]、视频摘要[7-9]和视频浓缩[10-11]等.视频快进利用不同的采样方式压缩视频长度,此类方法只适用于压缩变化缓慢的视频,并且容易丢失视频中的重要内容.视频描述将视频内容转换为文字供用户快速浏览,但文字表达的内容不够直观.视频摘要通过提取输入视频中的关键帧或关键镜头来表征原视频,但相邻关键帧或关键镜头之间内容关联较少,丧失了原视频的动态性和连贯性.

视频浓缩也称动态视频摘要,旨在将不同运动目标在时空域进行移动,并融合至相同视频背景,能够在短时间内展示多个运动目标,并能极大地缩短视频长度.鉴于其简洁性、高效性,一经提出就受到国内外研究人员的广泛关注,深度学习的兴起进一步加速了视频浓缩的飞速发展,涌现出了许多重要的研究成果.相较于其他视频压缩方法,视频浓缩不仅能够极大程度地缩短原视频,还能保留目标运动信息的完整性生成内容连贯的浓缩视频,具有较好的用户观看体验.

视频浓缩的概念最早于2006年由Rav-Acha等人[12]提出,Pritch等人[13]在此基础上将目标轨迹视作目标管采用活动聚类的方法生成浓缩视频,后续大多视频浓缩方法都是基于目标管进行拓展研究.视频浓缩主要包括运动目标检测,目标轨迹提取,轨迹优化重排和浓缩视频生成四个步骤.其中轨迹优化重排为视频浓缩的关键步骤,研究人员大多致力于改进优化重排方法以生成性能更高的浓缩视频,如Yang等人[14]提出一种基于八叉树的方法动态排列目标管,优化了浓缩视频中目标的开始时间.Ghatak等人[15]将SA和JAYA算法结合最小化能量函数来重排目标管.现有视频浓缩方法能够大幅度缩短原视频长度,但在优化重排时存在无法精准保留目标之间的交互行为,且难以平衡压缩和碰撞的问题,为此,众多学者展开了深入的研究.

现实中的目标存在如朋友手牵着手走路,孩童追逐嬉闹等交互行为.无法精准保留目标间的交互行为会导致浓缩视频内容偏离人眼视觉感知.为此,Lu等人[16]利用手工编辑方法对交互行为进行保留,但人工方法效率低下.Fu等人[17]将运动结构添加进能量函数,在最小化能量函数时保留交互行为.为进一步提高交互行为保留的准确性,Tian等人[18]定义了四种交互关系,Moussa等人[19]认为轨迹相交的目标具有交互行为.Li等人[20]将目标空间距离作为交互行为判断标准,但固定阈值难以泛化应用于其他类型视频. Namitha等人[21-22]利用最短距离保留交互行为,但此方法在处理擦肩而过的目标时容易产生误判.Zhang等人[23]考虑目标运动方向,利用全局向量计算目标移动方向,但难以处理目标移动方向大幅度改变的场景.如何精准保留目标之间的交互行为仍需进一步深入研究[24].

视频浓缩为缩短视频长度,不可避免会产生目标间的碰撞(重叠)[25].为减少碰撞,He等人[26]通过保留事件对目标进行重排,Sun等人[27]在空间域移动目标避免碰撞,而Nie等人[28]扩展视频背景来减少碰撞.Li和Nie等人[29-30]通过改变目标的大小和速度来平衡压缩和碰撞,但此方式容易出现行人比汽车更大的怪异结果.Pritch等人[31]牺牲目标顺序保证压缩性能,而Ra等人[32]以及Lin等人[33]则偏重于提升运行速度.一些研究者利用碰撞图[34-36]预防碰撞发生,而Chou等人[37]则将轨迹进行聚类平衡压缩和碰撞,但观看体验不佳.如何平衡压缩和碰撞是亟待解决的问题.

为解决上述问题,本文提出了一种基于多目标

平衡优化的监控视频浓缩方法(MBO).该方法从判断交互帧出发精准保留了目标之间的交互行为,并量化了目标碰撞和插入的位置,从而保证重排时压缩和减少碰撞的性能平衡,可将冗长的监控视频处理成简洁、信息密集且内容连贯的浓缩视频,以便用户快速浏览视频中的重要信息,具有重要的应用价值.本文主要贡献如下:

(1)为精准保留目标间的交互行为,提出了一种基于目标交互帧数量和动态阈值对比的交互行为判断方法,具有交互行为的目标被视作多目标单元统一处理.该方法结合了目标在每帧的移动方向,并且采用了动态阈值使所提方法能适应不同视频中目标间交互行为的判断.

(2)为量化目标间碰撞和多目标单元插入位置,定义了碰撞矩阵和插入位置占比率,以便重排时使缩减的视频长度和产生的目标碰撞平衡.

(3)为极大程度缩短视频时长的同时减少产生的目标碰撞,提出了一种压缩与碰撞的动态平衡方法用于优化重排.

2 MBO 方法

图 1 展示了本文所提出 MBO 方法的流程图.首先进行目标检测和跟踪获取目标管[38-39];其次判断交互行为,将具有交互行为的目标视作多目标单元;然后对多目标单元按出现时间进行排序后进行动态平衡优化重排;最后将重排后的多目标单元和视频背景重组生成浓缩视频.

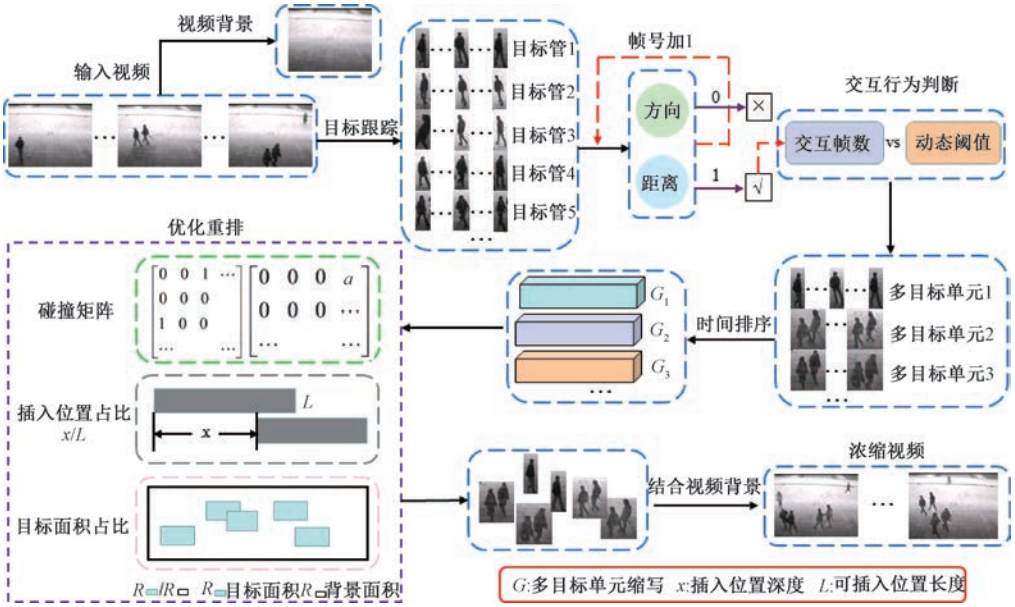

图 1 MBO 方法的流程图

2.1 交互性判断

本文将目标交互行为的判断转化为判断目标管之间的交互性.用 $T_i(1\leqslant i\leqslant n)$ 表示序号为 i 的目标管,n 为视频中目标管的个数.判断两个目标管是否具有交互性的公式如式(1)所示.

$$F(T_i,T_j)=\begin{cases}1, N(T_i,T_j)>TV(T_i,T_j)\\0, \text{else}\end{cases} \tag{1}$$

其中,$F(T_i,T_j)$ 表示目标管 T_i 和 T_j 的交互性判断结果,$F(T_i,T_j)=1$ 表示 T_i 和 T_j 之间具有交互性,而 $F(T_i,T_j)=0$ 表示 T_i 和 T_j 之间不具有交互性.$TV(T_i,T_j)$ 为目标管 T_i 和 T_j 之间的动态阈值.$N(T_i,T_j)$ 表示目标管 T_i 和 T_j 之间交互帧的数量,目标管之间的交互帧表示在该帧目标之间可能具有交互行为,计算公式如式(2)所示.

$$N(T_i,T_j)=\sum_{f=\min(T_i\cap T_j)}^{f=\max(T_i\cap T_j)} I(T_i,T_j,f) \tag{2}$$

其中,$f=\min(T_i\cap T_j)$ 表示 T_i 和 T_j 共同出现的最早帧;$f=\max(T_i\cap T_j)$ 表示 T_i 和 T_j 共同出现的最晚帧;$I(T_i,T_j,f)$ 表示目标管 T_i 和 T_j 在帧 f 交互性判断结果,计算公式如式(3)所示.

$$I(T_i,T_j,f)=\begin{cases}0, T_i\cap T_j=0\\DS(T_i^f,T_j^f)\cap O(T_i^f,T_j^f), \text{else}\end{cases} \tag{3}$$

其中,$I(T_i,T_j,f)=0$ 表明 T_i 和 T_j 在帧 f 不具有交互性;$T_i\cap T_j=0$ 表示 T_i 和 T_j 没有共同出现,所以不存在交互性;$DS(T_i^f,T_j^f)$ 表示 T_i 和 T_j 在

帧 f 的空间距离是否满足使 T_i 和 T_j 具有交互性的条件,而 $O(T_i^f,T_j^f)$ 表示 T_i 和 T_j 在帧 f 的移动方向是否满足条件. $DS(T_i^f,T_j^f)$ 的计算公式如式(4)所示.

$$DS(T_i^f,T_j^f)=\begin{cases}1, D(T_i^f,T_j^f)<d\\0, \text{else}\end{cases} \tag{4}$$

其中,$DS(T_i^f,T_j^f)=1$ 表示 T_i 和 T_j 在帧 f 的空间距离满足交互条件;$DS(T_i^f,T_j^f)=0$ 则表示不满足;d 表示设定的空间距离度量阈值,依据文献[20]设定为 1.17;$D(T_i^f,T_j^f)$ 表示 T_i 和 T_j 在帧 f 处的空间距离度量值,计算公式如式(5)所示.

$$D(T_i^f,T_j^f)=\frac{dis(T_i^f,T_j^f)}{avg(h(T_i,f),h(T_j,f))} \tag{5}$$

其中,$dis(T_i^f,T_j^f)$ 表示 T_i 和 T_j 在帧 f 检测框底部中点的欧式距离;$avg(h(T_i,f),h(T_j,f))$ 表示 T_i 和 T_j 在帧 f 检测框高度的平均值. $O(T_i^f,T_j^f)$ 的计算公式如式(6)所示.

$$O(T_i^f,T_j^f)=\begin{cases}1, 0^\circ\leqslant\theta(\boldsymbol{V}_i^f,\boldsymbol{V}_j^f)<90^\circ\\0, \text{else}\end{cases} \tag{6}$$

其中,$O(T_i^f,T_j^f)=1$ 表示 T_i 和 T_j 在帧 f 的移动方向满足使 T_i 和 T_j 具有交互性可能;$O(T_i^f,T_j^f)=0$ 则表示 T_i 和 T_j 在帧 f 的移动方向不满足使 T_i 和 T_j 具有交互性的可能;$\theta(\boldsymbol{V}_i^f,\boldsymbol{V}_j^f)$ 表示 T_i 和 T_j 在帧 f 时移动方向的夹角角度,目标同向移动时才满足具有交互性的条件. 其中 $\boldsymbol{V}_i^f,\boldsymbol{V}_j^f$ 分别为 T_i 和 T_j 在帧 f 的移动方向向量,计算公式如式(7)所示.

$$\boldsymbol{V}_i^f=(x_i^{f+1}-x_i^f, y_i^{f+1}-y_i^f) \tag{7}$$

其中,x_i^f 和 y_i^f 分别表示 T_i 和 T_j 在帧 f 检测框中心点的横纵坐标.

动态阈值 $TV(T_i,T_j)$ 的计算公式如式(8)所示.

$$TV(T_i,T_j)=\alpha * len(T_i^f\cap T_j^f) \tag{8}$$

其中,$len(T_i^f\cap T_j^f)$ 为 T_i 和 T_j 共同出现在视频中的帧数量. α 为动态阈值参数,依据实验设定.

具有交互性的目标管会被视作多目标单元统一处理. 如图 2 所示,目标 1,2 和 3 具有交互性;目标 4 和 5 具有交互性;目标 6 和其他目标无交互行为. 则通过交互行为保留将目标管 1,2 和 3 视为多目标单元 1,将目标管 4 和 5 视为多目标单元 2,将目标管 6 视作多目标单元 3.

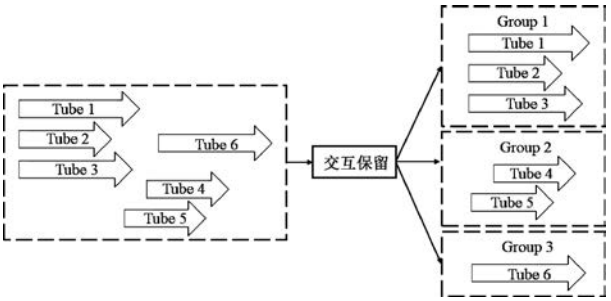

图 2　多目标单元生成示意图. Tube 表示目标管,Group 表示多目标处理单元.

2.2　碰撞矩阵

本文定义碰撞矩阵记录目标间产生的碰撞,碰撞矩阵分为记录碰撞数和碰撞度两种. 碰撞数矩阵用于记录产生碰撞的数量,碰撞度矩阵用于记录产生碰撞的程度. 其中矩阵 $\boldsymbol{f}$ 用来记录碰撞数,矩阵 $\boldsymbol{a}$ 用来记录碰撞度.

图 3 为从视频中截取的某一帧,一共有 6 个目标,其中目标 5 和目标 6 的检测框产生了重叠,表示发生了碰撞. 将图 3 转换为矩阵 $\boldsymbol{f}[6][6]$. 其中,矩阵行列数均为 6,对应图中 6 个目标. 目标 5 和 6 产生了碰撞则元素 $\boldsymbol{f}[5][6]$ 和 $\boldsymbol{f}[6][5]$ 值为 1,其他目标均未产生碰撞则矩阵中对应元素均为 0.

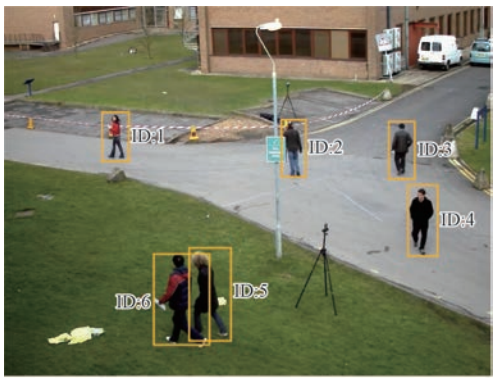

图 3　目标碰撞示意图

图 4 中检测框分别记为 $b1$ 和 $b2$. 用 A 和 B 分别表示检测框左上角和右下角的顶点,坐标为 (x,y). 当检测框 $b2$ 完全在检测框 $b1$ 的上下左右时,两个检测框不重叠,此时分别对应图 4 中 ID5,ID3,ID2 和 ID4 四种情况.

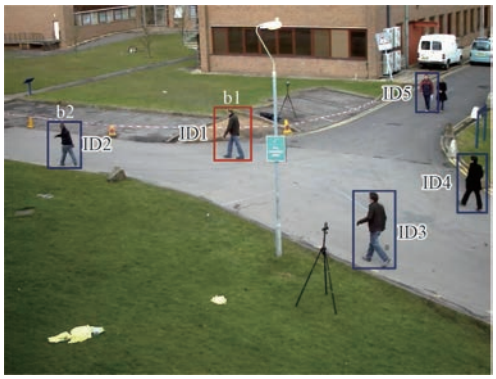

图 4　目标检测框示意图

四种情况下矩阵对应元素为 0，否则为 1，矩阵元素 $\boldsymbol{f}[i][j]$的计算公式如式(9)所示.

$$\boldsymbol{f}[i][j]=\begin{cases}0, M\leqslant N \text{ or } R\leqslant S\\ 1, \text{else}\end{cases} \tag{9}$$

其中，M 表示两个检测框右下角顶点横坐标的最小值，N 表示两个检测框左上角横坐标的最大值，R 表示两个检测框左上角纵坐标的最小值，S 表示两个检测框右下角纵坐标的最大值.

为记录碰撞数，定义 f_c 为第 f 帧检测框的碰撞速率，计算公式如式(10)所示.

$$f_c=\frac{2\times\sum_{j=1}^{i-1}\sum_{i=2}^{n}f[i][j]}{n\times(n-1)} \tag{10}$$

其中，n 为检测框的个数. f_c 越大表明视频帧内产生碰撞的目标数越多.

为了记录碰撞度，本文将矩阵 $\boldsymbol{a}[m][n]$每一个元素用来记录某一帧两个目标检测框重叠面积占两个目标检测框的总面积的比值，其中 m 为准备新插入来的目标数，n 为已经安排好开始时间的目标数. 若图 3 中 ID6 是新插入进来的目标，则记录碰撞度的矩阵为 $\boldsymbol{a}[1][5]$.

目标 6 为新插入的目标，则为新目标 1，对应矩阵 $\boldsymbol{a}$ 的行数 1. 剩下 5 个目标为已安排好开始时间的目标，对应矩阵 $\boldsymbol{a}$ 的列数 5. 矩阵中元素 $\boldsymbol{a}[1][5]$的值为新目标 1 和原目标 5 检测框面积的交集与并集的比值，新目标 1 与其他目标未发生碰撞则矩阵中对应元素的值为 0.

定义 f_a 为衡量第 f 帧碰撞度的目标碰撞面积率，计算公式如式(11)所示.

$$f_a=\frac{\sum_{j=1}^{m}\sum_{i=1}^{n}a[i][j]}{m\times n} \tag{11}$$

其中，m 和 n 分别为矩阵 $\boldsymbol{a}$ 的行和列数，f_a 的值越大表明插入过程中产生的碰撞越多.

2.3 插入位置占比率

依据 2.1 节所述方法，具有交互行为的目标将作为多目标处理单元统一处理. 为保证目标出现的顺序不变，本文将所有多目标单元按最先出现的目标进行先后排序，最先出现目标的多目标单元安排在浓缩视频的第一帧，后续单元进行优化重排安排开始时间.

如图 5 所示，箭头 x，y 表示视频背景的横纵坐标轴，f 表示视频的帧序号. 图 5 中的 1～5 为已经重排好的多目标单元，开始帧分别为 f_0、f_a、f_b、f_c 和 f_d. 现需安排多目标单元 6 的开始帧，为了保证不破坏目标出现的先后顺序，多目标单元 6 应该在 5 出现之后才出现，即开始帧最小为 f_d. 为了控制碰撞以及其他因素，多目标单元 6 需要找到一个合适的开始帧，而前五个多目标单元最迟在帧 f_e 都离开视频，所以单元 6 从帧 f_e+1 开始一定可以插入浓缩视频. 即只需要判断单元 6 在帧$[f_d, f_e]$区间内是否可以插入，若在此区间段没有找到插入的位置，则单元 6 的开始帧为 f_e+1.

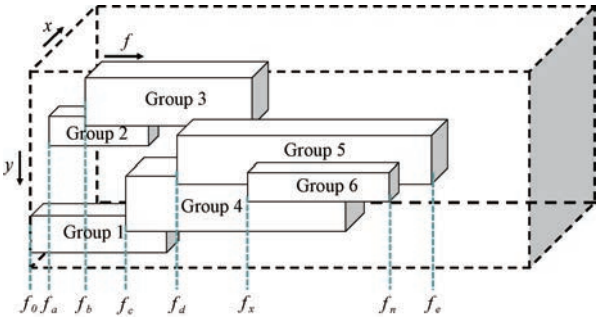

图 5 多目标单元插入示意图(x 为多目标单元 6 开始的帧号，n 为目标单元 6 结束的帧号)

在避免碰撞的情况下，为了尽可能的缩减视频长度，多目标单元应该放在比较合适的位置. 本文采用插入位置占比率来记录插入的位置，若多目标单元 6 插入在 f_x 的位置，则单元 6 的插入位置占比率计算公式如式(12)所示.

$$P_6=\begin{cases}\dfrac{x-d}{e-d}, d\leqslant x\leqslant e\\ 1, \quad x=e+1\end{cases} \tag{12}$$

其中，P_6 为多目标单元 6 的插入位置占比率，x 为单元 6 开始的帧号，d 为单元 6 前一个单元开始的帧号，e 为前五个多目标单元最迟结束的帧号. 当单元 6 开始帧为 $e+1$ 时 P_6 为 1. 第 i 个多目标单元的插入位置占比率 P_i 计算公式如式(13)所示.

$$P_i=\begin{cases}\dfrac{b_i-b_{i-1}}{\max(e_1,\cdots,e_{i-1})-b_{i-1}}, b_{i-1}\leqslant b_i\leqslant\max(e_1,\cdots,e_{i-1})\\ 1, \quad b_i=\max(e_1,\cdots,e_{i-1})+1\end{cases} \tag{13}$$

其中，b_i 和 b_{i-1} 分别为多目单元 i 和 $i-1$ 的开始帧号；$\max(e_1,\cdots,e_{i-1})$ 为前 $i-1$ 个多目标单元结束帧号的最大值. 本文采用平均插入位置占比率来衡量前面所有多目标单元插入位置占比率的综合值，计算公式如式(14)所示.

$$S_i=\frac{\sum_{m=1}^{i-1}P_i}{i-1} \tag{14}$$

其中，S_i 为前 $i-1$ 个多目标单元的平均插入位置占比率，S_i 越大表明前面多目标单元插入的位置平均比较靠后，视频长度会较长，则多目标单元 i 的插入位置应尽量靠前. S_i 越小表明前面多目标单元插入的位置都比较靠前，视频碰撞会较大，则多目标单元 i 的插入位置应尽量靠后.

2.4　优化重排

本部分介绍平衡压缩和碰撞的优化重排方法. 原视频经过交互行为保留后的多目标单元集合为 G，$G=\{\text{Group1},\text{Group2},\cdots,\text{Group}n\}$，其中 n 为视频多目标单元的数量. 为方便表示，后续用 G_i 表示排序后第 i 个多目标单元的缩写.

由 2.3 节知，多目标单元 G_i 能考虑的插入位置为 $[b_{i-1},\max(e_1,e_2,\cdots,e_{i-1})+1]$. 如图 6 所示，a 部分为单元 6 还未插入浓缩视频时部分帧的侧面图，图中黑色方框为视频背景，方框为一些多目标单元中目标的检测框. 如 a 部分 f_d 帧，具有 10 个目标检测框，将 f_d 帧到 f_e 帧的侧面图截取出来得到 b 部分，将单元 6 也以侧面图的方式组合起来得到 c 部分.

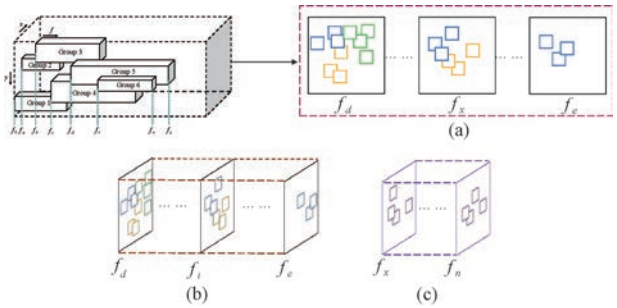

图 6　多目标单元展示图

现需要将 c 部分插入在 b 部分某个位置，插入时需要使产生的碰撞、视频内检测框总面积等满足用户需求. 如图 7 中 a 部分所示，当 f_n 小于等于 f_e 时，此时只需要考虑在多目标单元 6 长度内产生的碰撞以及视频内检测框总面积等因素即可；而当 f_n 大于 f_e 时，如图 7 中 b 部分所示，需要考虑的范围为 $[f_x,f_e]$. 当两种情况下均没有满足要求的位置时，则将开始帧序号设置为 f_e+1，此时不会产生多余的碰撞. 推广至一般情况，多目标单元 G_i 插入时需要考虑的长度计算公式如式(15)所示.

$$L_i=\begin{cases}e_i-b_i+1, & e_i\leqslant\max(e_1,\cdots,e_{i-1})\\ \max(e_1,\cdots,e_{i-1})-b_i+1, & b_i\leqslant\max(e_1,\cdots,e_{i-1})<e_i\\ 0, & b_i=\max(e_1,\cdots,e_{i-1})+1\end{cases}\tag{15}$$

其中，L_i 为多目标单元 G_i 插入时需要考虑的帧长度，e_i 为多目标单元 i 结束的帧号.

如图 7 所示，截取插入多目标单元 6 过程中的某一帧，c 部分为前面已安排好开始时间的某一帧，d 部分为多目标单元 6 中的某一帧，将二者结合得到 e. 可以看出此时 e 中目标检测框重叠较多，即产生的碰撞较多. 同时，e 中目标数量众多，占视频背景总比例较大，影响视觉观感. 因此在进行插入时需要控制碰撞和时长. 同时要兼顾视频中目标检测框总面积.

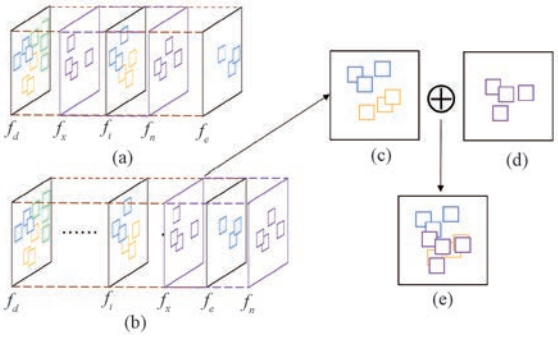

图 7　多目标单元插入展示图

G_i 插入时间的计算公式如式(16)所示.

$$b_i=\begin{cases}i, & C\cap A\cap O=1\\ \max(e_1,\cdots,e_{i-1})+1, & \text{else}\end{cases}\tag{16}$$

其中，C，A 和 O 分别表示碰撞数，碰撞度和目标检测框面积是否满足插入条件，计算公式如式(17)，式(18)和式(19)所示. 只有当三个条件的值均为 1 时，b_i 的值才被设置为 i.

$$C=\begin{cases}1, & \dfrac{\sum_{f=i}^{i+L_i}f_c}{L_i}\leqslant\omega\\ 0, & \text{else}\end{cases}\tag{17}$$

其中，ω 为碰撞数参数.

$$A=\begin{cases}1, & \dfrac{\sum_{f=i}^{i+L_i}f_a}{L_i}\leqslant(\alpha+\beta\times S_i)\\ 0, & \text{else}\end{cases}\tag{18}$$

其中，α 和 β 为动态平衡压缩和碰撞的参数.

$$O=\begin{cases}1, & \dfrac{\sum_{f=i}^{i+L_i}O_f}{L_i}\leqslant r(0.4,0.6)\\ 0, & \text{else}\end{cases}\tag{19}$$

其中，$r(0.4,0.6)$ 为 0.4 到 0.6 间的随机数，此时能在保证视觉观感的前提下较大程度压缩视频；O_f

为第 f 帧目标检测框总面积与视频背景面积的比值,计算公式如式(20)所示.

$$O_f=\frac{S_f^G}{S_v} \tag{20}$$

其中,S_f^G 为第 f 帧所有目标检测框总面积;S_v 为视频背景的面积.

G_i 在 $[b_{i-1},\max(e_1,e_2,\cdots,e_{i-1})]$ 内逐帧判断是否能插入,若无位置满足条件,则 $b_i=\max(e_1,e_2,\cdots,e_{i-1})+1$. 最后将所有重排后的目标和视频背景融合生成浓缩视频.

3 实 验

本节开展了大量实验用于验证所提 MBO 方法的有效性,以下将从数据集与评价指标,实验设置和实验结果与分析三个方面进行介绍.

3.1 数据集与评价指标

3.1.1 数据集

本文采用数据集 VISOR[40],CAVIAR[41],KTH[42] 以及文献[30]中的部分视频作为实验视频. 由于本文保留了目标间的交互行为,为了使实验结果更具有信服力,本文选取了存在交互行为的视频和不存在交互行为的视频进行实验对比. 表 1 展示具有交互行为视频的详细信息,表 2 展示不具有交互行为视频的详细信息.

表 1 具有交互行为视频数据

序号	大小	帧数	Fps	管数
V1	704×576	92	25	5
V2	384×288	3401	29	11
V3	320×240	4481	25	62
V4	640×360	51000	25	94
V5	360×270	5325	25	5
V6	384×288	390	25	4
V7	384×288	1674	25	15
V8	384×288	600	25	6

表 2 不具有交互行为视频数据

序号	大小	帧数	Fps	管数
V9	640×360	920	10	14
V10	640×360	176	10	4
V11	640×360	1376	10	8
V12	160×120	420	25	4
V13	160×120	401	10	6
V14	160×120	632	10	9
V15	368×276	825	15	3
V16	320×240	392	10	4

为选择实验参数进行了参数设定实验,表 3 展示了参数设定视频的详细数据. 三个表中均包含每个实验视频的背景大小,帧数,每秒传输帧数(Fps)和目标管数量的介绍.

表 3 参数设定视频数据

序号	大小	帧数	Fps	管数
V17	704×576	409	25	9
V18	704×576	929	25	14
V19	704×576	246	25	6
V20	704×576	481	25	11
V21	704×576	321	25	10
V22	704×576	273	25	9
V23	704×576	412	25	8
V24	704×576	202	25	6

VISOR 数据集主要包含 3DPes 和 SARC3D 两个部分,视频 V1～V3,V5 和表 3 中所有视频均为此数据集中的数据. CAVIAR 数据集包含了许多不同的场景,包括目标独立行走,与他人见面,进出商店购物,战斗等,表 1 中视频 V6～V8 为此数据集数据. KTH 数据集为经典的动作识别数据集,一共包含 2391 组数据,是被广泛使用的数据集,本文选取了其中的运动数据,视频 V12～V14 为此数据集中的视频数据. 视频 V4,V9～V11 和 V15 是来自文献[30]中的数据.

3.1.2 评价指标

本文采用 F-score 作为交互行为判断的评价指标,计算公式如式(21)所示.

$$F\text{-score}=\frac{2\times P\times R}{P+R} \tag{21}$$

其中,P 和 R 分别为精准率和召回率,计算公式如式(22)和式(23)所示.

$$P=\frac{TP}{TP+FP} \tag{22}$$

$$R=\frac{TP}{TP+FN} \tag{23}$$

其中,TP 表示目标在现实中具有交互行为,检测结果也具有交互行为;FN 表明目标在现实中具有交互行为,但是检测结果为不具有交互行为;FP 表明目标间在现实中没有交互行为,但是判断结果为具有交互行为. F-score 越大表明目标交互行为判断得越准确.

本文采用帧压缩率 FR 和碰撞率 OR 来衡量浓缩视频客观性能. FR 用来衡量视频浓缩缩短视频长度的性能,FR 值越小表明压缩视频性能越强. OR 用来衡量视频浓缩减少碰撞的性能,OR 值越小表明生成的浓缩视频产生的目标碰撞越少.

FR 的计算公式如式(24)所示.

$$FR=\frac{L_s}{L_o} \tag{24}$$

其中,L_s 为浓缩视频长度,L_o 为原视频长度.

OR 的计算公式如式(25)所示.

$$OR=\frac{1}{l\times w\times h}\sum_{i=1}^{l}np_i \tag{25}$$

其中,w,h 分别表示视频背景长度和宽度,l 表示视频帧的数量,np_i 表示视频第 i 帧中存在目标碰撞像素点的数量.

3.2　实验设置

动态阈值参数 α 的设定与交互行为保留的准确性密切相关.本文在视频 V17～V24 上进行参数设置的实验,α 的值从 0.1 至 0.9 进行选择,实验结果如图 8 所示.可以看出当动态阈值参数 α 为 0.5 时,多个视频上的平均 F-score 最高,所以参数 α 最终设置为 0.5.

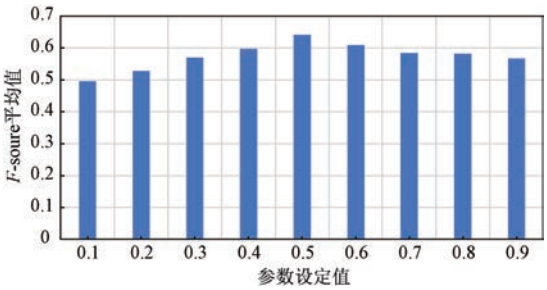

图 8　动态阈值参数设置结果

优化重排的参数决定了平衡压缩和碰撞的效果,而碰撞会影响用户的视觉观感.为此,我们找到了十名领域内人员,结合视频浓缩技术,让他们对参数进行选择,找到自我感觉视觉效果较好的视频.经过选择,参数 ω,α 和 β 最终分别设置为 0.6,0.05 和 0.2.

3.3　实验结果与分析

3.3.1　交互行为保留比较

本文将所提 MBO 方法与按空间距离和固定阈值判断的方法 FZ[20],按照最小距离以及动态阈值的方法 DT[21]还有结合全局移动方向的方法 OB[23]进行交互行为保留的比较,结果如表 4 所示.

从表 4 可以看出,除视频 V2 外,本文方法在所有视频中判断的 F 分数均最高,平均值为 0.787.而视频 V2 由于目标均在较远的区域,检测框较小使得目标帧移动方向判断不准确,导致本文所提方法的 F-score 不是最高.综上可知本文所提交互行为判断方法能够更为精准的判断并保留目标之间的交互行为.

表 4　交互行为判断 F-score 对比结果

序号	FZ[20]	DT[21]	OB[23]	MBO
V1	1.000	0.182	1.000	1.000
V2	0.125	0.125	1.000	0.333
V3	0.146	0.072	0.384	0.507
V4	0.117	0.207	0.486	0.565
V5	0.500	1.000	1.000	1.000
V6	0.400	0.286	0.400	1.000
V7	0.074	0.533	0.800	0.889
V8	0.154	0.667	0.857	1.000
平均	0.315	0.384	0.741	**0.787**

3.3.2　客观性能比较

浓缩视频的客观性能从 FR 和 OR 两个方面进行比较.为了验证所提方法的客观性能,将本文所提的 MBO 方法与粒子群优化方法 MM[19],基于事件重排优化的视频浓缩方法 ER[26]以及基于时空偏移的方法 ST[43]进行比较.

表 5 展示了在具有交互行为视频上客观性能比较结果,从表中可以看出,本文方法 FR 平均值最小,为 0.362,相较于其他三种方法分别降低了 0.065,0.106 和 0.055,表明所提方法能够将视频压缩得更短.同时,本文方法的 OR 平均值也最小,为 0.042,相较于对比方法分别降低了 0.029,0.051 和 0.010,表明本文方法所得浓缩视频产生的碰撞最少,具有更好的视觉效果.

表 5　具有交互行为视频的客观性能比较结果

序号	FR				OR			
	MM[19]	ER[26]	ST[43]	MBO	MM[19]	ER[26]	ST[43]	MBO
V1	0.598	0.663	0.630	0.565	0.041	0.034	0.024	0.021
V2	0.068	0.068	0.049	0.039	0.052	0.131	0.041	0.019
V3	0.532	0.612	0.479	0.455	0.063	0.108	0.098	0.055
V4	0.079	0.079	0.074	0.068	0.236	0.312	0.118	0.112
V5	0.069	0.069	0.096	0.095	0.084	0.085	0.065	0.063
V6	0.777	0.900	0.769	0.769	0.057	0.032	0.037	0.042
V7	0.666	0.609	0.689	0.291	0.024	0.017	0.016	0.011
V8	0.630	0.743	0.553	0.613	0.011	0.026	0.013	0.012
平均	0.427	0.468	0.417	**0.362**	0.071	0.093	0.052	**0.042**

表 6 展示了在不具有交互行为视频上客观性能比较结果，本文方法所得浓缩视频的 FR 和 OR 平均值也最小，分别为 0.215 和 0.112.这说明无论目标具不具有交互行为，本文方法均具有较好的客观性能.但并不是所有视频本文方法得到的 FR 值和 OR 均最低，尤其在具有交互行为的一些视频中.原因为具有交互性的目标往往距离较近，本身目标的检测框具有较多的重叠，较为精准判断目标间的交互行为会保留较多的碰撞.因此本文方法在精准保留交互行为的前提下，会导致一些视频的浓缩客观性能并不优异.而视频 V12 和 V15 虽不具有交互性的目标，但视频中目标均未同时出现，内容过于简单，本文动态平衡方法未发挥较好的优化作用，导致客观性能结果未达到最优值.综上，本文所提 MBO 方法能够在精准保留交互行为的基础上，使压缩视频和减少碰撞的性能平衡.

表 6 不具有交互行为视频的客观性能比较结果

序号	FR				OR			
	MM[19]	ER[26]	ST[43]	MBO	MM[19]	ER[26]	ST[43]	MBO
V9	0.257	0.290	0.248	0.220	0.104	0.113	0.091	0.087
V10	0.571	0.710	0.600	0.563	0.064	0.059	0.052	0.041
V11	0.153	0.142	0.126	0.125	0.076	0.082	0.062	0.059
V12	0.119	0.233	0.229	0.245	0.169	0.151	0.149	0.140
V13	0.195	0.180	0.190	0.147	0.203	0.212	0.194	0.177
V14	0.125	0.144	0.131	0.098	0.282	0.614	0.278	0.254
V15	0.238	0.238	0.361	0.250	0.100	0.099	0.072	0.097
V16	0.125	0.107	0.105	0.074	0.052	0.063	0.042	0.040
平均	0.223	0.256	0.245	**0.215**	0.131	0.174	0.121	**0.112**

3.3.3 主观性能比较

图 9 和图 10 分别展示了在视频 v5 和 v6 上原视频和本文方法所得浓缩视频的部分帧图像，其中子图(a)、(b)、(c)为原视频中截取的三帧图像，(d)为浓缩视频中截取的图像.

图 9(a)显示有目标正在搬运灭火器；图 9(b)显示有两个目标正在交谈，存在交互行为；图 9(c)显示有目标正离开视频场景.经过浓缩后，图 9(d)显示所提方法能够在产生较少碰撞的前提下将目标移动至同一帧内，缩短视频长度，图中的矩形方框表明所提方法正确保留了目标间的交互行为.

图 10(a)显示原视频存在一个站立不动的目标；图 10(b)和图 10(c)分别显示原视频存在一个正在行走的目标；图 10(d)显示所提方法能够使三个目标在同一帧显示，并且碰撞较少.综合而言，通过所提 MBO 方法得到的浓缩视频能够在保留交互行为的基础上，具有较好的视觉效果.

图 9 V5 原视频部分帧和浓缩视频部分帧图像

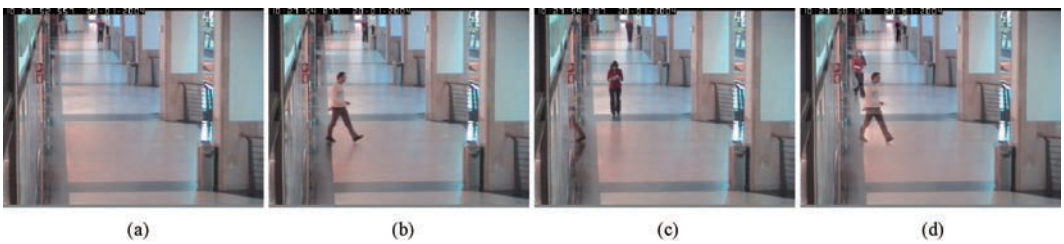

图 10 V6 原视频部分帧和浓缩视频部分帧图像

浓缩结果的好坏最终和用户的观看体验密不可分,为此我们邀请了 100 名参与者对不同方法所得结果进行打分评价,分值从 1 到 5,越高代表用户越满意,分值具体含义如表 7 所示.

表 7　分值含义表

分值	分值具体含义
5	浓缩视频结果特别好
4	浓缩视频结果较好
3	浓缩视频结果一般
2	浓缩视频结果较差
1	浓缩视频结果特别差

本文利用所有参与者评价的平均分进行主观评价比较,评价结果如图 11 和图 12 所示.从图中可以看出,存在交互行为的八个视频和不存在交互行为的八个视频上,本文所提方法的平均分均要高于其他方法,这说明本文方法得到的浓缩视频令用户观看体验更好,进一步验证了本文所提 MBO 方法的有效性.

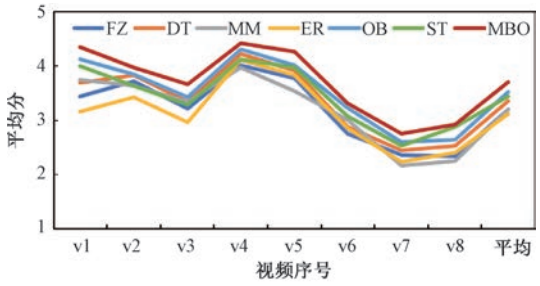

图 11　具有交互行为视频的主观评价结果

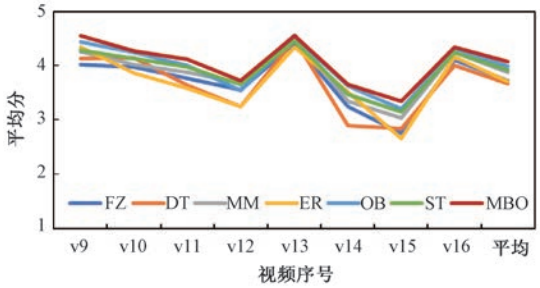

图 12　不具有交互行为视频的主观评价结果

4　结　论

本文提出了一种基于多目标平衡优化的监控视频浓缩方法(MBO).首先将交互帧数量和动态阈值比较保留目标间的交互行为;然后定义碰撞矩阵和插入位置占比率分别记录产生的碰撞和插入的位置;最后提出的优化重排方法能够平衡压缩视频和减少碰撞的性能.实验结果表明本文方法在精准保留目标间交互行为的前提下,能够提升压缩性能以及减少碰撞,使二者平衡.未来将针对更为复杂的场景作进一步研究.

参　考　文　献

[1] Zhang Yunzuo, Liu Yameng. Wu Cunyu. Attention-guided multi-granularity fusion model for video summarization. Expert Systems with Applications, 2024, 249:123568

[2] Liu Tianrui, Meng Qingjie, Huang Jun-Jie, et al. Video summarization through reinforcement learning with a 3D spatio-temporal U-Net. IEEE Transactions on Image Processing, 2022, 31:1573-1586

[3] Zhang Yunzuo, Guo Kainai, Tao Ran. Adaptive spatio-temporal tube for fast motion segments extraction of videos. IEEE Signal Processing Letters, 2022, 29:2308-2312

[4] Sun Yubao, Chen Xuhao, Mohan S, et al. Video snapshot compressive imaging using residual ensemble network. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(9):5931-5943

[5] Höferlin B, Höferlin M, Weiskopf D. et al. Information-based adaptive fast-forward for visual surveillance. Multimedia Tools and Applications, 2011, 55:127-150

[6] Guo Xianwei, Lai Hua, Yu Zhengtao, et al. Emotion classification of case-related microblog comments integrating emotional knowledge. Chinese Journal of Computers, 2021, 44(3):564-578
(郭贤伟, 赖华, 余正涛等. 融合情绪知识的案件微博评论情绪分类. 计算机学报, 2021, 44(3):564-578)

[7] Zhang Yunzuo, Zhang Jiayu, Liu Ruixue, et al. Key frame extraction based on quaternion Fourier transform with multiple features fusion. Expert Systems with Applications, 2023, 216(15):119467

[8] Li Xuelong, Zhao Bin. Video distillation (in Chinese). Science China Information Sciences, 2021, 51(5):695-734
(李学龙, 赵斌. 视频萃取. 中国科学:信息科学. 2021, 51(5):695-734)

[9] Srilakshmi S, Sai S Ch, Satvik J, et al. Video summarisation using shot boundary detection and TF-IDF vectorization//Proceedings of the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). Madurai, India, 2021:977-983

[10] Namitha K, Narayanan A, Geetha M. A synthetic video dataset generation toolbox for surveillance video synopsis applications//Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP). Chennai, India, 2020:493-497

[11] Zhang Zhensong, Nie Yongwei, Sun Hanqiu, et al. Multi-view video synopsis via simultaneous object-shifting and view-switching optimization. IEEE Transactions on Image Processing, 2020, 29:971-985

[12] Rav-Acha A, Pritch Y, Peleg S. Making a long video short: Dynamic video synopsis//Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). New York, USA, 2006:435-441

[13] Pritch Y, Rav-Acha A, Gutman A, et al. Webcam synopsis: Peeking around the world//Proceedings of the 2007 IEEE 11th International Conference on Computer Vision (ICCV). Rio de Janeiro, Brazil, 2007:1-8

[14] Yang Y, Kim H, Choi H, et al. Scene adaptive online surveillance video synopsis via dynamic tube rearrangement using octree. IEEE Transactions on Image Processing, 2021, 30:8318-8331

[15] Ghatak S, Rup S, Majhi B, et al. HSAJAYA: An improved optimization scheme for consumer surveillance video synopsis generation. IEEE Transactions on Consumer Electronics, 2022, 66(2):144-152

[16] Lu Shaoping, Zhang Songhai, Wei Jin, et al. Timeline editing of objects in video. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(7):1218-1227

[17] Fu Wei, Wang Jinqiao, Gui Liangke, et al. Online video synopsis of structured motion. Neurocomput. 2014, 135(5): 155-162

[18] Tian Yumin, Zheng Haihong, Chen Qichao, et al. Surveillance video synopsis generation method via keeping important relationship among objects. IET Comput Vis, 2016, 10(8): 868-872

[19] Moussa M M, Shoitan R. Object-based video synopsis approach using particle swarm optimization. Signal Image and Video Processing, 2021, 15:761-768

[20] Li Xuelong, Wang Zhigang, Lu Xiaoqiang. Video Synopsis in Complex Situations. IEEE Transactions on Image Processing, 2018, 27(8):3798-3812

[21] Namitha K, Narayanan A, Geetha M. Interactive visualization-based surveillance video synopsis. Applied Intelligence, 2022, 52:3954-3975

[22] Namitha K, Narayanan, A. Preserving interactions among moving objects in surveillance video synopsis. Multimedia Tools and Applications, 2020, 79:32331-32360

[23] Zhang Yunzuo, Zheng Tingting. Object interaction-based surveillance video synopsis. Applied Intelligence, 2023, 53: 4648-4664

[24] Han Lei, Li Junfeng, Jia Yunde. Human interaction recognition using spatio-temporal words. Chinese Journal of Computers, 2010, 33(4):776-784
(韩磊, 李君峰, 贾云得. 基于时空单词的两人交互行为识别方法. 计算机学报, 2010, 33(4):776-784)

[25] Hsia C H, Chiang J S, Hsieh C F. Low-complexity range tree for video synopsis system. Multimedia Tools and Applications, 2016, 75:9885-9902

[26] He Yi, Han Jun, Sang Nong, et al. Chronological video synopsis via events rearrangement optimization. Chinese Journal of Electronics, 2018, 27(2):399-404

[27] Sun Lei, Xing Junliang, Ai Haizhou, et al. A tracking based fast online complete video synopsis approach//Proceedings of the 21st International Conference on Pattern Recognition (ICPR). Tsukuba, Japan, 2012:1956-1959

[28] Nie Yongwei, Xiao Chunxia, Sun Hanqiu, et al. Compact video synopsis via global spatiotemporal optimization. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(10):1664-1676

[29] Li Xuelong, Wang Zhigang, Lu Xiaoqiang. Surveillance video synopsis via scaling down objects. IEEE Transactions on Image Processing, 2016, 25(2):740-755

[30] Nie Yongwei, Li Zhenkai, Zhang Zhensong. Collision-free video synopsis incorporating object speed and size changes. IEEE Transactions on Image Processing, 2020, 29: 1465-1478

[31] Pritch Y, Rav-Acha A, Peleg S. Nonchronological video synopsis and indexing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(11):1971-1984

[32] Ra Moonsoo, Kim W Y. Parallelized tube rearrangement algorithm for online video synopsis. IEEE Signal Processing Letters, 2018, 25(8):1186-1190

[33] Lin Longxin, Lin Weiwei, Xiao Weijun, et al. An optimized video synopsis algorithm and its distributed processing model. Soft Computing, 2017, 21:935-947

[34] Ruan Tao, Wei Shikui, Li Jia, et al. Rearranging online tubes for streaming video synopsis: A dynamic graph coloring approach. IEEE Transactions on Image Processing, 2019, 28(8):3873-3884

[35] He Yi, Qu Zhiguo, Gao Changxin, et al. Fast online video synopsis based on potential collision graph. IEEE Signal Processing Letters, 2017, 24(1):22-26

[36] He Yi, Gao Changxin, Sang Nong, et al. Graph coloring based surveillance video synopsis. Neurocomputing, 2017, 225: 64-79

[37] Chou C, Lin C, Chiang T, et al. Coherent event-based surveillance video synopsis using trajectory clustering//Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). Turin, Italy, 2015:1-6

[38] Zhong Anyu, Wang Rui, Zhang hua, et al. Consistency-aware domain adaptive object detection via orthogonal disentangling and contrastive learning. Chinese Journal of Computers, 2023, 46(4):827-842
(钟安雨, 王蕊, 张华等. 基于域内域间语义一致性约束的域自适应 目标检测方法. 计算机学报, 2023, 46(4):827-842)

[39] Yunzuo Zhang, Zhouchen Song, Wenbo Li, et al. Enhancement multi-module network for few-shot leaky cable fixture detection in railway tunnel. Signal Processing: Image Communication, 2023, 113:116943

[40] Vezzani R, Cucchiara R. Video surveillance online repository (visor): An integrated framework. Multimedia Tools and Applications, 2010, 50(2):359-380

[41] Fisher R CJ. Ec funded caviar project ist 2001 37540. Web-

site. http://homepages.inf.ed.ac.uk/rbf/CAVIAR. 2003
[42] U. o. Proc. ICPR'04, Cambridge, Ec funded caviar project ist 2001 37540, Website. URL http://www.nada.kth.se/cvap/actions/. 2004
[43] Zhang Yunzuo, Guo Kaina, Zheng Tingting. Surveillance video synopsis based on spatio-temporal offset. Journal of Electronic Imaging, 2023, 32(01):013013

ZHANG Yun-Zuo, Ph. D., associate professor. His current research interests include computer vision, artificial intelligence, and big data.

ZHU Peng-Fei, M. S. candidate. His current research interests include computer vision and image processing.

Background

Monitoring equipment can record and save events that occur within the monitoring area in real time, helping to combat illegal activities, and playing an important role in maintaining public safety and social stability. At the same time, with the increase in the number of cameras and 24-hour uninterrupted operation, the generated video data is exploding. How to save such a large amount of data and quickly find the desired information from it has become particularly difficult.

Video synopsis technology can help alleviate the above problems by displaying a large amount of video content in a short period. Video synopsis technology first extracts the motion trajectories of all objects then optimizes and rearranges the trajectories to obtain a new start time and space. Finally, the new trajectories are combined with the video background image to obtain a synopsis video. Condensed videos can dynamically display all objects without causing a sense of fragmentation for users while compressing video length to a large extent. Therefore, they have received widespread attention from researchers.

Most video synopsis methods use a single object as the processing unit, which can lead to the loss of interaction behavior between objects and make it difficult for users to understand video content. There are few methods to protect interactive behavior, and various defects lead to low accuracy in judging interactive behavior. How to accurately preserve the interactive behavior between objects in synopsis videos urgently needs to be addressed.

The primary purpose of video synopsis is to compress the video length, but compressing the length can cause collisions between objects and affect the user's viewing experience. To balance collision and compression, many researchers have researched object rearrangement methods. But the results show that most methods are not very effective, and how balancing collision and compression is particularly crucial in video synopsis.

To address the above issues, we propose a surveillance video synopsis method based on multi-objective balance optimization (MBO). Firstly, the number of object interaction frames is compared with the dynamic threshold to preserve the interactivity of the object. The object with interactive behavior is considered as a multi-objective unit and processed uniformly in subsequent steps; Secondly, the collision matrix and insertion position ratio are used to record the depth of the object collision and insertion position, respectively, collisions are divided into two parts: collision degree and collision number; Finally, a dynamic balance compression and collision method is proposed to optimize and rearrange the objects, and the rearranged objects are fused with the background to obtain a synopsis video. Numerous experimental results on multiple datasets have shown that the proposed MBO method in this paper can achieve better compression and collision reduction performance while accurately preserving interactivity.

This work is jointly supported by the National Natural Science Foundation of China (No. 61702347, No. 62027801), the Natural Science Foundation of Hebei Province (No. F2022210007, No. F2017210161), the Science and Technology Project of Hebei Education Department (No. ZD2022100), the Central Guidance on Local Science and Technology Development Fund (No. 226Z0501G). Graduate Innovation Program (YC2023081).