

基于时空灰度序特征的视频片段定位算法^{*}

王方圆^{1,2}, 张树武¹, 李和平¹

¹(中国科学院 自动化研究所 高技术创新中心, 北京 100190)

²(华数集团, 浙江 杭州 310012)

通讯作者: 王方圆, E-mail: fangyuan.wang@ia.ac.cn

摘要: 基于灰度序特征的视频片段定位算法是解决视频片段定位问题的典型算法. 这类算法存在的不足是: 特征的唯一性表示能力不够, 使得在召回率较高的情况下, 定位检索的精度下降得较快; 二次多项式级的时间复杂度使得响应时间过长, 并对查询视频长度敏感. 针对上述两个问题, 提出了一种基于时空灰度序特征的视频片段定位算法, 其关键步骤包括: (1) 在精确定位之前, 通过引入线性时间复杂度的基于时空二值模式直方图特征 (spatio-temporal binary pattern histogram, 简称 STBPH) 的实时过滤算法以及基于二值时间灰度序特征 (binary temporal ordinal measure, 简称 BTOM) 的快速过滤算法, 大幅度减少精确定位阶段需要进行比较的候选视频片段个数; (2) 在精确定位阶段, 通过引入唯一性表示能力更好且保持了较好鲁棒性的时空统一灰度序特征 (joint spatio-temporal ordinal measure, 简称 JSTOM) 进行序列匹配, 显著提高了定位检索的精度. 实验结果表明, 该算法能够快速、准确地进行视频片段定位, 大幅降低了对查询视频长度的敏感度.

关键词: 视频片段定位; 视频序列匹配; 视频拷贝检测; 灰度序特征

中图法分类号: TP391 文献标识码: A

中文引用格式: 王方圆, 张树武, 李和平. 基于时空灰度序特征的视频片段定位算法. 软件学报, 2013, 24(12): 2921-2936. <http://www.jos.org.cn/1000-9825/4415.htm>

英文引用格式: Wang FY, Zhang SW, Li HP. Video clip identification algorithm based on spatio-temporal ordinal measures. Ruan Jian Xue Bao/Journal of Software, 2013, 24(12): 2921-2936 (in Chinese). <http://www.jos.org.cn/1000-9825/4415.htm>

Video Clip Identification Algorithm Based on Spatio-Temporal Ordinal Measures

WANG Fang-Yuan^{1,2}, ZHANG Shu-Wu¹, LI He-Ping¹

¹(High-Tech Innovation Center, Institute of Automation, The Chinese Academy of Sciences, Beijing 100190, China)

²(WaSu Media Group Co. Ltd., Hangzhou 310012, China)

Corresponding author: WANG Fang-Yuan, E-mail: fangyuan.wang@ia.ac.cn

Abstract: Many state-of-the-art video clip identification algorithms are based on ordinal measures. However, they still have two problems: The weak uniqueness of video signature makes the precision decreases quickly as recall increases high enough; Quadratic-time complexity makes the response time too long and sensitive to the length of query video. To address these two problems, this paper proposes a video clip identification algorithm based on spatio-temporal ordinal measures. The key steps are: (1) Before the accurate identification starts, it employs a linear-time complexity real-time filtration method based on spatio-temporal binary pattern histogram (STBPH) and a fast filtration method based on binary temporal ordinal measure (BTOM) to filter out most candidate video clips in target video; (2) During the accurate identification process, it utilizes joint spatio-temporal ordinal measure (JSTOM) which is more unique and robust in improving the precision. Experimental results show that the approach improves the precision significantly and is very efficient and insensitive to the length of query video.

Key words: video clip identification; video sequence matching; video copy detection; ordinal measure

* 基金项目: 国家科技支撑计划(2011BAH16B01, 2011BAH16B02)

收稿时间: 2012-08-06; 修改时间: 2013-01-07; 定稿时间: 2013-03-22

随着视频的制作、传播等技术的快速发展,视频规模呈几何级增长.基于内容的视频浏览、检索、监控等技术能够快速有效地管理、利用丰富的视频资源,已成为近年来的研究热点.

视频片段定位是视频检索^[1-4]的关键技术之一,通常是在一个较长的目标视频中定位出与查询视频重复或近似重复的片段^[5-19],其典型应用包括广告视频片段定位与分析、拷贝视频检测等.比如,多数广告商希望监控自己投放的广告是否按照合同在正确的时段进行播出以及播放次数是否正确;如果视频中没有嵌入水印,传统的基于视频水印的拷贝检测技术则无法应用,此时,需借助视频自身的唯一性进行基于内容的视频拷贝检测^[6].

目前,大规模视频数据库上进行视频片段定位仍是一项极具挑战性的任务^[20].原因在于,拷贝视频可能存在各种复杂的变换,并且大规模视频数据库对实时性要求较高.虽然近年来提出了基于内容的图像或视频拷贝检测方法^[21,22],但多数难以直接扩展到视频片段定位^[17].

TRECVID^[23]举办的基于内容的拷贝视频检测(content based copy detection,简称 CBCD)大大推动了该领域的研究:

- 文献[24]对 TRECVID2011 各竞赛的目标、任务、数据集、评价机制和指标做了简要的介绍;
- 文献[20,25]在 TRECVID2010 CBCD 任务中取得了较高的名次,其检测策略在于融合多种特征和检测器的结果,以应对各种不同的变换;
- 文献[26,27]分别使用关键点轨迹和最大稳定极块作为视频特征,在 TRECVID2008^[23]及 CIVR2007^[28] CBCD 任务中共同使用的 MUSCLE-VCD-2007 数据集^[28]上取得了较好的结果.

然而,TRECVID 任务中目标视频通常不是很长,并且为了应对多种变换,使得参加 TRECVID 的多数算法的时间复杂度往往较高,很难直接扩展到目标视频长度很长情况下的拷贝视频检测应用.

本文主要针对拷贝检测中视觉内容不变而仅存在视频质量变化并且目标视频长度很长的情况下,如何快速准确地完成视频片段定位这一子问题展开研究.

此类视频片段定位算法可分为基于全局特征的方法^[5-17]和基于局部特征的方法^[18,19]两类.其中,基于局部特征的方法虽然能够处理各种复杂的变换,但这类方法的时间和空间复杂度往往很高^[10,17].多数算法采用的是视频中包含的颜色信息、运动向量信息或灰度序信息等全局特征.通常,单个关键帧(或帧间)的全局特征唯一性有限,所以视频特征多以特征序列的方式表示,通过嵌入视频的时间连续性约束,大大提高了特征的唯一性,并将视频片段定位问题转换为序列匹配问题.比如:

- 文献[13]考虑视频关键帧的颜色分布信息,采用 YUV 空间的颜色直方图作为一帧图像的特征;
- 文献[7,11]考虑视频中帧间的运动信息,构造运动向量直方图作为视频的关键帧特征;
- 文献[5]考虑关键帧内灰度平均值的空间分布,提出采用空间灰度序作为关键帧特征;
- 文献[12]提出采用 $\{-1,0,1\}$ 表示灰度序的局部时间约束,并通过加权调和的方式与空间灰度序融合,作为最终的视频特征;
- 文献[6]考虑视频灰度平均值在时间序列上的全局分布,提出采用时间灰度序作为视频特征;
- 文献[10,11]对基于颜色信息、运动信息和灰度序信息的各种特征进行了比较分析,指出基于灰度序信息的全局特征往往能够取得更好的定位效果.

此外,有一些算法采用的是视频中的梯度信息^[14],或是组合利用多种视频特征^[15].

近来,文献[17]提出的视频特征思想是利用颜色信息中 RGB 分量强度平均值相对关系的不变性.不过,其在 MUSCLE-VCD-2007 视频库上的实验结果要比空间灰度序特征^[5]和时间灰度序特征^[6]差.

上述各算法虽然采用的视频特征不同,但多数是以序列匹配的方式完成视频片段的定位.

传统的基于灰度序特征的视频片段定位算法虽然在多数情况下能够取得比其他全局特征更好的定位效果,但仍存在不足之处:

- (1) 无论是空间灰度序特征还是时间灰度序特征(文献[12]提出的时空灰度序特征效果较时间灰度序特征差^[6])的唯一性表示能力在定位检索召回率较高时都明显不够,表现为定位检索的精度比较低.在特征提取过程中,增加关键帧划分步长可以增强特征的唯一性,但会显著减弱特征的鲁棒性^[6,12];

- (2) 与其他全局特征类似,传统的灰度序特征也采用序列匹配的方式进行视频片段定位,序列匹配虽然有助于提高定位的精度,但其时间复杂度为二次多项式级复杂度,从而使定位检索的响应时间过长,并对查询视频长度敏感。

为解决上述问题,本文提出了一种基于时空灰度序特征的视频片段定位算法,其整体流程如图 1 所示。

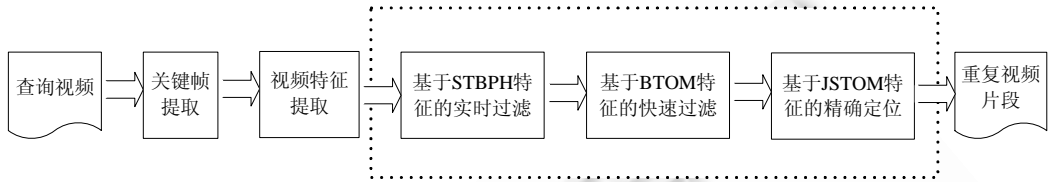


Fig.1 Process of the video clip identification algorithm based on spatio-temporal ordinal measures

图 1 基于时空灰度序特征的视频片段定位算法的处理流程

本文算法主要包括关键帧抽取、视频特征提取、视频片段定位 3 部分,这与多数视频片段定位算法的流程类似,而本文算法的主要贡献在于:

- (1) 提出了唯一性更好的时空统一灰度序特征,在保持鲁棒性的前提下(不增加关键帧划分步长),通过对视频关键帧序列中所有矩形图像区域的灰度平均值进行统一排序,增加了灰度序特征的顺序编码范围,从而增强了特征的唯一性.实验结果证明,在精确定位阶段利用时空统一灰度序进行序列匹配能够取得更高的定位精度;
- (2) 提出了两种简单有效的基于灰度信息的视频特征,即时空二值模式直方图特征和二值时间灰度序特征,并在精确定位之前,先利用这两种特征对目标视频中的候选视频片段进行过滤.其中:基于时空二值模式直方图的过滤算法为线性时间复杂度,平均过滤率约 85%;基于二值时间灰度序的过滤算法进一步使得平均过滤率达到 98%.实验结果证明,该过滤算法大幅减少了定位检索的响应时间,并能够大幅降低对查询视频长度的敏感度。

下面对本文算法的各步骤进行详细叙述。

1 视频关键帧序列抽取

常用的关键帧序列抽取方法有两种:一是基于镜头的关键帧抽取,二是均匀抽取。

- 基于镜头的关键帧抽取,抽取的关键帧个数少,更具有代表性,但同时也破坏了视频的时间连续性;
- 而时间连续性往往能够提高视频片段定位的准确率,并能够保证定位的时间精度^[1,6].因此,视频片段定位算法多采用均匀抽取的方式^[10,11]。

2 视频特征提取

本文算法共提取了 3 种基于灰度信息的视频特征,分别为:

- 二值时间灰度序(binary temporal ordinal measure,简称 BTOM)特征(如图 2(a)~图 2(c)、图 2(e)所示);
- 时空统一灰度序(joint spatio-temporal ordinal measure,简称 JSTOM)特征(如图 2(a)、图 2(b)、图 2(d)、图 2(f)所示);
- 时空二值模式直方图(spatio-temporal binary pattern histogram,简称 STBPH)特征(如图 3 所示)。

其中,STBPH 和 BTOM 用于对候选视频的快速过滤,而 JSTOM 用于对重复视频片段的精确定位.下面我们详细介绍其各自的定义及提取方法。



(a) 一个包含4个关键帧的视频片段,每个关键帧去除顶部和底部20%内容后,均匀划分成2x2个矩形区域

82.47	128.91	83.55	131.77	88.91	124.30	86.48	120.21
77.34	130.62	82.00	122.98	74.82	128.02	74.44	126.31

(b) 计算矩形区域内像素的灰度平均值

0	0
0	1

0	1
1	0

1	1
1	1

5	14
3	15

6	16
4	10

8	11
2	13

7	9
1	12

(c) 二值时间灰度序特征取值

(d) 时空统一灰度序特征取值

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 6 & 8 & 7 \\ 14 & 16 & 11 & 9 \\ 3 & 4 & 2 & 1 \\ 15 & 10 & 13 & 12 \end{bmatrix}$$

(e) 二值时间灰度序的矩阵表示

(f) 时空统一灰度序的矩阵表示

Fig.2 Illustration of binary temporal ordinal measure and joint spatio-temporal ordinal measure

图2 二值时间灰度序和时空统一灰度序示例

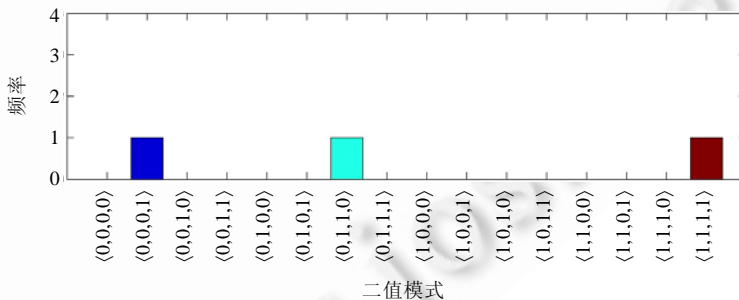


Fig.3 Illustration of spatio-temporal binary pattern histogram

图3 时空二值模式直方图示例

2.1 二值时间灰度序

二值时间灰度序仅使用{0,1}两个数字描述视频关键帧序列中相邻帧间相同位置的矩形图像区域灰度平均值的大小关系.由于其是帧间局部特征,目标视频中某一候选视频片段的二值时间灰度序特征即为目标视频的二值时间灰度序特征序列中对应的子序列.此外,其特征表示非常紧凑,二值化的表示方式便于应用速度较快的位运算.其提取步骤如下:

- (1) 将关键帧序列中的每一个帧图像转化成灰度图像(本文采用 8 位灰度图);
- (2) 将每一个灰度图像划分成 $k \times k$ 个均匀大小的矩形图像块(为避免视频中台标、挂角广告和滚动字幕等因素的干扰,本文去除了图像的顶部和底部各 20% 的图像),如图 2(a)所示;

- (3) 计算各矩形图像块的灰度平均值,如图 2(b)所示;
- (4) 按照公式(1)提取任意相邻两个关键帧之间的二值时间灰度序特征:

$$\text{sgn}(F_i(u,v) - F_{i+1}(u,v)) = \begin{cases} 1, & F_i(u,v) \geq F_{i+1}(u,v) \\ 0, & F_i(u,v) < F_{i+1}(u,v) \end{cases} \quad (1)$$

其中, k 为关键帧划分步长,常用的步长为 2 或 3, $F_i(u,v)$ 表示关键帧序列中第 i 个关键帧第 (u,v) 个矩形图像块的灰度平均值, $u,v \leq k$. 提取的二值时间灰度序特征如图 2(c)所示.

二值时间灰度序特征可以表示成矩阵形式,如图 2(e)所示.需要指出的是,本文二值时间灰度序的计算方式与文献[12]中时域特征(temporal measure)的计算方式有相似之处:文献[12]中利用 $\{-1,0,1\}$ 表示相邻帧对应矩形块灰度平均值的大小关系,并通过融合空间灰度序和时域特征相似度计算最终的相似度;而本文特征较文献[12]中时域特征的区别在于,二值时间灰度序特征是一种二值化的时域特征,二值化的表示不仅有利于位运算的使用,而且使得第 2.3 节中时空二值模式直方图特征以直方图形式表示特征和以增量计算方式构造特征成为可能.

2.2 时空统一灰度序

时空统一灰度序将视频关键帧序列中所有矩形图像区域的灰度平均值按照从小到大的顺序统一排序并编号,以顺序编码作为视频的表示特征.这种灰度平均值排序方式较空间灰度序的帧内排序^[5]和时间灰度序的帧间排序^[6]增加了顺序编码的取值范围,从而在不增加关键帧划分步长的前提下,增强了特征的唯一性.其唯一性分析如下:

如果一种特征的唯一性越好,那么其表示的两段随机视频的特征序列完全相同的概率越小.固定关键帧划分步长 k 和视频长度 M ,并假设两个视频的空间灰度序表示为 S_x^{so}, S_y^{so} , 时间灰度序表示为 S_x^{to}, S_y^{to} , 时空统一灰度序表示为 $S_x^{jsto}, S_y^{jsto}, S_x^{so}, S_y^{so}$ 相互独立, S_x^{to}, S_y^{to} 相互独立, S_x^{jsto}, S_y^{jsto} 相互独立,那么任意矩形图像区域的编码取值服从均匀分布.根据上述假设和各灰度序特征的定义可得:

$$\text{Prob}(S_x^{so} = S_y^{so}) = \prod_{i=1}^M \text{Prob}(S_x^{so}(i) = S_y^{so}(i)) = \left(\frac{1}{(k^2)!} \right)^{2M} = \left(\frac{1}{((k^2)!)^M} \right)^2 \quad (2)$$

$$\text{Prob}(S_x^{to} = S_y^{to}) = \prod_{i=1}^{k^2} \text{Prob}(S_x^{to}(i) = S_y^{to}(i)) = \left(\frac{1}{(M!)^k} \right)^{2k^2} = \left(\frac{1}{(M!)^{k^2}} \right)^2 \quad (3)$$

$$\text{Prob}(S_x^{jsto} = S_y^{jsto}) = \left(\frac{1}{(Mk^2)!} \right)^2 \quad (4)$$

由于阶乘函数较指数函数的复杂度更高,当 $M \gg k$ 时(通常关键帧序列的长度 M 远大于关键帧划分步长 k),由公式(2)~公式(4)可得:

$$\text{Prob}(S_x^{jsto} = S_y^{jsto}) < \text{Prob}(S_x^{to} = S_y^{to}) < \text{Prob}(S_x^{so} = S_y^{so}) \quad (5)$$

因此,时空统一灰度序特征的唯一性较空间灰度序特征和时间灰度序特征更强.上述分析过程中独立性假设条件过于严格,本文通过实验进一步验证了时空统一灰度序的唯一性.

根据时空统一灰度序特征的定义,其提取步骤如下:

- (1) 将关键帧序列中的每一个帧图像转化成灰度图像(本文采用 8 位灰度图);
- (2) 将每一个灰度图像划分成 $k \times k$ 个均匀大小的矩形图像块(为避免视频中台标、挂角广告和滚动字幕等因素的干扰,本文去除了图像的顶部和底部各 20% 的图像),如图 2(a)所示;
- (3) 计算各矩形图像块的灰度平均值,如图 2(b)所示;
- (4) 对所有矩形图像块的灰度平均值按照从小到大的顺序统一排序并编号,图 2(a)所示的视频序列的时空统一灰度序特征如图 2(d)所示.

时空统一灰度序特征也可以表示成矩阵形式,如图 2(f)所示.

2.3 时空二值模式直方图

时空二值模式直方图通过统计视频二值时间灰度序特征序列中某段子序列中包含的时空二值模式构造生成.时空二值模式定义,形式上与表示图像纹理特征的局部二值模式(local binary pattern)^[29]类似,具体是指二值时间灰度序特征中满足一定时空位置约束的二值取值组合.例如,当 $k=2$ 时,满足空间位置约束(按照从左至右、从上至下的矩形图像块排列顺序)的4个二值时间灰度序特征取值的组合是一个二值模式,见公式(6).

$$\left\{ \begin{array}{c} v_{11} \\ v_{12} \\ v_{21} \\ v_{22} \end{array} \right\}, (v_{i,j} \in \{0,1\}, i \in \{1,2\}, j \in \{1,2\}) \quad (6)$$

其中, i,j 为矩形图像块的位置坐标, $v_{i,j}$ 是对应位置矩形图像块的二值时间灰度序特征取值.公式(6)所示二值模式取值空间中共有 $2^{k \times k}$ 个不同的取值.

由于上述二值模式中每一个组成元素是对应位置的二值时间灰度序特征取值,而各元素的排列满足特定的空间顺序,因此称其为时空二值模式.

对于图2中所示视频关键帧序列,利用传统的直方图构造方法统计其二值时间灰度序中公式(6)所示的时空二值模式的分布,结果如图3所示.

由于视频片段定位过程中,查询视频的长度是变化的,使得目标视频中候选视频片段的长度也是变化的,因此,本文对于目标视频中候选视频片段的时空二值模式直方图采用动态构造的方式.

长度为 M 的查询视频,其在长度为 N 的目标视频中有 $N-M+1$ 个候选视频片段,如果对 $N-M+1$ 个候选视频片段均采用传统直方图统计方法,时间复杂度为 $O(MN)$.然而,候选视频片段之间并不是完全独立,时间序列上相邻的候选片段之间绝大多数关键帧是相同的,利用这一特性,本文提出了一种候选视频片段时空二值模式直方图增量式计算方法,步骤如下:

- (1) 采用传统直方图统计方法构造第1个候选视频片段的时空二值模式直方图;
- (2) 对于后续 $N-M$ 个候选视频片段,依次以增量计算方式生成,即对于第 $i(2 \leq i \leq N-M+1)$ 个候选视频片段的时空二值模式直方图,可以在第 $i-1$ 个候选视频片段的时空二值模式直方图基础上,减去目标视频中第 $i-1$ 个帧间二值时间灰度序特征对应的时空二值模式取值,然后加上第 $i+M-1$ 个帧间二值灰度序特征对应的时空二值模式取值生成.

由上述步骤可知,第1个二值模式直方图构造的时间复杂度为 $O(M)$,后续 $N-M$ 个直方图构造总的复杂度为 $O(N-M)$,故该算法整体时间复杂度为 $O(N)$.

3 视频片段定位算法

为了解决传统基于灰度序特征的视频片段定位算法存在的召回率较高时精度不高和定位检索响应时间过长并对查询视频长度敏感的问题,本文在第2节中引入的3种基于灰度信息的视频时空特征的基础上,提出了一种快速准确的视频片段定位算法:

- 该算法首先对目标视频中的候选视频片段进行基于时空二值模式直方图的实时过滤,该层过滤能够在线性时间复杂度内完成;
- 然后,对通过第1层过滤的候选视频片段进行基于二值时间灰度序的快速过滤.由于二值时间灰度序特征较以往灰度序特征简单,因此其速度也很快,由于采用序列匹配的方式,能够确保较高的过滤率;
- 最后,对通过上述两层过滤的候选视频片段进行基于时空统一灰度序的精确定位.

整个定位阶段的流程图如图4所示.

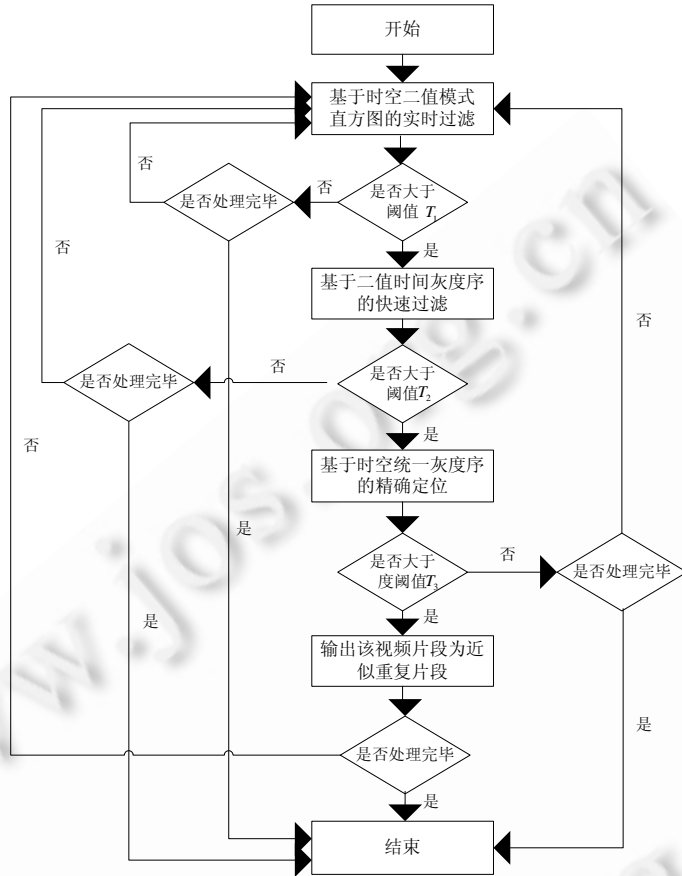


Fig.4 Process of video clip identification

图 4 定位检索阶段处理流程图

3.1 基于时空二值模式直方图的实时过滤

本文采用直方图相交度量^[1]衡量查询视频和候选视频片段之间的时空二值模式直方图相似度,见公式(7).

$$sim(V_q, V_i) = \frac{\sum_w \min(p_q^w, p_i^w)}{\sum_w \max(p_q^w, p_i^w)} \tag{7}$$

其中, V_q 为查询视频, V_i 为目标视频时间序列上第 i 个候选视频片段, p^w 为给定时空二值模式的第 w 种取值的个数.在统计 p^w 时,为减轻时空二值模式的硬量化效应,在第 w 种取值个数的基础上,加上与第 w 种取值仅有 1 位不同的对应时空二值模式取值的个数.

计算得到查询视频和候选视频片段之间的直方图相似度之后,通过与预先设定的过滤阈值 T_1 比较,相似度大于 T_1 的进入下一层继续处理.阈值 T_1 的选取与视频片段时空二值模式直方图的稀疏程度有关,由于时空二值模式的量化效应和直方图表示不考虑位置信息的特点,目标视频中即使是与查询视频片段重复的候选片段在直方图分布比较稀疏的情况下,相似度也不是很高,当然不相似的片段的相似度一般会更低,但对于不同稀疏程度的直方图设置相同的阈值,显然不合适.由第 2.3 节得知,影响直方图分布稀疏程度的因素有时空二值模式取值空间大小和查询视频长度,当取值空间固定时,根据不同的查询视频长度应选取不同的过滤阈值.实验中,我们给出了相应的经验阈值.

由上述描述可知,相似度计算和比较的操作次数为常数次,而由第 2.3 节知,目标视频中所有候选视频片段的时空二值模式直方图构造所需时间复杂度为 $O(N)$,因此,本层过滤算法的时间复杂度亦为 $O(N)$.

3.2 基于二值时间灰度序的快速过滤

利用提取的视频二值时间灰度序特征,计算查询视频片段与通过第 1 层过滤的候选视频片段之间的串匹配相似度,相似度大于给定阈值 T_2 的候选视频片段进入下一层继续处理.序列匹配相似度计算见公式(8).

$$\text{sim}(V_q, V_i) = 1 - \frac{1}{k^2} \sum_{(u,v)=(1,1)}^{(k,k)} \left\{ \frac{1}{M-1} \sum_{j=1}^{M-1} |S_q^{bto}(j) - S_i^{bto}(j)| \right\} \quad (8)$$

其中, $S_q^{bto}(j)$ 和 $S_i^{bto}(j)$ 为查询视频片段和目标视频中第 i 个候选视频片段中第 j 个空间位置为 (u,v) 的二值时间灰度序特征取值.

由于二值时间灰度序特征比传统灰度序特征更加简单,仅使用 $\{0,1\}$ 两个数值表示灰度平均值的大小关系,一方面减少了序列匹配的计算量,另一方面, $\sum_{j=1}^{M-1} |S_q^{bto}(j) - S_i^{bto}(j)|$ 可采用位运算完成,并且候选视频片段的二值时间灰度序特征序列直接对应目标视频的二值时间灰度序特征序列相应的子序列,因此,本层过滤的速度较基于空间灰度序或时间灰度序的序列匹配更快;并且由于采用了序列匹配算法,可以有效地确保最终的过滤效果.

本层阈值 T_2 对查询视频长度不敏感.

3.3 基于时空统一灰度序的精确定位

利用提取的视频时空统一灰度序特征,计算查询视频片段与通过第 2 层过滤的候选视频片段之间的序列匹配相似度,相似度大于给定阈值 T_3 的候选视频片段即认为匹配成功,输出为重复视频片段.序列匹配相似度计算见公式(9).

$$\text{sim}(V_q, V_i) = 1 - \frac{1}{C_{k,M}} \sum_{(u,v)=(1,1)}^{(k,k)} \sum_{j=1}^M |S_q^{jsto}(u,v,j) - S_i^{jsto}(u,v,j)| \quad (9)$$

其中, S_q^{jsto} 和 S_i^{jsto} 为查询视频和第 i 个候选视频片段的时空统一灰度序特征, $S_q^{jsto}(u,v,j)$ 和 $S_i^{jsto}(u,v,j)$ 为对应位置的特征取值. $C_{k,M}$ 是关于 k 和 M 的归一化因子,其值为两个排序完全相反的序列之间的差值,计算如下:

$$C_{k,M} = \sum_{i=1}^{Mk^2} |Mk^2 + 1 - 2i| \quad (10)$$

本层阈值 T_3 对查询视频长度不敏感.

4 实验结果

4.1 基准算法

根据文献[10]对比多种算法的结果,基于灰度序特征的算法在视觉内容相同、仅存在视频质量差异情况下的拷贝检测效果往往优于基于其他全局特征的算法.因此,本文主要以具有代表性的基于空间灰度序的视频片段定位算法(3×3-SOM)^[5]、基于空间灰度序和时域特征融合的视频片段定位算法(2×2-SOM+TM)^[12]和基于时间灰度序的视频片段定位算法(划分步长不同,分为 2×2-TOM,3×3-TOM)^[6]作为基准算法,与本文算法 2×2-STBPH+BTOM+JSTOM 和 3×3-STBPH+BTOM+JSTOM 比较.为此,根据文献[5,6,12],本文分别实现了其对运算法.此外,对于基于其他特征的方法,本文引用其文献中报告的 MUSCLE-VCD-2007 数据集上的评测结果或定位响应时间作为比较参考.

4.2 实验数据集与评价指标

本文算法共在两个视频数据集上进行了对比验证:第 1 个视频数据集为湖南卫视视频数据集,共包含 8 份长度为 16 小时(16 小时 15 分钟 33 秒)的目标视频和 277 个查询视频片段.8 份目标视频中,一份为采集的原始视频,其余 7 份为经过了重编码、分辨率变换、对比度变换或宽高比变换之后得到的拷贝视频.该数据集中 8 份目标视频的视频参数见表 1.而查询视频为从 16 小时原始视频中截取的 277 个不同的广告视频片段.查询视频的时长均介于 5s~30s,共播放 792 次,单个查询视频出现的次数介于 1~20 之间;第 2 个视频数据集为

CIVR2007 和 TRECVID2008 拷贝视频检测任务及文献[17,26,27]共同使用的 MUSCLE-VCD-2007 视频数据集。其中,所有视频格式为 MPEG1,分辨率为 288×352,帧率为 25 帧/秒,共包含 1 个目标数据库(含有 101 个目标视频)、2 个查询视频库(对应两个检测任务 ST1 和 ST2,分别包含 15 个和 3 个查询视频)。查询视频中既有拷贝片段,也有无关片段。拷贝片段又分为无攻击拷贝和有攻击拷贝,攻击方式包括颜色变化、模糊攻击、重编码、画面剪切、加字幕、画面翻转等。由于 ST2 任务中查询视频存在时序编辑变换,而本文算法与文献[5,6,12,17]类似,对时序编辑不鲁棒,故本文仅使用 ST1 任务中的 15 个查询视频进行验证。并且,为了验证本文算法在长视频中的定位检索效率,不失一般性,本文将 101 个目标视频拼接成一个长视频(约为 59 小时)作为最终的目标视频。

Table 1 Video parameter setting for the target videos of HNTV video dataset

表 1 湖南卫视视频数据集中目标视频的视频参数

No.	Name	Format	Resolution	Contrast ratio	Aspect ratio	Frame rate (fps)
A	原始视频	WMV	720×576	Orginal	4:3	25
B	MPG 格式视频	MPG	720×576	Orginal	4:3	25
C	AVI 格式视频	AVI	720×576	Orginal	4:3	25
D	320×240 分辨率视频	WMV	320×240	Orginal	4:3	25
E	1280×720 分辨率视频	WMV	1280×720	Orginal	4:3	25
F	对比度增加 25% 视频	WMV	720×576	Increase 25%	4:3	25
G	对比度减少 25% 视频	WMV	720×576	Decrease 25%	4:3	25
H	16:9 宽高比视频	WMV	720×576	Original	16:9	25

湖南卫视视频数据集中,我们手动标注出所有广告片段播出的起始时间、结束时间和播出次数作为定位检索的基准结果。由于查询视频可能在目标视频中出现多次,本文采用类似信息检索系统的精度(precision)和召回率(recall)衡量单次定位检索的效果^[5,6],用多次定位检索精度的平均值(mean precision)、多次检索召回率的平均值(mean recall)和召回率平均值-精度平均值曲线衡量算法的整体效果。用多次定位检索的平均响应时间(mean response time,定位检索阶段的 CPU 运行时间)来衡量算法的整体时间效率。各评价指标计算公式如下:

$$P_i = (H_i \cap G_i) / H_i \quad (11)$$

$$R_i = (H_i \cap G_i) / G_i \quad (12)$$

$$MP = \frac{1}{n} \sum_{i=1}^n P_i \quad (13)$$

$$MR = \frac{1}{n} \sum_{i=1}^n R_i \quad (14)$$

$$MRT = \frac{1}{n} \sum_{i=1}^n RT_i \quad (15)$$

其中, H_i 为第*i*次定位查询到的视频片段个数, G_i 为第*i*次查询对应人工标注的目标视频中与查询视频片段重复的视频片段总数, P_i 为第*i*次定位查询的精度, R_i 为第*i*次定位查询的召回率, MP 为多次定位查询精度的平均值, MR 为多次定位查询召回率的平均值, RT_i 为第*i*次定位查询的响应时间, MRT 为多次定位查询的平均响应时间, n 为查询视频片段总数。

MUSCLE-VCD-2007 视频数据集提供了基准结果(ST1GroundTruth)^[28]和通用的评价指标 Quality,见公式(16)。此外,我们采用响应时间和平均响应时间衡量算法的时间效率。

$$Quality = \frac{\text{正确检测到的视频片段个数}}{\text{查询视频片段总个数}} \quad (16)$$

4.3 参数设置

本文算法涉及的参数有视频关键帧抽取间隔 f 、关键帧划分步长 k 以及定位检索阶段各层过滤阈值 T_1, T_2, T_3 ,其中,本文选取 $f=5$;根据文献[6,12],增加关键帧划分步长可以增加特征的区分性,但会降低特征的鲁棒性,本文分别给出了 $k=2$ 和 $k=3$ 的情况。由第 2.3 节讨论知,基于时空二值模式直方图的过滤阈值 T_1 的选取与关键帧划分步长和查询视频长度有关,本文在湖南卫视视频数据集中 T_1 的设置见表 2(其中, M 为查询视频长度),在

MUSCLE-VCD 视频集上根据查询视频时长 T_1 简单设置为 0.80(≤ 20 分钟)和 0.90(> 20 分钟),所有实验中,统一设置 $T_2=0.58$;实验中,通过改变阈值 T_3 的取值,可获得不同召回率下的精度取值.对于基准算法中参数的设置,所有的中间参数均采用文献中提供的默认参数,对于最后的过滤阈值的处理与本文算法 T_3 类似,通过变化过滤阈值得到对应算法在不同阈值下的定位效果,更加全面的衡量算法的定位检索效果.

Table 2 Threshold values for T_1

表 2 过滤阈值 T_1 的取值

k	T_1									
	$M < 20$	$20 \leq M < 30$	$30 \leq M < 40$	$40 \leq M < 50$	$50 \leq M < 60$	$60 \leq M < 70$	$70 \leq M < 80$	$80 \leq M < 90$	$90 \leq M < 150$	$150 \leq M$
2	0.25	0.30	0.40	0.45	0.45	0.50	0.50	0.55	0.60	0.70
3	0.20	0.20	0.25	0.35	0.35	0.35	0.40	0.55	0.55	0.70

4.4 实验结果

4.4.1 湖南卫视视频数据集实验结果

本文算法 2×2 -STBPH+BTOM+JSTOM, 3×3 -STBPH+BTOM+JSTOM 和基准算法 3×3 -SOM, 2×2 -SOM+TM, 2×2 -TOM, 3×3 -TOM 在湖南卫视视频数据集 8 份 16 小时目标视频上 277 个查询视频的定位检索结果如图 5、表 3 和图 6 所示.其中,图 5 为各算法在 8 份目标视频中定位检索对应的召回率平均值-精度平均值曲线,横坐标表示召回率平均值,纵坐标表示精度平均值;表 3 为上述各算法在 8 组实验中 277 次定位检索的平均响应时间(s);图 6 为上述各算法定位检索响应时间与查询视频长度的相关性描述(统计 277 个广告视频中,时长约为 5s, 10s, 15s, 30s 的广告视频定位检索的平均响应时间而得).

由图 5 所示实验结果,可以得出如下结论:

- (1) 本文算法 2×2 -STBPH+BTOM+JSTOM 在第 B, F, G, H 组效果显著优于 3×3 -SOM 算法,其他组较 3×3 -SOM 差,原因在于 3×3 -SOM 采用的关键帧划分步长 $k=3$;而本文算法 3×3 -STBPH+BTOM+JSTOM 在各组实验的效果均显著优于 3×3 -SOM 算法;
- (2) 本文算法 2×2 -STBPH+BTOM+JSTOM 在各组实验的效果均显著优于 2×2 -SOM+TM;本文算法 3×3 -STBPH+BTOM+JSTOM 在第 H 组显著低于 2×2 -SOM+TM,其原因在于关键帧划分步长 $k=3$ 时,灰度平均值的空间相对顺序对宽高比变换鲁棒性较差^[6,12],其余组显著优于算法 2×2 -SOM+TM;
- (3) 本文算法 2×2 -STBPH+BTOM+JSTOM 在各组实验的效果均显著优于 2×2 -TOM;
- (4) 本文算法 3×3 -STBPH+BTOM+JSTOM 在第 A, B, C, D, E, G 组效果显著优于算法 3×3 -TOM,在第 F 组效果略低于 3×3 -TOM,而在第 H 组显著低于 3×3 -TOM,其原因在于关键帧划分步长 $k=3$ 时,灰度平均值的空间相对顺序对宽高比变换鲁棒性较差^[6,12];
- (5) 各组实验中,除第 F 组结果次优,本文算法 2×2 -STBPH+BTOM+JSTOM 或 3×3 -STBPH+BTOM+JSTOM 均能够取得最优的结果.

上述结论表明,本文算法采用的时空二值模式直方图特征、二值时间灰度序特征对于轻微视频质量变换具有很好的鲁棒性,而时空统一灰度序特征在保持较好鲁棒性的前提下,进一步提高了特征的区分能力,保证召回率的同时,有效提高了定位检索的精度.此外,实验结果也验证了文献[6,12]中关于关键帧划分步长为 3 时,灰度平均值的空间相对关系对宽高比变换不再鲁棒的结论,以及文献[6]中关于时间灰度序特征较文献[5,12]中使用的特征具有更高区分性的结论.

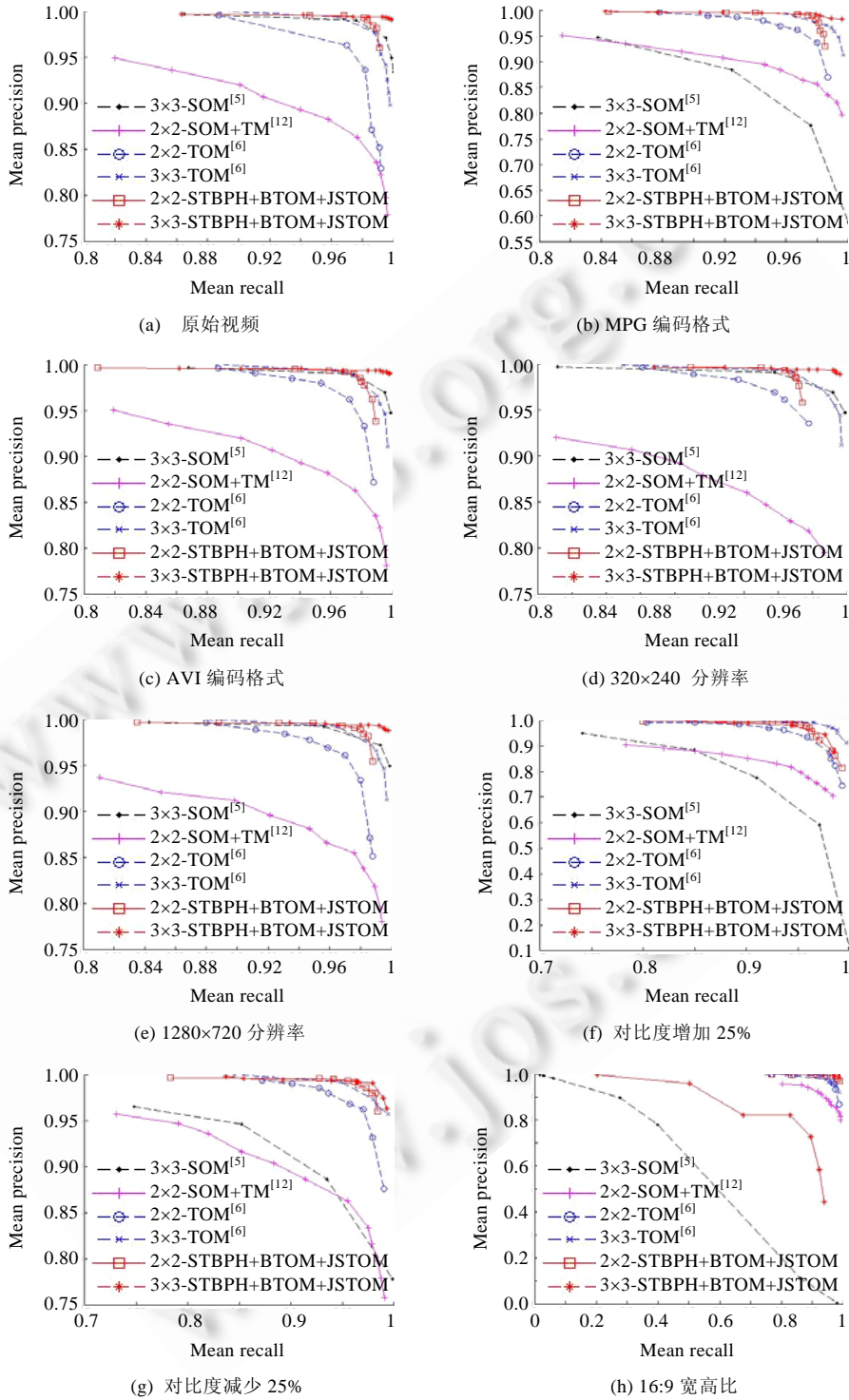


Fig.5 Mean recall-mean precision curves for HNTV video dataset
图 5 湖南卫视视频数据集上定位检索召回率平均值-精度平均值曲线

Table 3 Mean response time of each algorithm on HNTV video dataset

表 3 各算法在湖南卫视视频数据集上的平均响应时间

No.	Mean response time (s)					
	3×3-SOM ^[5]	2×2-SOM+TM ^[12]	2×2-TOM ^[6]	3×3-TOM ^[6]	2×2-STBPH+BTOM+JSTOM	3×3-STBPH+BTOM+JSTOM
A	3.113	1.063	3.827	9.017	1.024	1.085
B	3.278	1.043	3.835	8.944	1.031	1.093
C	3.282	1.102	3.854	9.080	1.102	1.107
D	3.261	1.097	3.867	9.230	1.021	1.053
E	3.382	1.028	3.902	9.393	1.078	1.028
F	3.218	1.075	3.895	9.267	1.053	1.102
G	3.129	1.059	3.879	9.108	1.109	1.097
H	3.202	1.046	3.898	9.275	1.082	1.109
Avg	3.233	1.064	3.870	9.164	1.063	1.084

Method	Mean response time (s)			
	Q =5	Q =10	Q =15	Q =30
3×3-SOM ^[5]	1.050	1.063	3.827	7.034
2×2-SOM+TM ^[12]	0.370	0.745	1.129	2.355
2×2-TOM ^[6]	1.354	2.690	4.125	8.768
3×3-TOM ^[6]	2.969	6.027	9.308	19.727
2×2-STBPH+BTOM+JSTOM	1.064	0.816	1.076	1.164
3×3-STBPH+BTOM+JSTOM	1.531	0.670	0.974	0.585

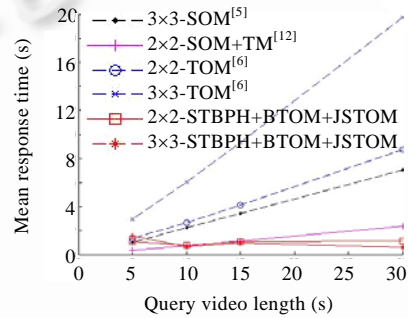


Fig.6 Response time vs query video length on HNTV video dataset

图 6 湖南卫视视频数据集上定位检索响应时间与查询视频长度的关系

由表 3 知,本文算法无论 2×2-STBPH+BTOM+JSTOM 还是 3×3-STBPH+BTOM+JSTOM,与基于空间灰度序的 3×3-SOM 和基于时间灰度序的 2×2-TOM,3×3-TOM 相比,都大幅度地降低了定位检索的平均响应时间.其中,较 3×3-TOM 约减少了 90%的耗时,较 2×2-TOM 和 SOM 约减少了 60%的耗时.

由图 6 知,基于空间灰度序的 3×3-SOM 算法、2×2-SOM+TM 算法和基于时间灰度序的 2×2-TOM,3×3-TOM 算法的定位查询响应时间基本与查询视频长度呈线性关系,而本文算法 2×2-STBPH+BTOM+STUOM 和 3×3-STBPH+BTOM+STUOM 的响应时间与查询视频长度基本无关(由于过滤率不同,时长较长的查询视频的响应时间可能耗时更少).

本文算法之所以能够大幅降低定位检索的响应时间,并使其对查询视频长度基本不敏感,表明本文算法过滤阶段采用的基于时空二值模式直方图的实时过滤和基于二值时间灰度序的快速过滤不但速度快而且过滤效果很好,其实验中 277 次定位检索的平均过滤率统计结果见表 4.

Table 4 Mean filtration rate

表 4 平均过滤率

	Filtration based on STBPH (%)	Filtration based on STBPH+BTOM (%)
2×2-STBPH+BTOM+JSTOM	84.7	98.7
3×3-STBPH+BTOM+JSTOM	86.9	99.1

上述实验的测试机器为一台 4G 内存 3.2GHz 主频的台式机.文献[5]没有报告平均响应时间.文献[6]在 2.0GHz 奔腾 4 机器上对长度为 50 帧的查询视频在长度为 285 535 帧目标视频中,定位查询时间为 9s(本文 16 小时视频的关键帧个数为 292 665).文献[3]在 3 小时 11 分钟视频中,查找广告片段最小响应时间为 97s.文献[8,17]响应时间随查询视频长度增加而线性增加.文献[16]对于 MPEG1 格式的视频,对长度为 10s 的查询视频速度较快,但不适于其他长度的查询视频或其他格式的目标视频.文献[7]关键帧抽取间隔为 25,在 24 小时视频中定位查询长度为 42s 音乐视频的最小响应时间为 29s.

4.4.2 MUSCLE-VCD-2007 视频数据集实验结果

以 MUSCLE-VCD-2007 数据集中 ST1 任务中 15 个视频为查询视频,以 101 个目标视频拼接得到的时长约为 59 小时的长视频为目标视频,本文算法 2×2-STBPH+BTOM+JSTOM 和 3×3-STBPH+BTOM+JSTOM 与基准算法 3×3-SOM,2×2-SOM+TM,2×2-TOM 和 3×3-TOM 的定位检索结果见表 5,响应时间见表 6。

Table 5 Results for the ST1 task on MUSCLE-VCD-2007
表 5 MUSCLE-VCD-2007 视频数据集执行 ST1 任务的结果

Method	Quality				
	Thres=0.50	Thres=0.60	Thres=0.70	Thres=0.80	Thres=0.90
3×3-SOM ^[5]	11/15	11/15	11/15	10/15	9/15
2×2-SOM+TM ^[12]	^	^	^	13/15	11/15
2×2-TOM ^[6]	^	15/15	12/15	11/15	11/15
3×3-TOM ^[6]	^	15/15	14/15	11/15	11/15
2×2-STBPH+BTOM+JSTOM	15/15	14/15	12/15	11/15	11/15
3×3-STBPH+BTOM+JSTOM	15/15	14/15	13/15	11/15	11/15

表 5 中,Thres 表示对应算法采用的相似度过滤阈值,“^”表示该阈值下对应算法的检测结果出现了误判(返回了多个检测结果)。由表中的实验结果分析可知,本文算法在该数据集上的拷贝检测效果与 2×2-TOM 算法和 3×3-TOM 算法^[6]基本相当,而优于 3×3-SOM 算法^[5]或 2×2-SOM+TM 算法^[12]。由于该数据集上查询视频与拷贝视频之间的视频质量差异较湖南卫视视频数据集更大,表明本文算法具有较好的鲁棒性。同时,由于该数据集中查询视频时长较长,目标视频中每一个查询视频只有一个或没有副本,本文算法采用的时空统一灰度序特征区分性强的优势没有在湖南卫视视频数据集上短时查询视频时明显。此外,文献[17]中报告的在该数据集上的评测结果为 11/15,低于本文算法;CIVR 2007 视频拷贝竞赛的结果^[28]中最优队伍得分为 13/15,与本文算法相当。

Table 6 Response time for the 15 queries in ST1 task
表 6 ST1 任务中 15 个查询视频定位检索的响应时间

ST1 Query	Response time (s)					
	3×3-SOM ^[5]	2×2-SOM+TM ^[12]	2×2-TOM ^[6]	3×3-TOM ^[6]	2×2-STBPH+BTOM+JSTOM	3×3-STBPH+BTOM+JSTOM
1	371.603	125.831	499.096	1 184.094	43.810	25.994
2	290.629	99.511	412.380	978.955	1.852	1.765
3	312.623	108.163	448.859	1 049.713	10.346	5.344
4	276.965	95.896	386.079	907.699	26.761	11.367
5	390.296	134.389	552.097	1 293.128	33.805	74..522
6	303.371	104.478	429.400	1 013.759	15.621	2.583
7	344.454	118.394	498.276	1 170.736	17.070	1.937
8	401.599	138.542	569.827	1 336.213	44.560	28.024
9	462.067	160.336	654.974	1 540.940	37.428	63.543
10	587.915	199.837	818.956	1 927.558	33.866	109.327
11	589.800	202.218	836.284	1 971.017	35.285	15.750
12	736.623	250.176	1 040.646	2 449.724	12.366	16.763
13	874.927	301.711	1 235.601	2 933.716	24.432	19.029
14	1 333.884	450.769	1 913.355	4 896.642	24.227	12.525
15	2 315.944	839.148	3 971.115	9 733.988	44.609	125.486
Avg.	639.513	221.960	951.130	2 292.525	27.069	31.388

表 6 的实验结果表明:

- 基于时间灰度序的定位算法^[6]定位响应时间最长,15 个查询的平均响应时间分别为 951.130s 和 2292.525s;
- 而文献[5]采用的 3×3-SOM 算法的平均定位响应时间为 639.513s;
- 文献[12]采用的 2×2-SOM+TM 的算法的平均定位响应时间为 221.960s;
- 而本文算法 2×2-STBPH+BTOM+JSTOM 和 3×3-STBPH+BTOM+JSTOM 的平均定位响应时间分别是 27.069s 和 31.388s。

即使是与基准算法中定位响应时间最小的 2×2-SOM+TM 的 221.960 相比,本文算法依然大约减少 85% 以上的时间开销。此外,基准算法 3×3-SOM,2×2-SOM+TM,2×2-TOM,3×3-TOM 的定位响应时间基本随查询视频长

度增加而线性增加(第 14、第 15 查询视频的时长约为 23 分钟和 43 分钟,其余视频介于 5 分钟~17 分钟之间),而本文算法的定位响应时间与查询视频长度的敏感性显然已大幅减少.CIVR 2007 视频拷贝竞赛的结果^[28]中,耗时最少的队伍为 14 分钟,而本文算法总耗时约为 6.8 分钟和 7.8 分钟(由于本文实验对 101 个目标视频进行了拼接,其耗时理论上大于各目标视频单独检测耗时之和).上述结果进一步验证了本文算法引入的基于时空二值模式直方图特征的实时过滤和基于二值时间灰度序特征快速过滤的有效性,表明本文算法能够大幅减少定位检索的响应时间及其对查询视频长度的敏感性.

4.5 实验分析与讨论

上述实验结果表明:本文算法能够快速、准确地完成视频片段定位任务,并大幅降低了定位检索响应时间对查询视频长度的敏感度.其原因在于:本文算法精确定位阶段采用的时空统一灰度序特征在保持较好的鲁棒性的前提下,具有比空间灰度序和时间灰度序更好的唯一性;而本文算法采用的基于时空二值模式直方图的实时过滤和基于二值时间灰度序的快速过滤,极大地减少了精确定位阶段需要进行的序列匹配次数.

显然,本文算法提出的过滤方法并不依赖于精确定位阶段采用的算法,因此其也可以用于加快基于空间灰度序特征或时间灰度序特征的视频片段定位算法的定位检索速度.

本文算法主要针对视觉内容一致、仅存在视频质量差异情况下的视频片段定位问题,实验结果验证了本文算法在此类情况下的有效性.然而,拷贝视频检测还包括诸如局部内容变换、视频时序编辑等情况,本文算法还没有考虑这些情况,这是我们下一步研究的重点.

5 结 论

视频片段定位算法是近年来被广泛关注的研究热点.本文针对现有算法存在的定位精度不高和响应速度过慢且对查询视频长度敏感两个主要问题,提出了一种基于时空灰度序特征的视频片段定位算法.该算法通过在精确定位阶段引入一种唯一性更好的时空统一灰度序特征,显著提高了定位检索的精度;并通过在精确定位之前引入基于时空二值模式直方图特征的实时过滤和基于二值时间灰度序特征快速过滤,在近似线性时间复杂度内对候选视频片段进行快速有效的过滤,大大减少了精确定位阶段需要进行的序列匹配次数,使得定位检索的平均响应时间大幅减少并与查询视频长度基本无关.在湖南卫视视频数据集和 MUSCLE-VCD-2007 视频数据集上进行的两组实验,验证了本文算法能够快速、准确地完成视频片段定位任务,并大幅降低定位检索响应时间对查询视频长度的敏感度.因此,本文算法具有较大的应用价值.

References:

- [1] Shang LF, Yang LJ, Wang F, Chan KP, Hua XS. Real-Time large scale near-duplicate web video retrieval. In: Proc. of the Int'l Conf. on ACM MM. 2010. 531–540. [doi: 10.1145/1873951.1874021]
- [2] Wu X, Ngo CW, Hauptmann AG. Practical elimination of near-duplicates from Web video search. In: Proc. of the Int'l Conf. on ACM MM. 2007. 218–227. [doi: 10.1145/1291233.1291280]
- [3] Peng YX, Ngo CW, Dong QJ, Guo ZM, Xiao JG. An approach for video retrieval by video clip. Ruan Jian Xue Bao/Journal of Software, 2003,14(8):1409–1417 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/1409.htm>
- [4] Zhuang YT, Liu XM, Wu Y, Pan YH. A new approach to retrieve video by example video clip. Chinese Journal of Computers, 2000,23(3):300–305 (in Chinese with English abstract).
- [5] Hua XS, Chen X, Zhang HJ. Robust video signature based on ordinal measure. In: Proc. of the Int'l Conf. on ICIP. 2004. 685–688. [doi: 10.1109/ICIP.2004.1418847]
- [6] Chen L, Stentiford FWM. Video sequence matching based on temporal ordinal measurement. Pattern Recognition Letters, 2008, 29(13):1824–1831. [doi: 10.1016/j.patrec.2008.05.015]
- [7] Schoeffmann K, Boeszoermenyi L. Video sequence identification in TV broadcasts. In: Proc. of the Int'l Conf. on MMM. 2011. 129–139. [doi: 10.1007/978-3-642-17832-0_13]

- [8] Shen HT, Shao J, Huang Z, Zhou XF. Effective and efficient query processing for video subsequence identification. *IEEE Trans. on Knowledge and Data Engineering*, 2009,21(3):321–334. [doi: 10.1109/TKDE.2008.168]
- [9] Yeh MC, Cheng KT. Video copy detection by fast sequence matching. In: *Proc. of the Conf. on CIVR*. 2009. 633–636. [doi: 10.1145/1646396.1646449]
- [10] Law-To J, Chen L, Joly A, Laptev I, Buisson O, Gouet-Brunet V, Boujemaa N, Stentiford F. Video copy detection: A comparative study. In: *Matsuoka S, Yonezawa A, eds. Proc. of the Conf. on CIVR*. 2007. 371–378. [doi: 10.1145/1282280.1282336]
- [11] Hampapur A, Hyun K, Bolle R. Comparison of sequence matching techniques for video copy detection. In: *Proc. of the Conf. on SPIE*. 2001. 194–201.
- [12] Kim C, Vasudev B. Spatiotemporal sequence matching for efficient video copy detection. *IEEE Trans. on Circuits and Systems for Video Technology*, 2005,15(1):127–132. [doi: 10.1109/TCSVT.2004.836751]
- [13] Naphade MR, Yeung MM, Yeo BL. A novel scheme for fast and efficient video sequence matching using compact signatures. In: *Proc. of the Conf. on SPIE*. 2000. 564–572.
- [14] Dohring I, Lienhart R. Mining TV broadcasts for recurring video sequences. In: *Proc. of the Conf. on CIVR*. 2009. 1–8. [doi: 10.1145/1646396.1646432]
- [15] Liu N, Zhao Y, Zhu ZF. Commercial recognition in TV streams using coarse-to-fine matching strategy. In: *Proc. of the Conf. on PCM*. 2010. 296–307. [doi: 10.1007/978-3-642-15702-8_27]
- [16] Yuan JS, Duan LY, Ranganath S, Xu CS. Fast and robust short video clip search for copy detection. In: *Proc. of the Conf. on PCM*. 2004. 479–488. [doi: 10.1007/978-3-540-30542-2_59]
- [17] Lei YQ, Luo WQ, Wang YG, Huang JW. Video sequence matching based on the invariance of color correlation. *IEEE Trans. on Circuits and Systems for Video Technology*, 2012,22(7):1–12. [doi: 10.1109/TCSVT.2012.2204811]
- [18] Joly A, Buisson O, Frelicot C. Content-Based copy retrieval using distortion-based probabilistic similarity search. *IEEE Trans. on Multimedia*, 2007,9(2):293–306. [doi: 10.1109/TMM.2006.886278]
- [19] Law TJ, Gouet B, Buisson O, Boujemaa N. Local behaviors labelling for content-based video copy detection. In: *Proc. of the Int'l Conf. on ICPR*. 2006. 232–235. [doi: 10.1109/ICPR.2006.767]
- [20] Tian YH, Huang TJ, Gao W. Multimodal video copy detection using multi-detectors fusion. *IEEE COMSOC MMTC E-Letter*, 2012, 5(7):3–6.
- [21] Hsiao JH, Chen CS, Chien LF, Chen MS. A new approach to image copy detection based on extended feature sets. *IEEE Trans. on Image Process*, 2007,16(8):2069–2079. [doi: 10.1109/TIP.2007.900099]
- [22] Bober M, Brasnett P. MPEG-7 visual signature tools. In: *Proc. of the Conf. on ICME*. 2009. 1540–1543. [doi: 10.1109/ICME.2009.5202798]
- [23] <http://trecvid.nist.gov>
- [24] Paul O, George A, Jon F, Brian A, Martial M, Alan F, Wessel K, Georges Q. TRECVID2011—An overview of the goals, tasks, data, evaluation mechanisms, and metrics. In: *Proc. of the TRECVID 2011*. 2011.
- [25] Guo X, Chen YB, Liu W, Mao YH, Zhang H, Zhou K, Wang LG, Hua Y, Zhao ZC, Zhao YY, Cai A. BUPT-MCPRL at TRECVID 2010. In: *Proc. of the TRECVID 2010*. 2010.
- [26] Wu X, Li JT, Tang S, Guo JB. Video copy detection based on spatio-temporal trajectory behavior feature. *Journal of Computer Research and Development*, 2010,47(11):1871–1877 (in Chinese with English abstract).
- [27] Zhang ZJ, Zou, JH. Video copy detection based on spatio-temporal feature. *Pattern Recognition and Artificial Intelligence*, 2012, 25(2):230–236 (in Chinese with English abstract).
- [28] Law TJ, Gouet B, Buisson O, Boujemaa N. Muscle-VCD-2007: A live benchmark for video copy detection. <http://www.rocq.inria.fr/imedia/civr-bench/>
- [29] Ojala T, Pietikoinen M, Harwood D. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 1996,29(1):51–59. [doi: 10.1016/0031-3203(95)00067-4]

附中文参考文献:

- [3] 彭宇新,董庆杰.一种通过视频片段进行视频检索的方法.软件学报,2003,14(8):1409-1417. <http://www.jos.org.cn/1000-9825/14/1409.htm>
- [4] 庄越挺,刘小明,吴翌,潘云鹤.通过例子视频进行视频检索的新方法.计算机学报,2000,23(3):300-305.
- [26] 吴潇,李锦涛,唐胜,郭俊波.基于时空轨迹行为特征的视频拷贝检测方法.计算机研究与发展,2010,47(11):1871-1877.
- [27] 张志杰,邹建华.基于时空域特征的视频拷贝检测方法.模式识别与人工智能,2012,25(2):230-236.



王方圆(1984—),男,山东滕州人,博士生,主要研究领域为多媒体检索技术,生物信息学.

E-mail: fangyuan.wang@ia.ac.cn



李和平(1978—),男,博士,高级工程师,主要研究领域为模式识别,多媒体处理.

E-mail: heping.li@ia.ac.cn



张树武(1964—),男,博士,研究员,博士生导师,主要研究领域为网络文化,数字版权服务与管理,现代网络服务与监管技术.

E-mail: shuwu.zhang@ia.ac.cn