E-mail: jig@irsa.ac.cn
Website: www.cjig.cn
Tel: 010-64807995

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
©中国图象图形学报版权所有

中图法分类号: TP391.4　　**文献标识码:** A　　**文章编号:** 1006-8961(2014)12-1685-11

论文引用格式: Wang J, Jiang X H, Sun T F. Review of video abstraction [J]. Journal of Image and Graphics, 2014, 19(12): 1685-1695. [王娟, 蒋兴浩, 孙锬锋. 视频摘要技术综述[J]. 中国图象图形学报, 2014, 19(12): 1685-1695.] [DOI: 10.11834/jig.20141201]

视频摘要技术综述

王娟[1], 蒋兴浩[1,2], 孙锬锋[1,2]

1. 上海交通大学电子信息与电气工程学院, 上海　200240; 2. 信息内容分析技术国家工程实验室, 上海　200240

摘　要: **目的**　类似于文本摘要, 视频摘要是对视频内容的总结。为了合理地评估视频摘要领域的研究进展, 正确导向视频摘要的继续研究, 本文归纳总结视频摘要技术的主要研究方法和显著性成果, 对视频摘要技术进行综述。**方法**　依据视频摘要的两个主要生成步骤: 视频内容分析和摘要生成分别介绍视频摘要的主要研究方法。同时, 分析了近5年视频摘要领域的研究状况, 对视频摘要发展的新趋势: 实时视频摘要和多视角视频摘要进行了阐述。最后, 还对视频摘要的评价系统进行了分类总结。**结果**　对视频摘要进行综述, 对摘要中的语义获取难题提出了2种指导性建议。并依据分析结果, 展望了视频摘要技术未来的发展方向。**结论**　视频摘要技术作为视频内容理解的重要组成部分, 有较大研究价值。而目前, 视频摘要在视频语义表达和摘要评价系统方面并不精确完善, 还需进一步的深入研究。

关键词: 视频内容分析; 摘要生成; 实时视频摘要; 多视角视频摘要; 视频语义获取

Review of video abstraction

Wang Juan[1], Jiang Xinghao[1,2], Sun Tanfeng[1,2]

1. *School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai*　200240, *China*;
2. *National Engineering Lab on Information Content Analysis Techniques, Shanghai*　200240, *China*

Abstract: **Objective** This paper summarized the significant methods and techniques for video abstraction and systematically presented all aspects of this technology to properly evaluate the research progress of video abstraction and to provide a correct orientation for deep research in this field. **Method** This paper introduced the main research techniques in detail and discussed the typical algorithms based on two main abstraction generation steps, namely, video content analysis and abstraction generation. Specifically, this paper analyzed the progress of video abstraction in the past five years and described the new demand in this field, that is, real-time and multi-view video abstraction. The evaluation benchmark of video abstraction was also analyzed. **Result** This paper summarized the video abstraction techniques and described the current research situation in this field. Two solutions to address the difficulty in obtaining video semantics were also proposed. Finally, the paper discussed the future development direction of video summary technology. **Conclusion** As an important part of understanding video content, video abstraction has great research value. To date, video abstraction still has defects that need to be addressed, particularly in the areas of video semantic presentation and abstraction evaluation benchmark.

Key words: video content analysis; abstraction generation; real-time video abstraction; multi-view video abstraction; video semantic acquisition

收稿日期: 2014-06-24; **修回日期:** 2014-07-21

基金项目: 国家自然科学基金项目(61272249, 61272439); 软件工程国家实验室开放研究基金项目(SKLSE2012-09-12); 国家教委博士点专项基金项目(20120073110053)

第一作者简介: 王娟(1991—　), 女, 上海交通大学电子信息与电气工程学院信息安全系在读硕士研究生, 主要研究方向为视频摘要, 视频内容理解。E-mail: wangjuanxidian@163.com

通信作者: 蒋兴浩, E-mail: xhjiang@sjtu.edu.cn

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 19, No. 12, Dec. 2014

0 引 言

近年来,随着互联网和多媒体技术的迅速发展,多媒体信息数据骤增。数字视频,作为主要的多媒体信息载体,广泛地应用于生活的方方面面。大量的视频一方面给生活带来了便利,以更丰富的形式获取信息,而另一方面,却给视频存储、传输、归档和检索带来巨大压力。

在这样的背景下,视频摘要技术应运而生。类似于文本摘要,视频摘要是对原始视频内容的总结,通过分析原始视频数据流,从中选取有意义的视频内容来组成紧凑的摘要。视频摘要不仅可以结合视频标注技术服务于视频检索;同时,它还能作为独立的产品应用,如电影预告片,方便生活。

视频摘要技术至 20 世纪 90 年代提出以来,作为一个研究热点和难点,得到了国内外众多研究团队的持续关注。目前,国外视频摘要领域主要的研究机构(https://webofknowledge.com/)有:佐治亚理工学院、大阪大学、卡内基梅隆大学、IBM 等。而国内的中国科学院、微软亚洲研究院、清华大学、浙江大学和国防科技大学等在摘要领域都有突出贡献。

针对不同视频内容,学者们提出了一些较为成熟的视频摘要系统。如:

1) 在新闻视频处理方面,MITRE 公司开发了广播新闻编辑和浏览系统 BNE(broadcast news editor)和 BNN(broadcast news navigator)[1]。前者用于捕获、分析、注释、分割以及存储新闻的文本、音频和视频数据,后者则是在 BNE 的基础上提供一个基于 Web 的浏览系统。系统综合利用了多模态处理技术,采用机器学习的方法从标准的新闻数据集合中学习新闻结构并检测某些特殊事件。

2) 1994 年曼海姆大学 Efelsberg 博士主持了 MoCA(movie content analysis)项目[2],主要针对电影视频进行分析。该项目在视频切分、视频流中文字的检测、定位与识别、人脸检测和视频摘要等方面都进行了研究。此外,在语音识别与自然语言理解等方面也做了大量的研究,并且对视频流中的广告检测和识别以及对不同电影类型的识别等视频的分类方面的问题,视频流中运动对象的分割和识别问题也进行了研究和探索。

3) 对于运动视频,哥伦比亚大学的新媒体技术中心设计的 VideoQ[3] 可以判断足球视频是否处于比赛状态,并能完成慢镜头检测、声音事件定位最后实现交互式浏览。另外,阿姆斯特丹大学开发的基于 Web 的足球视频分析系统——Goalgle[4] 足球视频搜索引擎,不仅可以方便地找到红牌、黄牌、进球和换人等事件,也可以对个别特定的足球运动员进行搜索操作。

4) 在监控视频方面,以色列耶路撒冷希伯来大学开发的 BriefCam(http://briefcam.com/)系统通过标定同一场景下不同对象出现的时间来浓缩监控视频的内容,将数 10 小时的视频缩短为几分钟,提高视频浏览效率。该系统已广泛地运用于我国安防事业。

除此之外,FX PaloAlto 实验室的 VideoManga[5] 系统、国防科技大学的 Videowser[6] 系统和卡内基梅隆大学的 Infor-media 系统(http://www.informedia.cs.cmu.edu)等都是视频摘要领域较为成熟的系统,为大家所熟知。

就具体的应用场景,视频摘要技术可以用于实现大视频库中关键内容的提炼,完成视频库的组织和整合。如利用视频摘要来组织、管理和整合图书馆视频库等;或者通过对相似主题的视频库提取摘要信息并进行匹配,从而完成主题事件的关联。如对大型活动不同场所安防视频的摘要信息进行匹配,可以对异常事件进行检测、搜索和回溯。同时,视频摘要技术还具有较大的商业应用前景。无论是电影视频预告片,或者是个人偏好的运动视频集锦都可作为成熟的产品应用。

从以上分析可以看出,视频摘要作为视频内容分析的一个重要组成部分,本身具有较大的研究价值。基于此,本文通过参阅大量文献对 20 多年视频摘要领域的研究成果进行总结。

1 视频摘要的概念与分类

理论上,视频是指由一系列静态图像按时间顺序或空间分布规则组合得到的图像集,可多角度表达语义信息。然而,视频摘要不仅对原始视频进行分析,还综合考虑了伴随着视频有意义的音频流和文本流等多媒体信息,进行语义理解,并对视频流或多媒体流进行摘要。基于此,本文中讨论的视频不仅包含典型的由图像集组成的视频,还包括伴随着视频有意义的音频流和文本流等多媒体信息。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

视频摘要是指利用计算机技术分析视频结构、理解视频内容,并从原始的多媒体数据中选取具有代表性的、有意义的部分,将它们以某种方式组合并生成紧凑的、用户可读的原始视频的缩略。一个好的视频摘要系统可在最少的时间使用户从原始视频序列中获得最大的信息量。

依据最终的呈现形式,视频摘要可分为静态视频摘要和动态视频摘要,如图 1 所示。

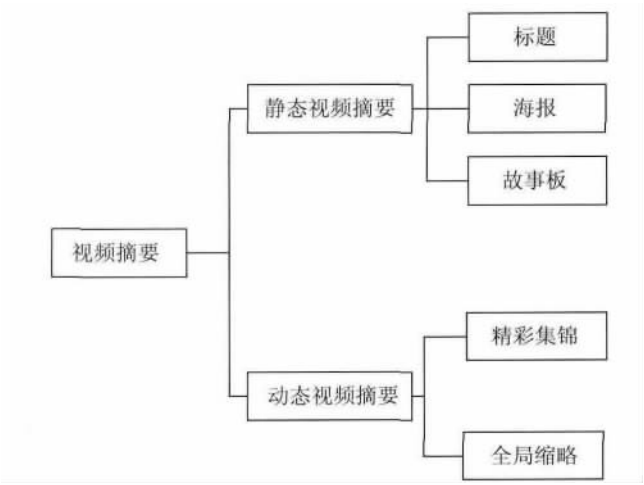

图 1　视频摘要分类

Fig. 1　The classification of video abstraction

静态视频摘要,又称为关键帧集,是由原始视频中具有代表性的图像帧组成,以直接、分层或缩放的方式进行组合。层次的视频摘要可以用于快速导航和索引视频帧,有利于视频检索。而以缩放的形式显示提取出的关键帧,如美国的 FX Palo Alto 实验室提出的漫画书式的会议视频摘要 VideoManga[5],能提高用户的观赏愉悦度,增强摘要的可读性。

静态视频摘要又可以分为标题、海报和故事板。标题是对视频内容的一段简短文字描述,是最简单的静态视频摘要。海报又称为视频代表帧,是从视频中抽出的能够代表视频内容的图像帧。而故事板是从视频中抽取的一段图像序列,按照时间顺序或者重要程度进行组合。在实际应用时,故事板常常结合文本信息来共同表征视频内容信息。

静态视频摘要的生成方法多种多样,不同算法在效率和结果上分别有所偏倚。最简单的静态视频摘要算法莫过于以固定时间间隔对视频序列进行抽样。而更进一步,通过对原始视频进行镜头分割,选取镜头的首尾帧或中间帧作为关键帧来合成摘要。随后,更多的研究工作集中于分析伴随着视频出现的多媒体信息流,如有意义的音视频和文本信息等,并依据相应准则选取出能代表视频语义的关键帧集,这种算法虽然复杂度较大,但是最终得到的摘要更加符合用户的观赏习惯。

动态视频摘要是从原始视频中选取可表达语义内容的视频片段拼接编辑得到。它本身也是一段视频,但比原视频要短得多。动态视频摘要可分为精彩集锦和全局缩略视频。精彩集锦一般由原始视频中最精彩的部分组成。如,足球比赛中的进球集锦。而全局缩略视频是对整个视频内容的概括,它通过对整个时间轴上的视频片段进行组合,使用户对视频内容进行全局掌握。

动态视频摘要生成的一般步骤为视频段分割、视频段选取和视频段的整合。视频段的分割主要是将原始视频依据视觉、音频或者文本上的特性分割成独立单元。而视频段的选取需要综合考虑音视频数据流的同步、视频段的重要度和摘要时长约束等因素。在电影和访谈视频中,音视频的同步要求较高,音频信息必须和视觉内容一一对应。视频段的重要度决定了优先权,而摘要时长约束决定了视频段的数目。最后,视频段的整合应在保证时域顺序的前提下,合理地使用各种编辑手段来最大限度地保证摘要的连贯性。

总体说来,静态视频摘要主要分析视觉内容,不考虑音频信息,它的构建与表现都相对简单,往往可灵活地组织以用于浏览和索引。动态视频摘要综合考虑多媒体信息流,通常含有丰富的音频、动作甚至文本信息,可更加清晰地表达原始视频的内容,更具有娱乐性和观赏性。

2　视频摘要技术框架

要想获取视频摘要,首先需要对原始视频的内容进行理解,在此基础上再依据一定准则提取摘要。图 2 给出了视频摘要生成的大致流程。由于视频类型和生成摘要类型的不同,视频摘要技术在具体的实施步骤上有一定的差异。但是总体上,视频摘要生成主要步骤可以归纳为视频内容分析和摘要生成两部分。

2.1　视频内容分析

视频内容分析作为视频摘要技术研究的第 1 步,主要是利用计算机技术对视频内、外部资源进行分析处理,从而获取视频语义信息。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 19, No. 12, Dec. 2014

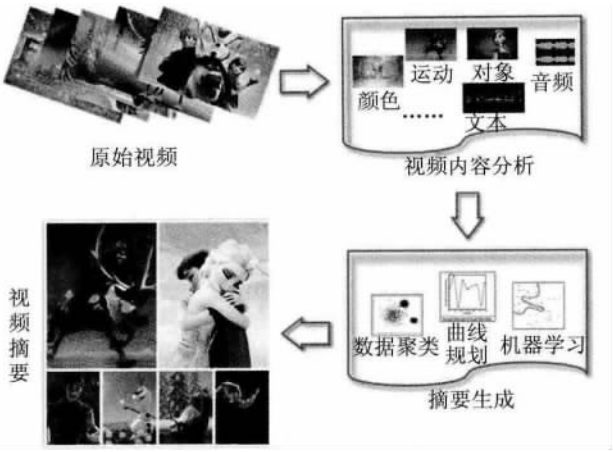

图2　视频摘要主要流程图

Fig. 2　The main flowchart of video abstraction

视频内容分析主要是对视频内部资源和视频外部资源进行分析。视频内部资源分析无非就是分析视频低级特征和由它们组合得到的对象和事件等高级特征。视频外部资源包括视频获取时的环境参数、视频观看群的习惯以及网络资源等等，妥善地分析视频内、外部资源可以改善视频数据和视频语义之间的鸿沟，较好地理解视频内容。图3展示了视频内容分析的主要分类。

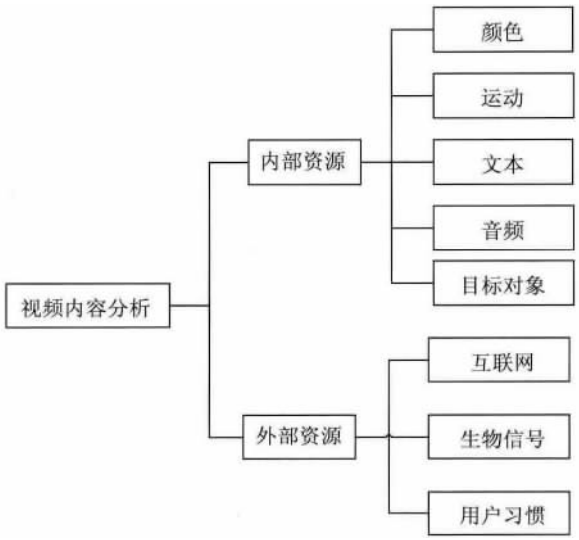

图3　视频内容分析结构图

Fig. 3　The structure of video content analysis

2.1.1　视频内部资源分析

颜色作为图像的重要属性，常被用于表征图像的统计信息。甚至对于一些指定域的视频，颜色信息可以直接表达语义，如足球视频中，绿色通常代表草地。由于HSV颜色间较RGB更适合人的感官视觉，Zhuang[7]和Almeida[8]通过分析视频帧在HSV空间中的颜色直方图信息来描述视频帧。不同的是，前者直接在空间域中分析，后者先对视频帧进行离散余弦变换(DCT)，然后取直流系数，即DC值构建DC图像，再对DC图像进行颜色特征分析。而考虑颜色对立空间Lab易于量化的特点，Coldefy[9]计算足球视频帧在Lab空间中颜色直方图的主要分量，从而判断草地区域并依据区域大小来确定镜头类型。

视频中的运动信息主要包括视频中目标对象的运动和拍摄相机的运动。通常情况下，运动的强度可以反映视频内容变化的程度。如在电影视频中，物体运动越剧烈，越可能对应故事情节的高潮片段。Wolf[10]率先提出采用光流法分析像素点在时域上的强度变换，进而推断视频帧像素点的运动速度和方向。Chau[11]在变化域中考虑运动信息，认为运动强度较大的帧一般含有较多的帧内编码宏块和运动矢量幅度较大的帧间编码宏块，据此分析宏块编码类型和运动矢量来衡量运动强度。而对于拍摄相机的运动检测，多利用仿射变换进行建模。Mei[12]通过比较连续帧来对如式(1)的仿射模型进行参数估计。

$$\begin{cases} v_x = a_0 + a_1 x + a_2 y \\ v_y = a_3 + a_4 x + a_5 y \end{cases} \tag{1}$$

式中，$a_i(i=0,\cdots,5)$为运动参数，而(v_x, v_y)为像素点(x, y)处的运动矢量。随后可得主要的相机运动pan、tilt、zoom和rotation的参数

$$\begin{cases} b_{\text{pan}} = a_0, b_{\text{tilt}} = a_3 \\ b_{\text{zoom}} = \dfrac{1}{2}(a_1 + a_5), b_{rot} = \dfrac{1}{2}(a_4 - a_2) \\ b_{\text{hyp}} = \dfrac{1}{2}(|a_1 - a_5| + |a_2 + a_4|) \\ b_{\text{err}} = \dfrac{1}{M \times N}\sum_{j=1}^{N}\sum_{i=1}^{M} |p(i,j) - p'(i,j)| \end{cases} \tag{2}$$

式中，b_{hyp}表征实际的相机运动参数，而b_{err}是仿射模型引入的估计误差。

同样地，Kim[13]针对MPEG视频分析相机运动，利用运动门限法将视频子镜头归类为5类主要相机运动。

体育视频中的计分板、演讲视频的演讲稿以及电影、新闻视频的字幕都是视频中的文本信息，包含有丰富的语义信息，如体育视频中的计分板既可显示球队信息又可显示比赛状态。视频中文本信息的

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

检测和分析对视频的事件理解有很好的指引作用。对计分板和字幕这类以图像形式显示的文本,主要的分析工作集中于文本的检测和识别。Lienhart 等人[14] 在其设计的 MoCA 电影视频摘要系统中,通过对字符区域进行聚类并对字符进行运动轨迹追踪,从而获取标题序列中文本的位图,最后将文本内容转化为 ASCII 存储。Zhang 等人[15] 对棒球比赛计分板中的比赛状态信息,如比分、局数等进行建模,利用状态转移图表征状态信息字符的时域变换关系,最后依据变化频率、字符区域等信息将字符进行分类。对于演讲稿这类纯文本的资源,通常可采用词频和词性分析来判定不同词的特征属性。Taskiran[16] 计算演讲稿中每个单词的词频,主要是分析每个单词在一段内出现的次数和该段的总单词数。对于特殊的词则进行单独处理,如停止词直接删除,对固定搭配词则计算其统计分布等,最终得到整个演讲稿的统计分析结果。不同于传统的 TF-IDF (term frequency-inverse document frequency),Lee[17] 为了分析单个词在整个文档集的分布,对 TF 和 IDF 进行了变种,计算在指定文本分类中每个词的 TF-IDF。

目前,在视频摘要领域,音视频结合多模态地分析视频内容已是一种较为常见的方法。视频中的音频对视频的主题和观看者的情绪都有一定的导向作用,尤其是在电影视频和体育视频中。而对音频的分析主要包括音频能量分析和音频归类等。Evangelopoulos 等人[18] 分析音频信号的幅频特性,基于幅频响应在时域和频域上量化音频特性,并结合音频能量分布对音频特征进行追踪。由于视频内容的差异,音频的归类也不尽相同。Jiang 等人[19] 对生活视频内包含的音频,在大量的音频数据训练的基础上,将音频归唱歌、掌声、说话和音乐 4 类。Xu 等人[20] 则将新闻视频中的音频分为静音、说话、音乐和噪音 4 类。

现实中,往往更加倾向于关注视频中特定的对象和目标。如监控视频中特定对象的行为和特定目标的状态,电影视频中和主要演员相关的事件,足球视频中球的运动等。Fu 等人[21] 分析监控视频中兴趣对象在时域和空域上运动的冗余性,获取对象的轨迹。在多目标方面,Lin 等人[22] 对电影视频出现的演员以及相互之间的交互关系进行分析,判断不同场景中对象的出现情况并建立场景社交网络,分析各个网络之间的依附关系。

此外,还可对视频结构进行分析,将视频分割为镜头和场景,得到场景数目和场景变换特征等。同时,利用一些先验知识可以获得简单的高级语义特征,如利用肤色检测人脸或由人脸出现的位置和顺序来判定会话事件[14]。

目前,大部分的视频内部资源分析都集中于分析视频低级特征。少数学者试图从低级特征中获取高级语义信息,如 Tamrakar[23] 利用静态特征 Gist、SIFT、colorSIFT 和动态特征 MOSIFT、SITP、DTF-HoG、DTF-MBH 对 15 个高级事件,如喂养动物、钓鱼、婚礼等进行识别。虽然分析视频低级特征直观又简便,但该方法对视频语义的获取太过粗糙,比如很难依据颜色信息去理解生成视频的内容。因此,未来更多的研究工作仍应该集中于探索如何从低级特征中获取高级语义。

2.1.2　视频外部资源分析

视频外部资源分析,主要是指借助外部资源,如网络、GPS 和各种传感器等,来辅助分析视频内容。外部资源分析可以从侧面反应视频内容的重要度和视频观看群的兴趣点,合理地利用外部资源可以为快速地获得有意义的用户偏好的视频摘要。

视频外部资源分析主要包括借助互联网来获取网络上对视频内容进行描述的文本摘要,或利用生物信号信息,如脑电波、心率等再结合用户浏览习惯来等等来综合分析用户对视频的关注度。下面就相应的技术分别进行探讨。

网络上的视频在呈现和传播时通常都包含对应的文本摘要用于对视频内容进行描述。合理地利用这些摘要不仅有利于视频事件理解,而且可以丰富最终的视频摘要。如 Babaguchi 等人[24] 借助足球网站来获取足球比赛的对应时间、比分和球队信息。而 Agnihotri 等人[25] 对音乐视频分析,通过 Internet 去搜索歌曲名称、歌手、歌曲类型和歌词等信息,辅助分析视频从而加深对原始视频的理解。

由于生物信号信息可以真实地反应用户对视频的直观感受,部分学者利用其来表征用户兴趣度。只是在信息获取上比较困难,通常需要借助特殊的仪器。Aizawa 等人[26] 基于人在激动时 α 波将会削弱而 β 波则会连续活跃的先验知识,借助头戴式的脑电波传感仪去分析用户在观看视频时的脑电波变化。在随后的研究[27] 中,Aizawa 将脑电波传感仪、GPS、加速传感器和回旋传感器相结合,对用户的头戴式相机拍摄的视频进行分析,获取对象运动的速

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 19, No. 12, Dec. 2014

度、位置和兴趣等信息。Peng 等人[28]结合视觉坐标、瞳孔直径和心率信息来检测目标视频中的重要场景。

用户的浏览习惯主要是指用户在观看视频时对播放器的操作，如快进、重放等。Yoshitaka[29]通过记录用户播放视频的时间线对应得到视频播放时用户的行为。Syeda-Mahmood[30]和 Mongy 等人[31]分别利用马尔可夫模型对用户观看视频时的暂停、快进、重放、结束等行为进行建模，前者利用最大似然估计模型的参数，而后者采用则通过小规模观看视频得到经验值。

除此之外，视频的生成环境、摘要的显示环境以及用户对摘要的定制要求，如摘要时长等等都可作为外部信息辅助分析视频内容，最终实现对视频内容的理解。

综上所述，外部资源分析的确为视频内容理解提供了额外的信息。但是，视频摘要本身面向商业应用较多，摘要系统的便捷性和实用性是必须考虑的问题。用户对由各种传感器或探测仪组成的庞大而粗糙的摘要设备的接受度不会很高，因此可以考虑如何利用身边常见的设备，如手机等来辅助进行外部资源获取。

2.2 摘要生成

视频内容分析旨在解析视频结构、理解视频内容。而摘要生成技术则是利用各种算法去除原始视频数据的冗余性，选取视频中具有代表性的、有意义的部分组合并生成最终的可视化摘要。图 4 给出了常见的摘要生成算法，主要有数据聚类法、曲线规划法和机器学习法。

数据聚类是一种无监督的模式分类方法，近年被广泛地运用于视频数据分析领域。该方法首先将视频数据流聚集成类，满足类间相似性尽量小而类内相似性尽量大的原则，再选取类中的中心数据作为类代表从而消除冗余性。Jain 等人[32]将聚类算法归为划分型和结构型两类，前者可一次划分数据确定所有分类，而后者需借助之前成功使用的聚类器以凝聚或者分裂的方式递归地进行分类。K-means 作为流行度最广的划分型聚类算法，实现较为简单。Amiri[33]利用一种改进的 K-means 算法来对镜头级的关键帧进行聚类。相较于传统的 K-means，该算法能自适应得到聚类数目。初始状态有 2 类，进入迭代后每一次计算聚类中心的距离，如果平均距离大于选定的阈值，则增加聚类数目，反之，停止算法，得到适宜的聚类数目。Guimarães 等人[34]使用基于图论的分裂式聚类算法，对原始视频帧基于帧间相似度构建图，再利用图论知识构建最小生成树从而对应视频中的关键帧。为了解决聚类算法中必须人为参与的问题，Frey 等人[35]于 2007 年提出了一种新的无监督聚类算法——仿射传播聚类(AP)，其基本思想在于通过消息传递，特指点之间的吸引度和归属度，实现数据点的自动聚类。AP 算法对大规模数据运算时速度较快，Xie 等人[36]利用 AP 聚类算法来对镜头提取的关键帧进行聚类，消除帧间冗余性。Shafeian 等人[37]为了得到个人偏好的摘要，依据用户的偏好设定 AP 矩阵中对角线元素的值，从而调整聚类结果。

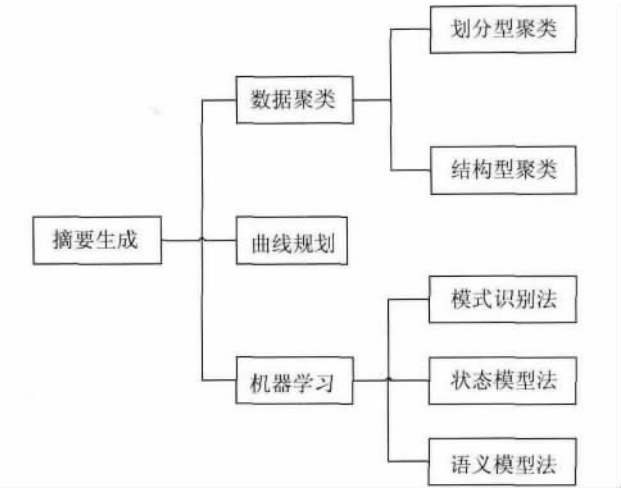

图 4 摘要生成结构图

Fig. 4 The structure of abstraction generation

其他常用的聚类方法还包括模糊聚类、谱聚类等。然而，并不存在哪一种算法适合于所有的数据类型。因此，数据聚类算法在实际应用时需针对数据特点进行相应算法的选择，而由于聚类过程中需要对样本进行不断划分和调整，在数据量较大时，聚类算法开销很大。同时，数据聚类忽略了视频帧的时域信息。

和数据聚类相似的是，曲线建模法也将视频帧视为多维特征空间的 1 个点。不过后者更注重视频帧的时域顺序，利用坐标轴中的曲线来拟合视频特征显示视频内容。Latecki 等人[38]基于帧间像素差异将视频帧映射为曲线，利用离散曲线演化将一段折线分割成折线集，再删除集合中相关性小的点来简化曲线。Han 等人[39]用 2 维曲线描述相机运动，将相机运动的不连续点对应为曲率变化较大的点，

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

并称其为视觉兴趣点。通过分析视觉兴趣点周围视频帧的运动、颜色和边缘等信息来进行事件探测。Albanese [40] 等人将足球视频分割为块,度量每个块中所包含的重要事件,并得到相应的优先权曲线。最后利用峰值检测算法进行块融合,消除冗余性得到紧凑的块集。

采用曲线建模可以简化视频的处理并直观地反应视频内容特性,但是这种方法只能对视频的主要变化进行描述,在语义上的表达并不完整。

机器学习指计算机利用数据建模来模拟和实现人类的学习行为,并获取新知识的过程。作为近年的研究热点,它已被广泛地运用于视频摘要领域。通过机器学习对复杂事件进行建模,完成视频事件的识别、理解和归类,从而消除视频冗余性,获得原始视频的精彩集锦。Lavee 等人 [41] 将基于学习的事件建模分为了模式识别法、状态模型法以及语义模型法 3 类。模式识别法主要是对训练集利用事件相关的先验知识进行训练学习,从而实现事件识别。Zawbaa 等人 [42] 利用 SVM 对足球视频中的回放标志进行训练和学习,识别回放镜头。该文作者认为进球事件的产生对应在回放标志之前依次出现停止、近镜头、观众镜头、球门区域和音频兴奋等场景,而在回放标志后则会出现长镜头、得分板等场景。而进攻事件则通常会出现停止、近镜头和球门区域等场景。通过检测到的回放镜头并判断其前后场景是否符合上述事实从而来进行足球视频语义事件监测。相较于模式识别法,状态模型法对事件本身的域知识要求较少,适应性更广。它主要通过语义信息对视频事件在时间和空间上的状态进行建模,找寻事件在状态空间的固有特性。常见的状态模型法有贝叶斯网络、隐式马尔可夫模型和条件随机场等。考虑 HMM 能较好地分析视频序列时域动态变化的特点,Huang 等人 [43] 利用其训练棒球视频并将视频镜头依据内容差异归为 15 类。不同于状态模型法,语义模型法并不需要构建出事件状态空间的全貌,只是去探测组成事件的各子事件之间的语义关系。该方法适应于复杂事件,研究相对较少。

大多数情况下,机器学习法需要人工参与辅助建模。同时该方法对建模时的特征选取较为敏感,不恰当的特征选取将使学习结果不尽人意。但是一旦选定了合适的特征,采用机器学习法可获得符合人类认知的高级语义信息。

3 视频摘要技术的新趋势

近 5 年,多媒体技术日趋成熟,网络视频数量正以无法估量的速度不断增加。视频摘要技术的研究也相应地有所偏倚,一方面市场应用对视频摘要的实时性和交互性的要求越来越高,而另一方面人们迫切地希望能依据个人喜好对视频在线提取摘要。基于此,实时视频摘要的研究越来越得到学者的青睐。

实时视频摘要通常将在线的网络视频分析作为分析对象。为了提高效率节省时间,一方面可以选取简单的特征或者改进的算法,如 STIMO [44] 采用改进的 Furthest-Point-First(FPF) 算法对预滤波后的视频帧提取的颜色直方图数据进行聚类,对 30 帧的故事板,STIMO 相较于采用 K-means 算法时间上快了 20 倍。而 OLAM(on-line abstraction module) [45] 对 10 min 的 CCTV 新闻在线生成摘要,通过选取简单的特征,处理时间仅为 1 min。另一方面,可通过直接在变换域中对视频进行处理节约解码时间,Almeida [8] 直接在变化域中对视频进行处理,提取 I 帧的 DC 值生成 DC 图像,再进行特征分析。

网络视频的不断增多,对实时视频摘要的要求也将越来越高。但妄图仅通过选取简单的特征来提高速度将使生成的摘要语义不完善,从而降低摘要质量。而目前压缩域的摘要技术研究较少,主要集中于 I 帧 DCT 系数分析、运动矢量分析和宏块类型分析 3 方面。并且,由于不同视频编解码方式的差异,在具体分析时算法仍须进行调整,较为复杂。

Fu 等人 [46] 在 2010 年第 1 次提出了多视角视频摘要,即对多视角、多摄像机拍摄的视频生成摘要。在实际生活中,对于一些重要场所,如银行、办公室,为了安全考虑通常会架设多台摄像机从不同角度进行监控摄像。对于此类视频,由于在同一时刻对同一场景有不同视角的镜头描述,因此在实际摘要时除了横向地在时域上选择具有代表性的内容,还应纵向地对同一时刻不同视角的镜头进行分析。

Fu 等人 [46] 通过构建镜头的时空序列图来展示同一时刻不同视角、同一视角不同空间的视频内容信息。首先,将不同视角获得的视频分割为镜头,然后利用超图显示不同视角的镜头关系,超边代表镜头间的相关性。再将超图转化为时空序列图,每个点值显示镜头的重要度值,各边的权重记录镜头间

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 19, No. 12, Dec. 2014

的相似度，最后利用随机漫步进行事件聚类，并最终得到多视角的故事板摘要。Leo 等人[47]在交通要道上间隔一定距离架设了的两台摄像机对路况进行监控，并对两个摄像机拍摄得到的视频提取摘要再结合运动分析进行异常事件监测。文献[47]提出利用概率潜在成分分析算法(PLCA)对视频中的高级运动行为识别并构建相机拓扑图来确定不同相机所包含运动的相关性，最终消除运动内部和运动外部的冗余性。

多视角视频摘要不仅需要对同一个视频中不同语义单元的进行识别区分，还必须实现同一语义单元在不同视角下的匹配和比较工作。研究难度较大，目前这方面的研究也较少。但是由于多视角拍摄视频广泛地存在于日常生活中，因此该研究具有较大的现实意义。

4 视频摘要评价系统

不同于目标识别等领域，截止目前视频摘要领域一直都缺乏统一的评价标准，定义所谓“正确”的摘要并不是一个简单的任务。Truong 等人[48]将现有的视频摘要评价方法归为结果描述、客观度量和用户学习 3 类。

结果描述作为最通用最简单的评价方法，并不包含与其他方法的比较。它主要是通过调整算法参数或改变摘要布局，来展示摘要的视觉差异。

客观度量通常是利用压缩率和保真度来判定摘要的优劣，即保证摘要在最大限度去除冗余性的前提下能重构还原原视频。如在 Liu[49]定义了镜头重构函数来显示关键帧的质量。

在为了完成用户交互任务而对视频进行摘要提取时，用户学习将是一种十分实用的评价方法。这种方法通常是选择一定数目的用户来观看然后比较不同摘要的优劣并进行统计判断。Furini 等人[44]先对故事板摘要质量从 1(bad quality)到 5(good quality)设定了 5 个等级，再选了 20 个不同背景的用户来进行测试。这种方法一方面由于测试群体可能不具有代表性，而另一方面评定的标准比较模糊，因此也很难衡量摘要技术的实际价值。

而也有部分学者依据自身理解，从主客观两方面提出了高质量视频摘要所具有的特性。He 等人[50]认为一个好的视频摘要需要考虑简明性、覆盖率、合理性和连贯性。其中，简明性和覆盖率是通过摘要长度和摘要的内容从客观上来评价摘要，而后两者则是主观地评判摘要的合理性和流畅度。Valdés 等人[45]对生成的动态视频摘要提出的 4 个评判的准则，分别是大小、连续性、冗余性和运动强度。其中，连续性是统计连续分段被选入最终摘要的比例，冗余性则是计算选择的视频段之间的相似性，而运动强度则是视频段优先权的一种体现，因为通常情况下，运动强度越大的视频段被选入最终的摘要的概率越大。

NIST 曾针对其 TRECVID 2007 年和 2008 年(http://trecvid.nist.gov/)的任务提出了对未编辑的 Rushes 视频摘要的评价标准。该评价系统针对原始的大规模视频集，合理地标定了视频集中的基准事件，并提供了主、客观的评分标准。

具体的，主观方面 2007 年与 2008 年略有差异，不过主要方法都是让用户观看摘要对摘要整体内容进行感知，再回答下述问题，程度由“非常同意”到“非常不同意”共 5 个等级完成对摘要的评价。

TRECVID 2007 年与 2008 年 Rushes 视频主观评价标准如下:

1) 2007 年 Rushes 视频主观评价:

(1) 摘要是否易于观赏和理解;

(2) 摘要中是否包含多余的视频分段。

2) 2008 年 Rushes 视频主观评价:

(1) 摘要中是否包含相同的内容;

(2) 摘要中是否包含颜色条、纯黑或纯白帧;

(3) 摘要呈现是否连贯。

客观上这两年的评价方法是一样的。主要是判定摘要中包含的原始事件的比例和摘要时长等数据。

TRECVID 2007 年和 2008 年 Rushes 视频客观评价标准如下:

1) 摘要中包含事件与原始视频中包含的事件比例;

2) 比较摘要和标准所花的时间;

3) 摘要时长;

4) 生成摘要耗费的时间;

5) 摘要可用性的质量得分。

Valdés 等人[51]在 TrecVid 2008 年任务的基础上提出了一种基于视觉分析和机器学习的动态摘要评价系统。该作者利用 TrecVid 2008 中的视频集和学者们提交的摘要结果来进行学习训练，对摘要中包含的基准事件、摘要的冗余性和连贯性进行分析

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

并综合评定摘要的质量。

最后,在视频摘要技术与视频检索等相结合时,摘要的质量也可以通过查全率、查准率等指标从侧面来进行量化。

一个客观而全面的摘要评价标准可以很好地评定不同摘要算法的优劣,对推动摘要技术的发展也有很大的意义。然而,由于摘要在不同的应用背景下要求不尽相同,这使得建立统一的评价标准变得较为困难。基于此,不妨具体问题具体评定,忽略视频内容差异和摘要的形式,仅针对特定的摘要应用环境来给出合理的评定标准。如 TRECVID 2007 年和 2008 年中,仅针对未编辑视频给出了主客观两方面的评价标准。

5 结　语

综上所述,本文依据视频内容分析和摘要分析两个摘要生成主要步骤总结了视频摘要的主要研究技术,并介绍了近 5 年摘要领域的新研究趋势: 实时视频摘要和多视角视频摘要。最后,还讨论不同的摘要评价标准。

从前文的讨论可以看出,类似于视频标注和视频检索,视频摘要同样面临着视频语义获取困难的问题。考虑如何从视频数据流中获取高级语义,并缩小视频数据流和语义信息之间的鸿沟,给出以下建议:

1) 视频内部和外部特征相结合,共同表达视频语义信息。视频摘要技术的特征分析历程是从早期的分析视频底层特征,如颜色、纹理等,到结合音视频和文本多模态来综合理解视频的内容,再到分析生物信号特征,如心率、瞳孔的直径等。这是视频特征分析由内到外的一种扩展。视频内部特征是从客观上描述视频的本质内容,而视频外部特征,尤其是用户相关信息,可以主观地表达视频内容的实用性和重要度。二者相结合,能更加轻易地表达视频语义并且获得有意义的视频摘要。

2) 视频摘要和文本摘要相结合,全面地描绘视频内容信息。文本摘要是利用计算机对文本信息进行理解并对其主要内容进行概括,具有较强的直观性,易于理解。在体育视频和电影视频的摘要分析中,可以通过文本的识别和检测来获得视频的重要信息。而在新闻视频的摘要提取中,利用文本分析方法生成文本关键字,并与视觉信息组成最终的摘要的例子也并不少见。对于一般的视频,识别视频实体并进行语义标注,再结合文本模板可以生成相应的文本摘要。将视频摘要和文本摘要相结合可以更加直观和全面地描绘视频内容信息,从而能快速地理解视频内容。

依据文中第 3 节对摘要新趋势的分析可知,视频摘要技术未来的一个重要研究方向将会是视频摘要的商品化。视频摘要本身具有很大的商业前景,而当将其作为成熟商业产品进行包装时,仍需考虑许多现实问题:

1) 视频摘要的实时性和交互性。不同用户拥有不同的偏好,从而对摘要长度、摘要内容的要求有所差异。允许用户依据个人偏好定制需求获得专属摘要将是视频摘要商业化必须考虑的问题。随之,为了保证用户的满意度,摘要的实时性也必须得到满足。

2) 视频摘要的便携性。生成高质量的摘要,离不开硬件设备。而即使是一台笔记本,由于缺乏移动性,也将大大影响摘要技术的传播。一旦生成摘要的硬件设备太过繁复,则视频摘要终将沦为“实验室”产品。因此,怎样让用户利用身边常见的电子设备,如手机、平板等,随时随地对自己拍摄的或者自己感兴趣的视频生成高质量摘要,也是摘要商品化需解决的问题。

参考文献(References)

[1] Maybury M T. Broadcast news understanding and navigation [C] //Proceedings of the Fifteenth Conference on Innovative Applications of Artificial Intelligence. Trier, German: DBLP, 2003: 117-122.

[2] Pfeiffer S, Lienhart R, Kühne G, et al. The MoCA project. [M] //Informatik'98. Berlin, Heidelberg: Springer, 1998: 329-338.

[3] Chang S F, Chen W, Meng H J, et al. VideoQ: an automated content based video search system using visual cues [C] //Proceedings of the 5th ACM International Conference on Multimedia. New York, USA: ACM, 1997: 313-324.

[4] Snoek C G M, Worring M. Time interval maximum entropy based event indexing in soccer [C] // Proceedings of IEEE International Conference on Multimedia and Expo. Washington DC, USA: IEEE, 2003: 481-484.

[5] Uchihashi S, Foote J, Girgensohn A, et al. Video manga: generating semantically meaningful video summaries [C] //Proceedings of the seventh ACM International Conference on Multimedia (Part 1) . New York, USA: ACM, 1999: 383-392.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 19, No. 12, Dec. 2014

[6] Wu L Q, Li G H. Video's structured browsering and querying system: Videowser [J]. Mini-micro Systems, 2001, 22(1): 112-115. [吴玲琦, 李国辉. 视频结构化浏览和查询系统: Videowser [J]. 小型微型计算机系统, 2001, 22(1): 112-115.] [DOI: 10.3969/j.issn.1000-1220.2001.01.030]

[7] Zhuang Y, Rui Y, Huang T S, et al. Adaptive key frame extraction using unsupervised clustering [C] // Proceedings of International Conference on Image Processing. Washington DC, USA: IEEE, 1998, 1: 866-870. [DOI: 10.1109/ICIP.1998.723655]

[8] Almeida J, Torres R D S, Leite N J. Rapid video summarization on compressed video [C] // IEEE International Symposium on Multimedia. Washington DC, USA: IEEE, 2010: 113-120. [DOI: 10.1109/ISM.2010.25]

[9] Coldefy F, Bouthemy P. Unsupervised soccer video abstraction based on pitch, dominant color and camera motion analysis [C] //Proceedings of the 12th Annual ACM International Conference on Multimedia. New York, USA: ACM, 2004: 268-271.

[10] Wolf W. Key frame selection by motion analysis [C] // Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Washington DC, USA: IEEE, 1996, 2: 1228-1231. [DOI: 10.1109/ICASSP.1996.543588]

[11] Chau W S, Au O C, Chong T S. Key frame selection by macroblock type and motion vector analysis [C] // Proceedings of International Conference on Multimedia and Expo. Washington DC, USA: IEEE, 2004, 1: 575-578. [DOI: 10.1109/ICME.2004.1394257]

[12] Mei T, Tang L X, Tang J, et al. Near-lossless semantic video summarization and its applications to video analysis [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2013, 9(3): #16.

[13] Kim J G, Chang H S, Kim J, et al. Efficient camera motion characterization for MPEG video indexing [C] // Proceedings of International Conference on Multimedia and Expo. Washington DC, USA: IEEE, 2000, 2: 1171-1174. [DOI: 10.1109/ICME.2000.871569]

[14] Lienhart R, Pfeiffer S, Effelsberg W. Video abstracting [J]. Communications of the ACM, 1997, 40(12): 54-62.

[15] Zhang D, Chang S F. Event detection in baseball video using superimposed caption recognition [C] //Proceedings of the 10th ACM International conference on Multimedia. New York, USA: ACM, 2002: 315-318.

[16] Taskiran C M, Pizlo Z, Amir A, et al. Automated video program summarization using speech transcripts [J]. IEEE Transactions on Multimedia, 2006, 8(4): 775-791.

[17] Lee S, Kim H. News keyword extraction for topic tracking [C] // Proceedings of the 4th International Conference on Networked Computing and Advanced Information Management. Washington DC, USA: IEEE, 2008, 2: 554-559. [DOI: 10.1109/NCM.2008.199]

[18] Evangelopoulos G, Zlatintsi A, Potamianos A, et al. Multimodal saliency and fusion for movie summarization based on aural, visual, and textual attention [C] //IEEE Transactions on Multimedia, Washington DC, USA: IEEE, 2013: 1553-1568. [DOI: 10.1109/TMM.2013.2267205]

[19] Jiang W, Cotton C, Loui A C. Automatic consumer video summarization by audio and visual analysis [C] // Proceedings of International Conference on Multimedia and Expo. Washington DC, USA: IEEE, 2011: 1-6. [DOI: 10.1109/ICME.2011.6011841]

[20] Xu S, Feng B, Xu B. Multi-modal topic unit segmentation in videos using conditional random fields [C] // Proceedings of International Conference on Acoustics, Speech and Signal Processing. Washington DC, USA: IEEE, 2013: 2287-2291. [DOI: 10.1109/ICASSP.2013.6638062]

[21] Fu W, Wang J, Zhao C, et al. Object-centered narratives for video surveillance [C] // Proceedings of the 19th IEEE International Conference on Image Processing. Washington DC, USA: IEEE, 2012: 29-32. [DOI: 10.1109/ICIP.2012.64666787]

[22] Lin C, Tsai C, Kang L, et al. Scene-based movie summarization via role-community networks [C] //IEEE Transactions on Circuits and Systems for Video Technology, Washington DC, USA: IEEE, 2013: 1927-1940. [DOI: 10.1109/TCSVT.2013.2269186]

[23] Tamrakar A, Ali S, Yu Q, et al. Evaluation of low-level features and their combinations for complex event detection in open source videos [C] // Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Washington DC, USA: IEEE, 2012; 3681-3688. [DOI: 10.1109/CVPR.2012.6248114]

[24] Babaguchi N, Kawai Y, Kitahashi T. Generation of personalized abstract of sports video [C] // Proceedings of International Conference on Multimedia and Expo. Washington DC, USA: IEEE, 2001: 619-622. [DOI: 10.1109/ICME.2001.1237796]

[25] Agnihotri L, Dimitrova N, Kender J R. Design and evaluation of a music video summarization system [C] // Proceedings of International Conference on Multimedia and Expo. Washington DC, USA: IEEE, 2004, 3: 1943-1946. [DOI: 10.1109/ICME.2004.1394641]

[26] Aizawa K, Ishijima K, Shiina M. Summarizing wearable video [C] // Proceedings of International Conference on Image Processing, Washington DC, USA: IEEE, 2001, 3: 398-401. [DOI: 10.1109/ICIP.2001.958135]

[27] Aizawa K, Tancharoen D, Kawasaki S, et al. Efficient retrieval of life log based on context and content [C] //Proceedings of the 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences. New York, USA: ACM, 2004: 22-31.

[28] Peng W T, Chu W T, Chang C H, et al. Editing by viewing: automatic home video summarization by viewing behavior analysis [J]. IEEE Transactions on Multimedia, 2011, 13(3): 539-550.

[29] Yoshitaka A, Sawada K. Personalized Video summarization based on behavior of viewer [C] // Proceedings of the 8th International Conference on Signal Image Technology and Internet Based Systems. Washington DC, USA: IEEE, 2012: 661-667. [DOI:

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

10. 1109/SITIS. 2012. 100]

[30] Syeda-Mahmood T, Ponceleon D. Learning video browsing behavior and its application in the generation of video previews [C] //Proceedings of the 9th ACM International Conference on Multimedia. New York, USA: ACM, 2001: 119-128.

[31] Mongy S. A study on video viewing behavior: application to movie trailer miner [J]. The International Journal of Parallel, Emergent and Distributed Systems, 2007, 22(3): 163-172.

[32] Jain A K, Murty M N, Flynn P J. Data clustering: a review [J]. ACM Computing Surveys, 1999, 31(3): 264-323.

[33] Amiri A, Fathy M. Hierarchical keyframe-based video summarization using QR-decomposition and modified k-means clustering [J]. EURASIP Journal on Advances in Signal Processing, 2010: #102.

[34] Guimarães S J F, Gomes W. A static video summarization method based on hierarchical clustering [M] //Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Berlin, Heidelberg: Springer, 2010: 46-54.

[35] Frey B J, Dueck D. Clustering by passing messages between data points [J]. Science, 2007, 315(5814): 972-976.

[36] Xie X, Wu F. Automatic video summarization by affinity propagation clustering and semantic content mining [C] // The 2008 International Symposium on Electronic Commerce and Security. Washington DC, USA: IEEE, 2008: 203-208. [DOI: 10. 1109/ISECS. 2008. 118]

[37] Shafeian H, Bhanu B. Integrated personalized video summarization and retrieval [C] // Proceedings of the 21st International Conference on Pattern Recognition. Washington DC, USA: IEEE, 2012: 996-999.

[38] Latecki L J, DeMenthon D, Rosenfeld A. Extraction of key frames from videos by polygon simplification [J]. Proc. of Signal Processing and its Applications, 2001: 643-646.

[39] Han S H, Kweon I S. Scalable temporal interest points for abstraction and classification of video events [C] // Proceedings of IEEE International Conference on Multimedia and Expo. Washington DC, USA: IEEE, 2005: 670-673. [DOI: 10. 1109/ICME. 2005. 1521512]

[40] Albanese M, Fayzullin M, Picariello A, et al. The priority curve algorithm for video summarization [J]. Information Systems, 2006, 31(7): 679-695.

[41] Lavee G, Rivlin E, Rudzsky M. Understanding video events: a survey of methods for automatic interpretation of semantic occurrences in video [J]. IEEE Transactions on Systems, Man, and Cybernetics: Part C: Applications and Reviews, 2009, 39(5): 489-504.

[42] Zawbaa H M, El-Bendary N, Abraham A. SVM-based soccer video summarization system [C] // Proceedings of the 3rd World Congress on Nature and Biologically Inspired Computing (NaBIC). Washington DC, USA: IEEE, 2011: 7-11. [DOI: 10. 1109/NaBIC. 2011. 6089409]

[43] Huang C L, Chang C Y. Video summarization using hidden Markov model [C] // Proceedings of International Conference on Information Technology: Coding and Computing. Washington DC, USA: IEEE, 2001: 473-477. [DOI: 10. 1109/ITCC. 2001. 918841]

[44] Furini M, Geraci F, Montangero M, et al. STIMO: STIll and MOving video storyboard for the web scenario [J]. Multimedia Tools and Applications, 2010, 46(1): 47-69.

[45] Valdés V, Martínez J M. On-line video abstract generation of multimedia news [J]. Multimedia Tools and Applications, 2012, 59(3): 795-832.

[46] Fu Y, Guo Y, Zhu Y, et al. Multi-view video summarization [J]. IEEE Transactions on Multimedia, 2010, 12(7): 717-729.

[47] Leo C, Manjunath B S. Multicamera video summarization and anomaly detection from activity motifs [J]. ACM Transactions on Sensor Networks (TOSN), 2014, 10(2): #27. [DOI: 10. 1145/2530285]

[48] Truong B T, Venkatesh S. Video abstraction: A systematic review and classification [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2007, 3(1): #3.

[49] Liu T, Zhang X, Feng J, et al. Shot reconstruction degree: a novel criterion for key frame selection [J]. Pattern Recognition Letters, 2004, 25(12): 1451-1457.

[50] He L, Sanocki E, Gupta A, et al. Auto-summarization of audio-video presentations [C] //Proceedings of the 7th ACM International Conference on Multimedia: Part 1. New York, USA: ACM, 1999: 489-498.

[51] Valdés V, Martínez J M. Automatic evaluation of video summaries [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2012, 8(3): #25. [DOI: 10. 1145/2240136. 2240138]

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net