

基于动画特征的视频摘要方法*

谭洁¹, 吴玲达², 应龙²

(1. 解放军南京政治学院 计算机教研室, 南京 210003; 2. 国防科学技术大学 信息系统与管理学院, 长沙 410073)

摘要: 针对动画视频与新闻视频、体育运动视频的不同特点, 提出了一种适合动画视频的视频摘要技术。首先通过对动画视频的结构分析得到动画视频的可视特征与层次结构; 然后根据动画视频内容的重要度来选取视频中的重要片段; 最后通过粒度选择, 按照时序方式组合得到故事板和缩略视频形式的视频摘要。实验表明该方法能有效地获得动画视频摘要。

关键词: 动画视频摘要; 视频结构分析; 缩略粒度; 故事板; 缩略视频

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2009)10-3960-03

doi:10.3969/j.issn.1001-3695.2009.10.104

Video summarization based on animation distinct property

TAN Jie¹, WU Ling-da², YING Long²

(1. Computer Center, Nanjing Politics College, Nanjing 210003, China; 2. College of Information System & Management, National University of Defense Technology, Changsha 410073, China)

Abstract: Based on distinct property of animation video which was different from that of other videos, especially news video and sports video, this paper proposed a video abstract technique, suitable for animation video. It analyzed the construction of animation video. Then got the visual character and clear construction, defined the granularity of video scene. Based on content evaluation model of scene significance and the granularity of scene, found the important scenes of the video. Then, in the temporal order, proposed the video abstract including storyboard and video skim. Experimental results show that the proposed method extracts animation video summarization efficiently.

Key words: animation video summarization; video construction analysis; skim granularity; storyboard; video skim

0 引言

2005 年以前国内卡通动画市场中进口卡通动画的市场占有率达 90%, 全国卡通动画从业者不足 1 万人, 只及韩国的三分之一。每年卡通动画片的缺口达 25 万 min。卡通动画产业在国内生产总值中的比重只有十万分之一。通过最近几年的发展, 国内的动漫产业取得了长足的进步。动漫产业的快速发展促生了海量的动画视频, 以湖南三辰卡通集团出品的大型科普动画片《蓝猫淘气 3 000 问》为例, 到 2008 年 10 月, 该片已经完成了超过 2 800 集, 每集素材平均存储量不低于 4 GB。与新闻、广告、监控视频、家庭视频等数字视频类似, 动画视频也面临视频摘要的需求。动画视频摘要不仅能快速浏览动画视频数据, 帮助观众迅速了解动画所描述的内容; 同时也能够帮助动画工作者检索需要的素材, 以加速新动画的生成。

随着多媒体技术的发展和大量数字视频的涌现, 引发了一系列包括视频存档、编目、索引以及有效存取等新技术。其中, 视频摘要技术的出现是为了解决如何快速浏览大容量的视频数据, 如何获取和表现视频内容的问题。视频摘要, 即以自动或半自动的方式对视频的结构和内容进行分析, 从原视频中提取出有意义的部分, 并将它们以某种方式进行组合, 形成简洁的能够充分表现视频语义内容的概要^[2]。目前, 比较著名的

研究成果包括: 卡内基·梅隆大学的 Informedia 项目^[4,5] 针对视频的镜头检测、图像分析, 以及标题型字幕文本和伴音信息分析进行了研究; Mannheim 大学的 MoCA 项目^[6] 研究了自动提取视频的结构和语义内容。国内, 微软亚洲研究院对视频的场景分割、字幕探测与识别等技术进行了研究; 中国科学院对压缩域下新闻视频的分割以及字幕探测技术进行了研究; 清华大学、北京大学对镜头分割、关键帧提取、播音员镜头探测等技术进行了研究等。目前对动画视频摘要方面的研究还属于崭新的课题, 国内外在该领域的研究相对较少。

本文分析了动画视频的结构特点, 比较了动画视频和其他视频, 如新闻视频、体育视频的差异, 提取了动画视频独有的特征, 提出了适合动画视频摘要的方法。

1 动画视频的结构特点

动画是逐帧制作的, 视、听结合, 通过以一定的速率回放所记录的画面而产生运动视觉的一种媒体表现形式^[3]。动画视频与目前视频检索领域研究相对较多的新闻视频和体育运动视频结构相比较, 动画视频的结构有其自身的特点。

连续的视频数据通常由相关联的场景组成, 每个场景又包含若干镜头。根据视频结构特点, 将视频数据按照由粗到细的顺序划分为四个层次结构: 故事层、场景层、镜头层和图像帧

收稿日期: 2009-01-05; **修回日期:** 2009-02-10 **基金项目:** 国家科技支撑计划资助项目(2007BAH14B01); 国家自然科学基金资助项目(60802080)

作者简介: 谭洁(1971-), 女, 湖南长沙人, 讲师, 主要研究方向为多媒体技术、计算机动画; 吴玲达(1962-), 女, 上海人, 教授, 博士, 主要研究方向为多媒体信息系统与虚拟现实技术; 应龙(1976-), 男, 江苏淮安人, 讲师, 博士, 主要研究方向为计算机动画(kenstella@hotmail.com)。

层,如图 1 所示。目前,视频结构分析的研究集中于从视频数据中提取出中间两层结构单元,即镜头和场景。并且,由于动画的故事情节相对集中,动画视频的结构分析中故事层的分析处于相对次要的位置。

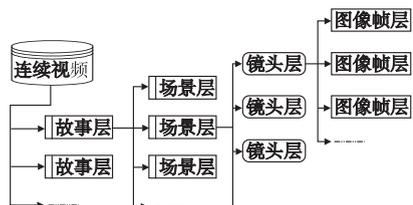


图 1 连续视频的层次化结构

目前,视频摘要的研究领域中对于新闻视频和体育运动视频的研究相对成熟。新闻视频除了开头和结尾外,其主题内容是一系列的播音员镜头和新闻单元(有时中间会附加广告镜头)^[7]。同时,视频上会出现标题型的字幕,这些字幕一般会持续到一则新闻的结束。对其结构分析主要在于检测出单独的播音员镜头和新闻事件片段(以及广告内容)。通过播音员镜头的图像特点分析,如播音员一般采取坐姿播音、播音员在整个画面中占据主要的画面比例,根据这些特征可以检测出当前画面中是否是播音员镜头。同时根据广告片段具有强烈的视觉、听觉效果的特征判断出新闻视频中存在的广告镜头。

体育运动视频的视频摘要一般是对精彩镜头摘要(high-light summary)。体育运动视频的精彩镜头会伴有观众的欢呼声、掌声等,同时会有慢镜头回放或对运动员的特写镜头。以足球比赛为例,一个射门镜头之后往往会给出射门者的进球庆祝或没有进球惋惜的镜头,同时伴有观众的欢呼声音和解说员快速解说的声音,并且显示射门者的基本资料,之后会出现射门场面的慢镜头回放^[8],有时候还会有一些赛事标志的镜头或赛事宣传片出现。

动画视频一般具有丰富的叙事情节,除了拥有一般视频的基本特征外,相对于新闻和体育视频,其自身的结构特征包括:

a) 动画视频字幕一般是故事人物的对白或者旁白,位置相对固定,并且贯穿于整个视频内容,区别于新闻视频中的概括性标题字幕。

b) 视频结构分段特征不明显。动画视频中不存在新闻视频中播音员镜头和新闻事件的显著区别,也不存在体育视频中运动员特写镜头或者慢动作回放等标志性的镜头。

c) 动画视频中为了表现其夸张的效果会出现时间相对较长的单色调、高亮度、低亮度等镜头,反之也会有一些场景变化很快的镜头。

上述特征给动画视频的结构分析和视频摘要工作带来一些新的困难。但是,如果利用好上述特征也会给动画视频摘要带来一些契机。

2 动画视频摘要的生成

动画视频摘要是以动画的内容分析为基础的,对动画视频的内容进行结构分析,再对视频内容片段按照重要程度进行选取,最后将选取的片段根据设定的缩略粒度进行组合,得到动画视频摘要,其过程如图 2 所示。

2.1 动画视频结构分析

对动画内容进行结构化分析和处理,将其划分为镜头、场景、故事等不同层次的基本结构单元,并提取用于描述不同层

次的关键帧,最终使得动画视频具有层次结构,像文本编目一样便于快速浏览和检索。其中需要解决的关键技术包括颜色特征提取、镜头边界探测、关键帧提取等。



图 2 动画视频摘要生成

2.1.1 颜色特征提取

视频的可视特征以颜色为主,动画视频尤其如此。颜色特征主要包括颜色直方图、颜色对、主色调等。图像的整体颜色分布常用颜色直方图表示。直方图特征描述了图像颜色(或灰度)的统计特性,反映了图像颜色的统计分布和基本色调,但是直方图不包含位置特征,因此不同的图像有可能具有相同的直方图特征。颜色直方图能够反映整幅图像的颜色特征,但不能反映空间特性。采用颜色对的方法可以用来描述图像颜色的空间分布特性。所谓颜色对,是将一幅图像分成若干小块,首先计算每个颜色块颜色的平均值作为当前块的代表色,不同块之间的颜色差即为颜色对(可以用欧式距离来表述),最后,统计指定颜色差的颜色对数目。这一统计结果中就含有图像颜色分布的空间信息。将获取的颜色特征存储,以备后续分析所用。

2.1.2 镜头边界探测

准确的镜头边界探测结果是视频内容结构分析的重要基础。根据视频的可视特性或者音频特征,分析两个相连镜头之间视频内容的变化情况,可以把镜头边界分为突变和渐变的两种。产生突变镜头边界的原因通常有两种:a)画面内容的跳转,如近景和远景镜头的直接切换,或其他蒙太奇的手法都会造成突变的镜头边界;b)采用比较硬的过渡特技造成的,如为了强调表现对象的变化效果,在同一内容、同一景别的镜头之间采用了“Slide”的切换特技也会产生突变的镜头边界。

渐变镜头边界的产生同样有类似的两个原因:a)画面本身的连续性,如在两个近景对象处于同一背景当中,在两个近景镜头切换前都先缓慢过渡到远景画面,这样在镜头边界处的变化会很缓慢;b)由于采用了平滑的过渡特技,在任意变化剧烈的镜头之间只要采用“Cross Dissolve”特技都可以基本消除镜头边界的突变。突变和渐变应该采取不同的探测方法。对于镜头的突变,由于突变通常只发生在连续的两帧之间,当连续的两帧视觉内容出现巨大变化时,可以判定突变边界的存在。对于镜头渐变,可以在双重比较法的基础上进行渐变过程探测^[11]。

2.1.3 关键帧选取

镜头关键帧是能反映镜头主体内容的代表帧,对镜头内容的分析往往可以转换为对镜头关键帧内容的分析。关键帧的选取可以采用基于时间线和基于内容的两种方法。动画视频的制作过程首先就是从关键帧开始的,并且一部卡通片的关键帧数目是相对稳定的。国内,通常一部 9.5 min 的卡通片包含 150 个左右的镜头数目,每个镜头内画 3~5 个关键帧。在掌握了这个统计规律后,可以结合时间和内容两种策略选取关键帧。首先,根据镜头的大概数目设定镜头边界探测的阈值,由镜头探测的结果将动画视频切分成视频片段,选取片段的开始帧或者中间帧作为该片段的代表帧;然后再依据片段的长度设定相应的阈值,决定当前片段是否再提取其他的代表帧,将这

些提取的代表帧组合在一起即形成了代表整个动画视频的关键帧。

2.2 动画视频内容重要度

动画视频摘要需要在结构化分析的基础上对视频内容重要度进行判定,选择概括性强、能够表现主要内容的视频片段,组合编辑得到符合需求的视频摘要。动画片段选择解决两个问题:a)动画视频内容重要度判定,重要度是指视频中各实体要素能够表现整个视频内容的程度指标;b)动画视频片段的缩略粒度选择,缩略粒度是指生成视频摘要过程中摘要内容的精细程度。

在视频摘要之前,首先要明确动画的一些重要因素,包括动画人物、事件以及重要场景等。一般在一部动画片里,动画人物有着固定的人物形象,包括外形特征、颜色特征以及特有的动作特点。事件是动画中的重要内容,也是推动故事情节发展的主线,特别是那些能表现主题思想的激烈运动场面,是动画片的标志性内容。动画片为了表现其精彩的故事和夸张的效果,会选择一些特定的场景来发生和容纳这些重要的事件。

对新闻视频摘要的重要度可以用字幕帧、人脸帧以及镜头关键帧的重要度的线性组合来表示^[9]。本文利用动画主要人物和镜头的重要度来建立动画视频镜头的重要度模型。镜头重要度用下式表示:

$$I_{sh} = \lambda_c I_c + \lambda_{kf} I_{kf} \quad (1)$$

其中: I_c 、 I_{kf} 分别表示动画人物、镜头关键帧的重要度, λ_c 、 λ_{kf} 分别为相应的权值。动画人物和镜头关键帧的重要度通过其视频出现的次数以及它们的经验权值获得。

场景重要度主要体现在镜头的重要度和时间长短。场景重要度用下式表示:

$$I_{sc} = \lambda_{sh} I_{sh} + \lambda_{sc} T_{sc} \quad (2)$$

其中: T_{sc} 表示动画视频中场景的时间, λ_{sh} 、 λ_{sc} 分别为镜头、场景时间的权值。通过分析场景的重要度来判断视频片段的重要度。

2.3 动画视频片段组合

从最终表现形式上来看,对于动画视频来说,静态的故事板和动态的缩略视频比较适合做动画视频摘要。视频摘要效果主要看摘要的概括性、缩略性以及可理解性。视频摘要可理解,在能概括整个视频内容的基础上时间应尽可能地短,即使是重要的镜头,也没有必要把它完整地包含到摘要中,因此,选择合适的缩略粒度有利于控制视频摘要的可理解性和缩略性。同一镜头内的视觉特征一般不会发生太大变化,将一个镜头全部放到摘要中会产生冗余;而理解镜头需要一段时间,太短的镜头又难以表达原始的语义。基于这种思想,根据用户对动画内容的熟悉程度可以按照指定的缩略粒度对动画视频进行摘要。例如,开始设定以原视频的 25% 进行视频摘要,如果多数用户可以理解本段视频就可以适当再缩短摘要视频的时间,反之,可以适当增加摘要视频的时间。

2.3.1 故事板

对于故事板形式的动画视频摘要,其生成策略如下:确定组成故事板的图像数。考虑到缩略的粒度,采用时间比例的方法,约定平均每分钟的节目需要多少幅图像来表现;采用重要度判定模型,将场景重要度归一化,判断场景所占摘要时间比例,得到该场景下图像的数目,计算方法如式(3):

$$N(Sc)_k = T \cdot n \cdot (I_{sc})_k / \sum_k (I_{sc})_k \quad (3)$$

其中: $N(Sc)_k$ 为图像数; T 为摘要时间; n 为帧率; $(I_{sc})_k$ 为第 k 场景的重要度。 K 为整个原始视频所分割的镜头数。

根据镜头重要度判定模型来求得镜头关键帧的重要度,约定重要度在某个阈值之上的关键帧作为故事板图像的来源。计算出图像数目后,只需简单地选取重要度大的关键帧,按照时序组合即生成故事板。

2.3.2 缩略视频

缩略视频是与原视频最为接近,表现力最为丰富的一种摘要形式。其生成策略如下:根据确定的缩略粒度,即希望获得的缩略程度(如希望缩略视频的长度占原视频长度的 25% 左右),依据场景的重要度,判断场景在摘要中的时间,时间可由式(4)确定

$$T(Sc)_k = T \cdot (I_{sc})_k / \sum_k (I_{sc})_k \quad (4)$$

计算镜头的重要度,同时根据缩略粒度设定镜头重要度阈值,以获取相应的镜头来源。考虑到镜头的可理解性,对短镜头尽量不再剪辑,对于长镜头,可以选择其中的部分关键帧作为该段视频的代表帧。例如,将小于 4 s 的镜头不做剪辑,全部保留,而大于 4 s 的镜头根据关键帧重要度来截取。最后将获得镜头的关键帧按照时序方式组合,生成缩略视频。

3 实验结果分析

本文利用每集时长为 9.5 min、MPEG-4 标准压缩的动画视频《虹猫蓝兔七侠传》系列动画片对算法进行了验证。首先利用任意挑选出的 15 集作为学习样本,通过对镜头重要度权值的调整使得每集提取的关键镜头数目在 140 ~ 160。图 3 是在设置参数后对重新选取的一集视频的分析结果。该集共计包括四个小的故事片段。其中片段 2 包括五个主要场景。图 4 表示的其中一个场景包括了五个重要的镜头。图 5 表示的是镜头的代表帧。图 6 表示的是部分故事板形式的视频摘要。从实验结果看,就该系列动画片而言,当缩略视频的时间占到全部视频长度的 20% 时,缩略视频即能表现全部视频的大概意思。本文也对其他类型的动画视频做了类似实验,基本上缩略视频的时间都可以控制在整个视频长度的 30% 以内。



图 3 片段中包含的场景



图 4 场景中包含的镜头



图 5 镜头的代表帧



图 6 故事板形式的视频摘要

4 结束语

本文针对动画视频的特点,提出一种适合动画视频摘要的方法,最终获得故事板和缩略视频形式的视频摘要,满足视频摘要的缩略性和可理解性的要求。本文没有分析字幕、语音这些动画视频的重要组成部分,这些也可能成为下一步研究的契机。随着动漫产业的快速发展,动画视频摘要有极大的商业应用前景,如何集成现有技术到动画视频

6 表示 GMM 参数迭代估计阶段(2 为初始化 GMM,3~6 为更新 GMM),7 表示最小割算法切割原始图像阶段。

由分割结果如图 2(b)和(c)可知,两个算法分割结果相当,可见本文算法继承了 GrabCut 算法精确分割的优点。由图 2(d)可进一步看出,两个算法在 GMM 参数迭代估计阶段的时间消耗分别为 1.741 s 和 55.292 s,占各自算法总耗时的 24.01% 和 92.01%,该阶段的耗时在 GrabCut 算法总耗时中占据绝大多数比例,而在本文算法中变为了次要位置;在算法总耗时上,本文算法仅为 GrabCut 算法的 12.06%,可见,本文算法已经极大地提高了 GMM 参数迭代估计的效率,解决了 GrabCut 算法效率问题的核心矛盾,使得整个算法效率得以显著提高。

再从伯克利图像分割测试图库中选取一组图像,图像大小均为 481 × 321,分别用本文算法和 GrabCut 算法进行对比分割测试,其中 GMM 参数迭代估计进行四次,分割耗时结果如表 1 所示。实验数据进一步表明了本文算法的有效性和分割效率的显著提高,而且耗时比平稳,表明了本文算法针对各种不同图像进行切割有着比较稳定的改进效果。

表 1 图像分割耗时对比

图像编号 (文件名.JPG)	本文算法 /s	GrabCut 算法/s	耗时比/%
3096	1.312	6.953	18.87
124084	1.953	8.547	22.85
113016	1.844	11.000	16.76
163014	1.218	6.094	19.99
227092	1.656	9.782	16.93

3 结束语

图像分割是图像工程中从图像处理到图像分析转变的关键,也是一种基本的计算机视觉技术,因此,快速准确分割图像是图像工程及计算机视觉领域的一个基础课题。GrabCut 算法海量的像素级处理单元以及为达到一定分割精度而采用的迭代求解模式,使得算法分割效率不高。由于 GrabCut 算法效率问题的核心矛盾在于确定 GMM 参数,本文在研究分析基于图割的 GMM 参数迭代估计过程的基础上,引入多尺度分析方

法,以塔式分解的多尺度图像序列代替固定尺度的原始图像序列估计 GMM 参数,理论分析与实验结果表明了本文算法的正确、可行及有效性。

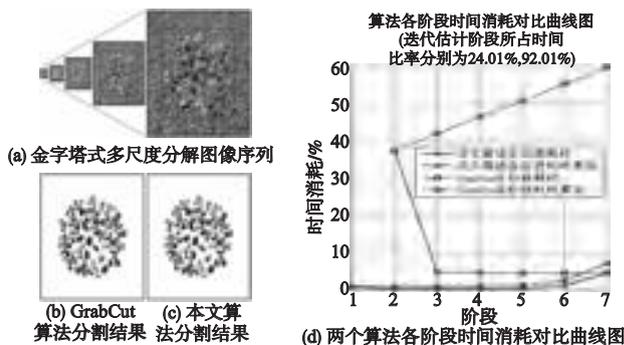


图 2 GrabCut 和本文算法分割图像的实验结果对比

参考文献:

- [1] BOYKOV Y, JOLLY M P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images[C]// Proc of International Conference on Computer Vision. 2001: 105-112.
- [2] BOYKOV Y, FUNKA-LEA G. Graph cuts and efficient N-D image segmentation[J]. International Journal of Computer Vision, 2006, 70(2): 109-131.
- [3] JUAN O, BOYKOV Y. Active graph cuts[C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2006: 1023-1029.
- [4] WU Xiao-yu, WANG Yang-sheng. Interactive foreground/background segmentation based on graph cut[C]// Proc of Congress on Image and Signal Processing. 2008: 692-696.
- [5] ROTHER C, KOLMOGOROV V, BLAKE A. GrabCut: interactive foreground extraction using iterated graph cuts[J]. ACM Trans on Graphics, 2004, 23(3): 309-314.
- [6] BENVENISTE A, NIKOUKHAH R, WILLSKY A S. Multiscale system theory[C]// Proc of the 29th IEEE Conference on Decision and Control. Honolulu: IEEE Press, 1990: 2484-2487.

(上接第 3962 页)摘要系统中,使得动画视频分析与检索系统能够真正满足商业化应用的需求,也是今后研究的重点。

参考文献:

- [1] 王辰,刘桂清,老松杨,等.面向事件的影片摘要生成方法[J].中国图象图形学报,2005,10(5):642-649.
- [2] LI Ying. An overview of video abstraction techniques[R]. Palo Alto: Image Systems Laboratory, HP Laboratory, 2001.
- [3] 贾否,王雷.动画运动[M].北京:中国传媒大学出版社,2005:1-19.
- [4] DEMENTHON D. Video summarization by curve simplification[C]// Proc of ACM Multimedia. 1998: 211-218.
- [5] CHRISTEL M, HAUPTMANN A. Adjustable filmstrips and skims as abstractions for a digital video library [C]// Proc of ADL'99. [S. l.]: IEEE Press, 1999: 98-104.

- [6] LIENHART R, PFFEIFFER S. Video abstracting [J]. Communications of the ACM, 1997, 40(12): 55-62.
- [7] 黄伟.数字资产管理系统中的视频摘要技术[J].广播电视信息,2002,7(1):45-57.
- [8] XIE Le-xing, CHANG SF, DIVAKARAN A. Structure analysis of soccer video with hidden markov models[C]// Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. 2002: 4096-4099.
- [9] MA Yu-fei. A user attention model for video summarization [C]// Proc of ACM Multimedia'02. 2002.
- [10] 朱志辉.基于视频摘要生成技术的研究[J].微电子学与计算机,2006,23(2):76-78.
- [11] 周政,刘俊义,马林华,等.视频内容分析技术[J].计算机工程与设计,2008,29(7):1766-1769.