

## 一种面向动画视频的摘要方法<sup>\*</sup>

袁志民<sup>1+</sup>, 吴玲达<sup>1</sup>, 陈丹雯<sup>1</sup>, 谭洁<sup>2</sup>, 周文<sup>1</sup>

<sup>1</sup>(国防科学技术大学 信息系统与管理学院,湖南 长沙 410073)

<sup>2</sup>(南京政治学院 计算机教研室,江苏 南京 210003)

### A Method for Animation Video Abstract

YUAN Zhi-Min<sup>1+</sup>, WU Ling-Da<sup>1</sup>, CHEN Dan-Wen<sup>1</sup>, TAN Jie<sup>2</sup>, ZHOU Wen<sup>1</sup>

<sup>1</sup>(College of Information System and Management, National University of Defense Technology, Changsha 410073, China)

<sup>2</sup>(Computer Center, Nanjing Politics College, Nanjing 210003, China)

+ Corresponding author: E-mail: yuanzhimin@nudt.edu.cn

Yuan ZM, Wu LD, Chen DW, Tan J, Zhou W. A method for animation video abstract. *Journal of Software*, 2009,20(Suppl.):51-58. <http://www.jos.org.cn/1000-9825/09007.htm>

**Abstract:** In recent years, animation is increasingly rapid developing, and the processing and abstract of animation video have become a hotspot. Animation video abstract is an important step in the research of animation video. Based on distinct property of animation video which is different from that of other videos, especially news video and sports video, a video abstract method, suitable for animation video, is proposed. In this paper, the construction of animation video is analyzed. Then the visual character and clear construction are gotten. The granularity of video scene is defined. Based on content evaluation model of scene significance and The granularity of scene, we find the important scenes of the video. Then, in the temporal order, the video abstract including storyboard and video skim is proposed. Experimental results indicate that the proposed method extracts animation video abstract efficiently. The two forms of video abstracts are produced by this method can generalize and condense the animation video effectively.

**Key words:** animation video abstract; video construction analysis; significance; granularity; storyboard; video skim

**摘要:** 动画视频摘要是目前热门的动画视频处理和检索技术的重要组成部分.针对动画视频与新闻视频、体育运动视频显著不同的特点,提出了一种适合动画视频的摘要方法.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,然后分析动画视频内容的重要度选取视频中的重要片段,最后通过缩略粒度选择,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.实验证明,该方法比较有效地获得了动画视频摘要,两种形式的摘要对动画视频能够很好的概括和缩略.

**关键词:** 动画视频摘要;视频结构分析;重要度;缩略粒度;故事板;缩略视频

\* Supported by the National Natural Science Foundation of China under Grant No.60802080 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2009AA01Z335 (国家高技术研究发展计划(863))

Received 2008-09-20; Accepted 2009-07-17

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的新技术.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<sup>[2]</sup>.同时,动漫产业的快速发展促生了海量的动画视频,与其他新闻、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挑战.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.

视频摘要这方面比较著名的研究,国外有卡内基梅隆大学的Infermedia项目<sup>[4,5]</sup>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;Mannheim大学的MoCA项目<sup>[6]</sup>是为了自动提取视频的结构和语义内容.国内有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;中科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.而目前对动画视频摘要的研究还属于崭新的课题,国内外对这方面的研究还比较少.

因此,本文比较动画视频与其他普通视频(特别是新闻视频和体育视频)的结构特点,研究了动画视频独有的特征,提出了适合动画视频摘要的技术.

## 1 动画视频的结构特点分析

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<sup>[3]</sup>.将动画视频与目前主要用于视频检索的新闻视频和体育运动视频进行结构比较,从而获得动画视频的结构特点.

通常一段视频数据可以划分为几个场景,每个场景又包含几个镜头.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4个层次结构:故事层、场景层、镜头层和图像帧层.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,故对动画视频的结构分析弱化故事层.

目前的视频摘要技术研究主要针对于新闻视频和体育运动视频,主要是这两类视频具有明显的结构特征.再将动画视频与这两种视频相比较,得到动画视频的特有的特点.

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<sup>[7]</sup>,有时中间会附加广告镜头.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.图1是对凤凰卫视的“时事快报”的关键帧提取部分结果,图1(a)是新闻视频的开头.对其结构分析主要在于检测出单独的播音员镜头(如图1(b)所示)与新闻镜头(以及广告镜头).通过播音员帧的图像特点(端坐的播音员)来检测播音员镜头,而新闻镜头一般在播音员镜头后颜色转换比较慢的画面(如图1(c)(d)(e)所示),同时根据广告片段具有的强烈的视觉效果判断广告镜头.



Fig.1 Some frames of news video

图1 新闻视频的抽取帧

体育运动视频的视频摘要一般是对精彩镜头摘要(highlight summary).体育运动视频会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者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者没进球惋惜的镜头伴有观众的欢呼声音和解说者快速解说的声音,并且显示

射门者的基本资料(如图 2 所示),同时回放这个射门的慢镜头<sup>[8]</sup>(如图 2(c)所示),有时候还会有一些重大赛事的标志性图案的镜头.



Fig.2 Some frames of soccer video

图 2 体育视频的抽取帧

动画视频一般具有叙事的情节,除了拥有一般视频的基本特征外,相对于新闻视频、体育视频,其自身特有的结构特征:

(1) 动画视频中的字幕一般是视频中人物的语言或心理表达,不具有像新闻视频的概括一段新闻的标题式字幕,如图 3(a)、图 3(b)所示,因此仅采用字幕的分析不能很好的分割动画视频的镜头.

(2) 视频结构不明显,其内容一般是叙事性的情节,镜头之间的内容顺序变化或者内容相关的联系,一般不会出现如新闻视频中单独的播音员镜头和新闻镜头,或体育视频中的运动员特写镜头、慢动作回放镜头以及标志物镜头,如图 3(a)、图 3(b)所示,动画视频的镜头边界从突变和渐变两个方面来考虑,在第 2.1.2 节中作具体分析.

(3) 动画视频中为了表现其夸张的效果会有一些很长的镜头颜色单调、太亮或太暗等的镜头,有时候则反之会有一些场景变化很快的镜头,如图 3(b)、图 3(c).

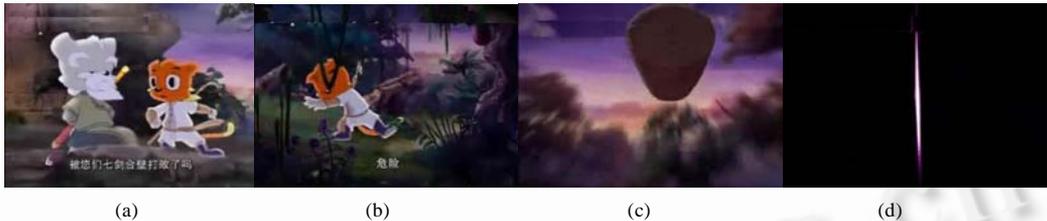


Fig.3 Some frames of animation video

图 3 动画视频的抽取帧

这些特征给动画视频的结构分析和视频摘要工作带来了很大的难度.

## 2 动画视频摘要的生成

对于动画来说,其视频摘要是以动画的内容分析为基础的,只有对动画视频的内容进行结构分析,才能以比较紧凑的方式表现动画视频的摘要.即是先对视频进行结构分析,再对视频内容片段选取,最后将选取的片段组合,得到动画视频摘要,其过程如图 4 所示.

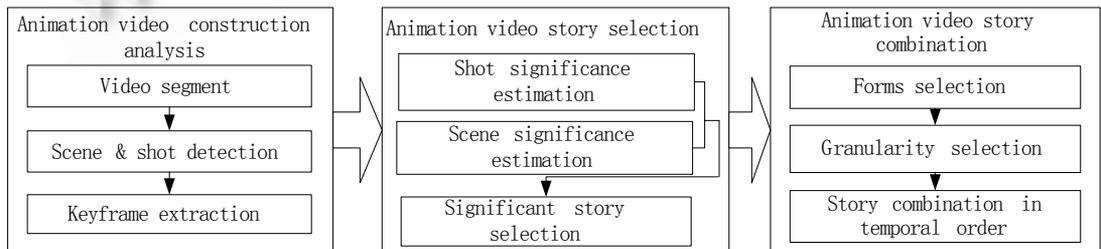


Fig.4 The process of animation video abstract

图 4 动画视频摘要生成

## 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.最终使得动画视频象文本一样便于浏览和检索.其中需要解决的关键技术包括可视特征提取分析、镜头边界检测、关键帧提取等.

### 2.1.1 可视特征提取分析

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.采用颜色对的方法来模型化具有一定对象位置和明显的对象边界的图像.所谓颜色对,就是将一幅图像分成若干小块,对相邻的两个小块,计算它们各自的直方图的灰度平均值,如果两个平均值之间的欧几里德距离大于某一阈值,则认为这两个小块构成颜色对.将获取的可视化特征存储,以备接下一步骤分析所用.

### 2.1.2 镜头边界检测

准确的镜头边界探测结果是进一步视频内容结构分析的重要基础.根据视频的可视特性,分析两个相续镜头处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.镜头突变指两个相邻的镜头之间没有过渡由前镜头的瞬间直接转换为后面相邻镜头的方法,镜头渐变则是镜头之间有渐渐过渡过程,一般由溶化、淡入淡出、和其他特效编辑手法造成.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相续的两帧之间,因此当相续的两帧的视觉内容(帧差)出现巨大变化时,可以判定突变边界的存在.对于渐变镜头,利用经典的“双阈值法”<sup>[11]</sup>的基础上进行渐变过程探测.

### 2.1.3 镜头关键帧选取

镜头关键帧是能反映镜头的整体内容的图像,对镜头内容的分析往往可以转化为对镜头关键帧图像内容的分析.选取关键帧来表示动画视频的主题或部分内容,而不是动态的细节,则关键帧应该具有代表性,避免冗余.本文采取一种与镜头内容变化相关的关键帧选取方法,其要点是:取每个镜头的第一帧为关键帧(若是渐变衔接的镜头则取渐变后的第一帧);对该镜头的其他帧基于颜色的标准逐一检查,如果某一帧与最近检测出的关键帧的差别超过阈值,则选取该帧做关键帧.

## 2.2 动画视频片段选取

动画视频摘要要在结构化分析的基础上,对视频内容的重要信息进行判定,选择概括性强、能够表现主要内容的视频片段(或镜头),组合编辑得到符合需求的视频摘要.因此是否简洁而全面的概括了动画视频的重要信息是判断动画视频摘要的重要标准.基于此,本文定义动画视频摘要的两个重要概念:① 动画视频内容重要度,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;② 动画视频摘要的缩略粒度,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、动画的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.重要的事件是动画中的重要内容,也是推动动画情节的重要手段,特别是视频中能够表现主题内容的激烈运动场面,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.同时,一般较长的场景是在动画视频中更有可能重要,但是场景中的内容变化不大,即使时间长也不应该从中抽取多个关键帧,避免出现视频摘要的冗余.

文献[9]对新闻视频摘要的重要度模型用字幕帧、人脸帧以及镜头关键帧的重要度的线性组合来表示.本文利用动画主要人物和关键帧的重要度来建立动画视频镜头的重要度模型.首先确定镜头的关键帧重要度,通过动画视频帧特征集合中的特征(如颜色、形状等底层特征),加上适当的约束条件,映射到关键帧重要度 $I_{kf}$ 表示为

$$I_{kf} = \text{map}(f_i, g(f_i)), \quad f_i \in Fs^N, i \in \{1, 2, \dots, N\} \quad (1)$$

其中,  $F_s^N$  是镜头的  $N$  维特征集合;  $f_i$  是特征集合中的特征;  $map(,)$  是从特征到重要度的映射;  $g(f_i)$  是对特征的约束条件.

镜头的重要性主要体现在镜头中的动画人物、关键帧的相对重要性以及镜头的持续时间等方面,则其重要度  $I_{Sh}$  表示为

$$I_{Sh} = \lambda_c I_c + \lambda_{kf} I_{kf} + \lambda_{shT} \frac{T_{Sh}}{T_{Sc}} \quad (2)$$

其中,  $I_c$  表示动画人物的重要度,通过其视频主角分析以及出现的次数综合分析获得;  $\lambda_c$ ,  $\lambda_{kf}$  与  $\lambda_{shT}$  分别为动画人物重要度、关键帧重要度以及镜头在场景中持续时间比的权值,且  $\lambda_c + \lambda_{kf} + \lambda_{shT} = 1$ ;  $T_{Sh}$  表示镜头的持续时间;  $T_{Sc}$  表示镜头所在场景的持续时间,  $T_{Sc} = \sum_i (T_{Sh})_i$ .

场景的重要性主要体现在其内镜头的重要度、场景的持续时间以及场景中镜头的数目等方面.则场景重要度  $I_{Sc}$  表示为

$$I_{Sc} = \lambda_{sh} I_{Sh} + \lambda_{scT} \frac{T_{Sc}}{T_{Sf}} + \lambda_{scN} \frac{N_{Sc}}{N_{Sf}} \quad (3)$$

其中,  $\lambda_{sh}$ ,  $\lambda_{scT}$ ,  $\lambda_{scN}$  分别为镜头重要度、场景在动画视频片段中持续时间比以及场景中镜头数目与视频片段中镜头数目比的权值,且  $\lambda_{sh} + \lambda_{scT} + \lambda_{scN} = 1$ ;  $T_{Sc}$  表示动画视频中当前场景的持续时间;  $T_{Sf}$  表示动画视频片段的持续时间,  $T_{Sf} = \sum_i (T_{Sc})_i$ ;  $N_{Sc}$  表示动画视频中当前场景中镜头的数目;  $N_{Sf}$  表示动画视频片段的镜头数目,  $N_{Sf} = \sum_i (N_{Sc})_i$ .

通过分析场景的重要度来判断视频片段的重要性,依据公式(3)计算场景的重要度,并且判断镜头的重要度,最后从场景中的镜头中抽取图像帧,形成动画视频片段的选取.

### 2.3 动画视频片段组合

从最终表现形式上来看,视频摘要包括了静态的视频摘要和动态的视频摘要.对于动画视频来说,静态摘要的故事板和动态摘要的缩略视频比较适合做动画视频摘要.而视频摘要效果主要看摘要的概括性、缩略性以及可理解性.视频摘要在可理解,能概括整个视频内容的基础上尽可能的缩短.即使是重要的镜头,也没有必要把它完全包含到摘要,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是一种冗余;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.

#### 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:首先,确定需要多少幅来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;同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,具体如下式:

$$N(Sc)_k = \frac{T \cdot n \cdot (I_{Sc})_k}{\sum_k (I_{Sc})_k} \quad (4)$$

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.求和时  $k$  取整个原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,设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,只需简单的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形式.其生成策略如下:首先根据确定下来的缩略粒度,以及场景的重要度,判断场景在摘要中的时间,如下式:

$$T(Sc)_k = \frac{T \cdot (I_{Sc})_k}{\sum_k (I_{Sc})_k} \quad (5)$$

计算镜头的重要度,同时根据粒度设定镜头重要度阈值,来获取相应的镜头来源;同时考虑到镜头不能变化

太快,将小于 4 秒的镜头全取,而大于 4 秒的镜头根据关键帧重要度来分析.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

利用测试模型验证该动画视频摘要技术的可行性.测试环境是 Pentium Core2 1.8GHz 处理器,2G 内存;利用 VC++编写的测试模型,对长 14.5 分钟、MPEG4 标准压缩的动画视频《虹猫蓝兔七侠传》第 1 集的片段测试,用时 3'12",得到 45 个场景和 116 个关键帧.其中,动画视频分析采用 24 维颜色直方图作为特征来描述帧的内容,距离度量采用直方图相交距离,进行关键帧的提取和场景的检测.

图 5、图 6 分别是得到的部分场景、关键帧和最终的故事板.其中,故事板能基本表现视频,概括出视频的故事情节.



Fig.5 Key frames from experiment

图 5 实验得到的场景关键帧



Fig.6 Storyboard from experiment

图 6 实验得到的故事板

缩略视频部分,依据经验值选择原视频的 20%,得到缩略视频长度 2'52".为了评估缩略视频的摘要质量,我们请 10 位不同年龄、不同专业、不同爱好(爱好动画片与否)的评估人员对得到的缩略视频进行评价.评价的指标包括:① 概括性,反映整个视频内容的程度,② 缩略性,对视频的缩略程度,冗余度比较低,③ 突出性,重要内容突出程度.评估人员先观看整个视频,然后在观看缩略视频,并对各个部分打分.满分为 10,最低分为 0,最后取 10 人打分的平均值,获得的评价结果见表 1.

由表中可以发现,缩略视频能较好的体现缩略视频的概括性、缩略性和突出性.在 1、5 片段中由于场景变换不明显造成选取帧过少,不够突出;而在 6、7 片段中,由于包含了大量的动作,场景转换过快,并且夸张效果明显,大部分帧保留造成缩略性变低.还有,选取帧的过程过渡帧一般不会选取,造成在部分镜头转换的时候显得生硬.

**Table 1** Estimating the important part of video skim from experimentation**表 1** 实验得到的缩略视频主要内容的评分

编号	主要内容	概括性的打分	缩略性的打分	突出性的打分
1	虹猫在树林歌唱	6.3	8.0	5.0
2	虹猫和麒麟打闹	8.3	7.9	7.6
3	虹猫和大象戏水	8.2	8.0	7.7
4	突来袭击	9.0	8.0	9.3
5	虹猫和父亲交谈往事	6.5	8.5	6.7
6	父亲和魔兵交战	8.5	6.6	8.1
7	黑心虎打败父亲	8.3	6.9	9.1
8	麒麟带虹猫逃亡	8.5	7.3	8.3
9	火烧大山	7.8	9.4	7.3

#### 4 结束语

本文利用动画视频与其他视频不同的特点,总结研究出一种适合动画视频摘要的技术,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.同时,本文没有从基于语义的方向研究,由于没有分析字幕、语音这些动画视频的重要部分,所以生成完整、准确、令人满意的视频摘要是比较困难的,下一步要加强这方面的研究.由于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,如何集成现有成熟技术到动画视频摘要系统中,使得动画视频分析与检索系统能够真正商业化应用,也是今后研究的重点之一.

#### References:

- [1] Wang C, Liu GQ, Lao SY, Jiang J. Event-Oriented film abstract. Journal of Image and Graphics, 2005,10(5):642-649 (in Chinese with English abstract).
- [2] Li Y, Zhang T, Tretter D. An overview of video abstraction techniques. Image Systems Laboratory, HP Laboratory Palo Alto, HPL-2001-191, 2001.
- [3] Jia F, Wang L. Action in Animation. Beijing: Communication University of China Press, 2005. 1-19 (in Chinese).
- [4] Dementhon D, Kobla V, Doermann D. Video Summarization by Curve Simplification. ACM Multimedia, 1998: 211-218.
- [5] Christel M, Hauptmann A, Warmack H, Crosby S. Adjustable filmstrips and skims as abstractions for a digital video library. In: Proc. of the ADL'99. IEEE Press, 1999. 98-104.
- [6] Lienhart R, Pfeiffer S, Effelsberg W. Video abstracting. Communications of the ACM, 1997,40(12):55-62.
- [7] Huang W. Techniques of video Abstract in management of digital finance. Information on Broadcast of Television, 2002,7:45-57 (in Chinese with English abstract).
- [8] Xie L, Chang SF, Divakaran A, Sun H. Structure analysis of soccer video with hidden Markov models. In: Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP), Vol.5. 2002,5:4096-4099.
- [9] Ma Y, Lu L, Zhang H, Li M. A user attention model for video summarization. In: Proc. of the ACM Multimedia 2002, Juan-les-Pins. 2002.
- [10] Zhu ZH. Research on video abstraction techniques. Microelectronics & Computer, 2006,23(2):76-78 (in Chinese with English abstract).
- [11] Zhou Z, Liu JY, Ma LH, Chen D. Video content analysis technique. Computer Engineering and Design, 2008,29(7):1766-1769 (in Chinese with English abstract).

#### 附中文参考文献:

- [1] 王辰,刘桂清,老松杨,蒋杰.面向事件的影片摘要生成方法.中国图形图像学报,2005,10(5):642-649.
- [3] 贾否,王雷编著.动画运动.北京:中国传媒大学出版社,2005.1-19.
- [7] 黄伟.数字资产管理系统中的视频摘要技术.广播电视信息,2002,7:45-57.
- [10] 朱志辉.基于视频摘要生成技术的研究.微电子学与计算机,2006,23(2):76-78.
- [11] 周政,刘俊义,马林华,陈东.视频内容分析技术.计算机工程与设计,2008,29(7):1766-1769.



袁志民(1982—),男,河南淮阳人,博士生,主要研究领域为基于内容的多媒体检索技术,监控视频对象行为分析.



谭洁(1971—),女,讲师,主要研究领域为计算机技术.



吴玲达(1962—),女,教授,博士生导师,主要研究领域为多媒体技术,虚拟现实技术.



周文(1983—),男,硕士生,主要研究领域为动画自动生成技术.



陈丹雯(1983—),女,博士生,主要研究领域为多媒体检索技术,图像分类与识别.

www.jos.org.cn

www.jos.org.cn