

新闻视频中广告片段精确定位方法研究

李士进¹⁾ 郭跃飞²⁾ 李昊¹⁾ 秦明举¹⁾ 冯钧¹⁾

¹⁾(河海大学计算机及信息工程学院, 南京 210098) ²⁾(复旦大学计算机科学与工程系, 上海 200433)

摘要 由于广告节目制作风格的多样性,新闻视频中广告片段的检测和定位是一个非常具有挑战性的问题。提出了一种新闻视频中广告片段定位的新方法。首先针对镜头切换和非镜头切换数目不平衡问题,设计了一种新的分类方法进行镜头检测,然后利用聚类分析粗略地标识广告块,最后通过分析相邻镜头的平均持续时间和镜头关键帧的视觉特征准确定位广告边界。实验结果表明,该方法具有较高的定位精度。

关键词 广告片段定位 聚类 时间和视觉特征 镜头检测 不平衡数据

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2009)07-1432-08

Study on Precise Commercial Segments Location in News Video

LI Shi-jin¹⁾, GUO Yue-fei²⁾, LI Hao¹⁾, QIN Ming-ju¹⁾, FENG Jun¹⁾

¹⁾ (School of Computer & Information Engineering, Hohai University, Nanjing 210098)

²⁾ (Department of Computer Science & Engineering, Fudan University, Shanghai 200433)

Abstract The detection and location of commercials in news video has been a challenging problem because of the diversity of the production styles of commercial programs. In this paper, the authors present a novel algorithm for the location of commercial segments in news program. With the proposed method, firstly shot transition detection is conducted, which is based on the imbalanced data classification methodology, and then clustering analysis is employed to label commercial blocks roughly. Finally the accurate boundaries between the commercial segments and news story are located by analyzing a new set of features, including the average duration of preceding and subsequent shots and the visual features of the shots, such as color, saturation and edge distribution. The experiment results show that the proposed algorithm is effective with high precision.

Keywords commercial segments location, clustering, temporal and visual features, shot detection, imbalanced data

1 引言

新闻节目是电视台的重要媒体资产,新闻视频分析吸引了众多的研究人员^[1-3]。在新闻节目中,由于商业利益的驱使,电视台经常会在其中插播一些广告。但是,在诸如视频点播和媒体资产管理等应用中,为了更好地对新闻视频检索和浏览,这些广告块应该被剔除。以新闻节目视频点播为例,用户

常常希望只看到他们所点播的新闻节目而没有插播广告。因此,一个能自动检测、精确定位广告片段的系统具有很高的实用价值。

近年来,研究人员提出了很多广告视频检测的算法。这些方法大致可分为两大类:一类方法是利用特定电视台节目播出的先验知识,例如黑帧^[4]、电视台台标^[5]等作为广告检测的线索,这种方法通常具有很高的检测准确率,但是只能应用于特定电视台的视频;另一类方法主要是利用基于内容的特

基金项目:国家自然科学基金项目(60673141)

收稿日期:2008-08-13;改回日期:2008-10-05

第一作者简介:李士进(1973~),男,副教授。2000年于南京理工大学获模式识别与智能系统专业博士学位。主要研究领域为图像/视频分析、模式识别。E-mail: lishijin@hhu.edu.cn

征来标记新闻视频中的广告块,这些特征通常包括视觉的^[6-7],听觉的^[8]和纹理^[9]信息,从而可以运用到不同风格的新闻视频。文献[10]提出了一种基于视觉、音频及其派生特征的基于机器学习的广告检测方法。Duan等人提出了一种多模态分析算法来分割广告片段,并且进行了高层语义分析^[11]。张亮等人提出了一种基于镜头的鲁棒视频广告检测方法^[12],在对电视节目镜头分割的基础上,提取出一些镜头统计特征(如镜头切换率,帧差的均值和方差,以及帧的主颜色空间占有率等),然后利用支持向量机(SVM)将镜头分成普通节目镜头和广告镜头,最后通过引入后处理机制,实现对错分镜头的纠错,将广告镜头整合成广告视频段。刘宇驰等人研究了基于探测视频流中重复镜头的方法进行新闻视频中非新闻段的去除^[13],取得了较好的效果。

文献[6]提出了一个广告片段检测定位算法,该算法首先利用滑动窗口分析了在很短时间内有剧烈切变的镜头从而得到候选的广告块,然后用视频场景探测器来标识它们的边界,这一步有时需要人工干预。与文献[6]相比,本文方法是基于对新闻和广告节目详细的结构分析和内容分析。

目前很多地方电视台在播出新闻节目时,都会在节目中插播数目不等、长度不等的广告。以江苏省某地方频道的新闻节目为例,在每一个新闻报道之前都有一个主持人镜头来对新闻做一个简单介绍,由一系列广告组成的广告块有可能插播在某个新闻报道和主持人镜头之间。本文主要针对这类节目进行研究,希望能在完整长度的新闻节目中自动标记定位所有插播的广告块,并且利用新闻节目和广告片段的时间和视觉特征差异来准确定位广告块的边界。本文的主要贡献在于:一是提出了基于不平衡数据分类思想的镜头检测新方法;二是在粗定位广告片段时,通过对镜头持续时间的聚类,可以估计新闻镜头和广告镜头的统计特征,从而可以避免人工调整阈值;三是在识别广告片段时,首次引入了两种新的视觉特征。以往的很多研究都没有充分考虑视觉特征,而这些视觉特征可以提高广告片段检测定位的精度。

本文提出的新闻视频中广告片段检测新方法分为粗标记和精确识别两个阶段。在粗标记广告块阶段,首先假定在新闻节目中,所有镜头都可以分成几组,然后通过全局聚类算法标记可能的广告块,这一步可以克服文献[6]所带来的手工阈值调整问题。

为了准确标记候选广告块的边界,我们视其为镜头模式识别问题:广告类镜头和新闻报道类镜头。在精确识别阶段,提取了5种特征,其中包含了两种新的视觉特征。

2 镜头检测

镜头边界检测是基于内容视频检索中的第1步,在视频分析中扮演着重要角色^[14-16]。镜头变换一般分为两类:突变(亦称切变)和渐变(如淡入淡出,溶解,擦除等)。目前研究人员已经提出了很多方法,但尚未完全解决镜头检测问题^[17]。现有文献中提到的方法主要有两大类:基于阈值的方法和基于机器学习的方法。在基于阈值的方法中,如何调整阈值对算法的成功至关重要,因此普通的工程技术人员很难将现有文献中提到的方法付诸实现。最近,亦有研究人员提出利用机器学习的方法实现镜头检测,如隐马尔可夫模型(HMM)^[18]和支持向量机^[19-22]。Zhang等人提出了一种基于HMM的镜头检测方法^[18],该方法为不同的视频变换建立各自的HMM模型。Ling等人提出了一种基于SVM的渐变镜头检测方法^[19],该方法基于视频帧的方差投影函数提取特征。Ren和Singh提出一种基于K-最近邻分类器和神经网络的方法^[20]来预测镜头变换,该方法对包含18个特征如宏块系数和神经元系数的特征集合进行分类。Cao和Cai提出了一种压缩域中基于SVM的快速镜头检测算法^[21],利用一个滑动时间窗所包含的所有帧形成特征向量,每一帧由6个压缩域的特征来表示。该算法虽然具有很高的精度,但受压缩算法和视频编码质量的限制较大。Ngo提出利用时间切片中的纹理特征进行“溶解”检测^[22]。

上述基于机器学习的方法主要用来解决渐变检测的问题,而切变检测一般仍采用基于阈值的方法^[23]。针对新闻视频中的广告检测,笔者基于统一的机器学习策略,设计了一种新颖的镜头分类方法^[32]。该方法由两部分组成,第1步,利用SVM将镜头分为无场景变化、切变、渐变(和快速运动)3类。之所以这样划分,是因为切变代表了相邻帧的显著变化,无场景变化则表明相邻帧的内容基本不变,而包括了渐变和快速运动的第3类,由于其反映了连续帧之间更大的变化,因此成为以往的研究重点。第2步,利用与第1步不同的特征,解决快速运

动和渐变的分类问题。除了以上提到的两步分类策略,还研究了以往工作中很少涉及的不平衡数据的分类问题。

2.1 完整的分类步骤

新闻视频包括频繁的快速镜头移动、大段含有剧烈运动的广告、几乎静止的演播室和主持人镜头等一系列镜头种类。针对新闻视频和新闻故事片段中的广告检测,考虑 3 种主要的变换形式:切变、淡入/淡出、溶解。

用两层分类器来检测不同类型的镜头变换:无场景变化、切变和渐变。第 1 步,将镜头分为无场景变化、切变和大场景变化——包括渐变和快速运动。之所以这样划分,是因为发生渐变和快速运动时,最长公共子串(LCSS)特征^[24]序列和直方图相似性序列在表现上是相同的,并不能很好地对其区分。以往的研究人员为此做了大量的工作,以降低假阳性(false alarm)并提高渐变的检测精度。第 2 步,根据颜色矩(color moments)特征序列将渐变从快速运动情况中分离出来。完整的步骤如图 1 所示。由于 SVM 在处理小样本问题上表现出的良好能力,采用其作为本文的分类器。使用 Libsvm^[25]对样本进行分类。

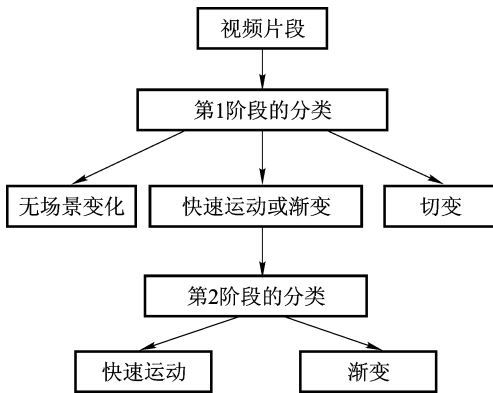


图 1 镜头检测的完整分类步骤

Fig. 1 The complete scheme for shot detection

2.2 不平衡样本分类问题

在新闻视频中,通常切变和渐变镜头并不多,最多的是无场景变化类的镜头。因此,在机器学习过程中会导致不平衡样本的分类问题^[26]。已有工作中训练样本是由人工选择的^[19-21],并没有考虑到不平衡样本分类问题。Lienhart 试图采用渐变合成的方法来解决该问题^[27],这种合成能够产生上千个不同的视频变换样本。然而,这些样本并非是节目中

真实存在的,而且在新闻节目中也并不存在这么多类型的变换。采用一种串行聚类的方法来减少无场景变化类样本。

为了使正反例样本达到平衡,通常采用重采样的方法^[26],包括对多数样本的下采样和对少数样本的过采样。如果过采样切变或者渐变,由于对训练样本的过度逼近^[26],过拟合的可能性就会增加。但是,随机的下采样也有缺点,当抛弃大量有效样本的同时,一些对分类有用的信息也随之丢失。按照 Chawla 等人的观点^[26],在重采样中,对多数样本的下采样优于对少数样本的过采样。因此,采用下采样方法来解决镜头检测中的不平衡样本分类问题。另外,采用串行聚类来代替随机的下采样,能保留更多对分类有利的信息。

由于视频中只有少量的镜头变换,所以可以在第 1 阶段分类中减小无场景变化类和大场景变化类中样本的规模,在第 2 阶段分类中减小快速运动类中样本的规模。本文采用串行聚类算法^[28-29]。该算法的主要思想如下:逐个输入训练样本,对任意一个新样本,如果它到已经存在的簇中心的最小距离大于预先设定阈值,就增加一个新簇;否则,它就属于离它最近的中心所在的簇。该算法要预先确定一个阈值。这个阈值基于最后得到的簇数量,在不同阶段分类中,这些簇所对应类别的样本数量应该大致相等。

3 新闻视频中广告片段定位

如图 2 所示,提出一种分 2 步逐步求精的检测算法。首先,使用上一节提出的方法进行镜头检测,提取每一镜头的中间一帧作为关键帧。文中所处理

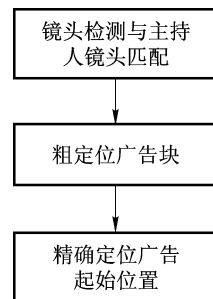


图 2 新闻视频中广告片段定位算法框架

Fig. 2 The framework for commercial segments location in news program

的新闻视频是一个由主持人镜头开始的完整新闻节目,所以定义第1个镜头的关键帧作为模板,而后与后续每一个镜头关键帧进行匹配从而检测所有的主持人镜头。

然后,初步标记广告块并通过聚类分析获得候选广告块集合。最后,使用SVM来识别候选广告块集合中各个镜头的时间和视觉特征,从而准确标记广告块起始位置。

3.1 初步标记广告块

3.1.1 镜头的全局聚类

文献[6]使用滑动窗口来计算参数“一分钟内镜头切变数”(镜头持续时间的倒数)。最后通过设定一个阈值将连续的窗口进行标记从而得到候选广告块。本文中,仍然考虑镜头的持续时间,因为它是广告镜头区别于新闻镜头的一个非常重要的特征。但是,这里采用全局聚类的方法而非固定阈值来标记广告块。

首先,假定新闻视频中的镜头可以分为主持人镜头、新闻报道镜头、广告镜头和其他的诸如天气预报和新闻预告镜头等。也就是说,主持人镜头构成集合A,新闻报道镜头构成集合B,广告镜头构成集合C,其他的镜头构成集合D。

这4组镜头之间应该是可以相互区分的:A中镜头的持续时间最长;B中镜头是连续分布的,但是镜头的持续时间却长短不一;C中镜头也是连续分布的并且镜头的持续时间较短;D中的镜头持续较长的时间,并且相互之间持续的时间较为相似。

对于整个视频,使用一个长度为 T_w 的滑动窗口,滑动的速度设为 $T_w/2$ 。这里 T_w 设为30s。计算每一时刻滑动窗口内镜头的数目、镜头持续时间的均值和方差。这些窗口的属性数据构成了聚类所用的特征向量。

最后,使用K-均值聚类算法将每一个窗口的镜头分成4类,分别对应到A、B、C、D 4个集合。

3.1.2 标记广告块

聚类以后,使用一种近似的方法初步标记广告块:

M 个连续的窗口将被标记成一个广告种子,如果以下条件成立:

- (1)在 M 个连续的窗口中没有属于A的镜头。
- (2)在 M 个连续的窗口中大部分镜头都属于C,这里设该比例为80%。

由于大部分广告块持续时间约为2~3分钟,所

以 M 的取值应该大于等于 $240/T_w$ 。

然后,对先前得到的主持人镜头集合和广告种子集合求交集。位于两个主持人镜头之间的广告块可以被初步标记。被标记的广告块的第1个镜头认为是候选块的起始位置,广告种子的最后一个镜头位于候选广告块的最后位置。

3.1.3 确定候选广告块起始位置

通过聚类分析,能够识别位于两个主持人镜头之间的广告块,并且分别获得新闻报道镜头和广告镜头持续时间的统计规律,并将所有镜头聚成4类。

在新闻节目中,插播的广告可以位于新闻报道之后,也可以位于主持人镜头之间。由于广告镜头持续的时间很短,位于主持人镜头之间的大部分滑动窗口将被标记成候选广告块。所以,在这种情况下直接将这样的块分类成广告。下面主要解决位于新闻报道之后插播广告片段的精确定位。

新闻镜头和广告镜头持续时间的统计特征可以通过上文全局聚类算法获得。令 D_n 代表新闻镜头的平均持续时间, D_a 代表广告镜头的平均持续时间。可以设置一个阈值 t 来确定待分类的镜头:

$$t = D_n \times p + D_a \times (1 - p) \quad (1)$$

其中, $p \in [0, 1]$,反映镜头持续时间相对于新闻镜头的偏移程度,这里设为0.6。

本文的启发式规则是从候选广告块的最后一个镜头开始向前搜索直到遇到一个镜头,它的持续时间大于阈值 t 。该镜头的下一个镜头就作为候选广告块的起始位置。这样可以大大减小参与分类的镜头数目。通常,只有1~3个新闻镜头需要送入分类器,这样能够消除很多新闻报道中间短镜头的错误判断。

3.2 精确定位广告块边界

广告块的粗略位置被检测出来后,下一个问题转化为寻找候选广告块中第1个广告镜头的位置作为广告块的起始位置。

新闻报道和广告都有自己相应的风格。当一个广告块开始的时候,先前的视觉风格就会消失,伴随而来的是一种新的具有很高镜头切变频率的风格。也就是说在两种视觉风格之间有一个镜头切变。本文提出5种特征来识别这种切变,其中的两种特征是在广告检测中本文首次提出。

前两个特征是镜头持续时间的比率。计算两个值:当前镜头和前面若干镜头持续时间的比率,当前

镜头和后续若干镜头持续时间的比率。比率值依赖于在广告镜头之前是一个新闻报道镜头还是一个广告镜头。我们知道新闻镜头的持续时间一般较长,而广告镜头的持续时间一般较短。由于我们处理的是综合新闻节目,每段新闻报道都很长,一般有 20 多个镜头。设定在当前镜头之前之后有 $k=10$ 个镜头来计算持续时间的比率,具体计算方法如下:

$$F_1 = \frac{\text{当前镜头持续的时间}}{\text{前 } k \text{ 个镜头的平均持续时间}} \quad (2)$$

$$F_2 = \frac{\text{当前镜头持续的时间}}{\text{后 } k \text{ 个镜头的平均持续时间}} \quad (3)$$

镜头的 HSV 累加直方图由镜头中所有帧的直方图累加得到。相比单一关键帧直方图而言,它更强调整个镜头的总体颜色特征。因此,为了更好地度量两个镜头之间的相似度,将镜头累加直方图之间的相似度而非关键帧的直方图作为第 3 个特征来表示颜色信息。相应的直方图和前面若干镜头的最大相似度记为 F_h 。

镜头持续时间的比率和累加直方图相似度并不能很好区分广告和新闻镜头,因为很可能出现广告的第 1 个镜头与前面新闻报道镜头在颜色和持续时间上都很相似,所以又提出了两个额外的特征来加强区分度。

在很多广告中,为了给观众留下深刻印象,其画面颜色是生动活泼的,计算 HSV 空间中 S 分量的平均值和亮度值大于 0.4 的具有高饱和度的像素所占的比率,记为 F_s 。这些特征在文献[30]中被用来检测电视节目中的卡通片,在文献[31]中用来区分油画和照片。广告镜头关键帧的 F_s 值较大。

在文献[31]中,亮度边缘和彩色边缘的比率被用来从照片中区分油画。颜色信息在广告画面中扮演了更重要的角色。采用 Canny 边缘检测算子分别在亮度通道和 RGB 通道中提取边缘信息。最终该特征的定量计算如下:

$$F_e = \frac{\text{亮度边缘而非彩色边缘像素数}}{\text{总边缘像素数目}} \quad (4)$$

广告镜头关键帧的 F_e 值比新闻镜头关键帧的 F_e 值小。

最后,对于候选广告镜头得到一个特征向量: $(F_1, F_2, F_h, F_s, F_e)$ 。

下一步,同样使用 SVM 来对这些特征进行分类。对于广告和新闻的分类问题,一段完整的新闻节目中有很少的广告片段,一般在 1 h 的新闻节目中插播 3~4 个广告片段。因并不需要也不可能收集成千上万的视频来训练分类器。所以,从一些真实的新闻节目中合成“新闻+广告”片段作为训练样本。每一个新闻报道后面都跟着一个广告片段,从而得到一个既有新闻又有广告片段。

4 实验结果

测试数据来自江苏省某地方电视台的 4 段新闻节目,每段新闻节目持续时间大约 50 min。训练数据来自另外两段不同时间的新闻节目。

首先进行镜头检测,利用本文第 2 节提出的镜头检测方法,对 4 段新闻节目进行镜头检测,切变检测查全率为 98.5%,正确率达 92.9%,渐变检测查全率为 89.4%,正确率为 87.3%,表明本文设计的基于两层分类器的镜头检测方法可以满足后续的广告片段检测和定位的需求。

以上只给出了基于串行聚类重采样的检测结果,与随机下采样以及未重新采样方法的对比实验,可见文献[32]。

为了验证本文提出的两种新特征在区分广告和新闻镜头方面的有效性,从训练视频中随机抽取了 100 幅新闻镜头关键帧和 100 幅广告镜头关键帧,分别提取这 200 幅图像的 F_s 和 F_e 特征,并进行 20 等份量化,每份代表 0.05,统计它们的分布,结果分别如表 1 和表 2 所示。

表 1 广告和新闻画面的 F_s 特征的分布直方图

Tab.1 The histogram of the distributions for feature F_s of commercials and news frame

	量化级别																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
广告	8	8	9	7	11	3	8	5	4	7	4	5	1	2	3	4	3	3	3	2
新闻	51	14	8	4	3	3	2	2	3	4	1	2	1	0	1	0	0	1	0	0

表2 广告和新闻画面的 F_c 特征的分布直方图Tab.2 The histogram of the distributions for feature F_c of commercials and news frame

	量化级别																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
广告	3	0	3	4	6	10	18	10	6	5	4	5	3	4	4	3	0	3	4	5
新闻	0	0	0	1	2	2	5	8	5	6	7	5	11	7	9	6	7	4	7	8

从表1可以看出,在200幅图像中有73%的新闻画面 F_c 小于0.15,而 F_c 大于0.15的广告画面则有75%。从表2可以看出,在200幅图像中有74%的广告画面 F_c 小于0.6,而 F_c 大于0.6的新闻画面则有60%。我们对 F_c 大于0.85的广告画面进行分析,发现大多是产品名称画面,即画面中间只有产品名称,背景为某种纯色,这会导致 F_c 偏高。而在实际检测中,本文并未发生这类误检,原因是这类画面一般出现在产品广告的最后镜头。后续研究也可以通过提取文字区域特征进行剔除。

最后对4段新闻节目进行广告片段检测,结果如表3所示。从表3可知,本文方法可以很准确定位大部分广告。实验中误检的一个镜头是某新闻报道的最后镜头,该镜头是一书面合同的特写,和该新闻报道前面的镜头都不相似,因此被错误识别为广告镜头。漏检的一个镜头是某移动电话的广告,其画面颜色不明显,且和前面的新闻节目镜头也很相似。在实际节目中这种情况很少发生,在我们以后的工作中可以进一步完善。

表3 4段新闻视频中广告片段检测结果

Tab.3 The detection results of the proposed algorithm on four news programs

总帧数	插播广告片段数	漏检数	误检数
79 815	4	0	0
82 499	4	0	1
80 902	4	0	0
79 735	4	1	0

本文方法的时间开销,主要包括镜头检测和广告片段精确定位。本文提出的镜头检测方法主要是解决训练样本不平衡问题,但由于要计算每帧的直方图和 LCSS 特征,和基于直方图的阈值比较方法相比,时间复杂度要高些。实验环境如下: Intel Pentium 4 PC, CPU 2.4 GHz, RAM 512 MB, Windows XP, VC ++ 6.0, 视频解码采用微软 VFW AVIFILE 库。对一段 10 min(600 s)新闻视频(AVI 文件),提

取每帧直方图特征共需要 490.98 s,而加上 LCSS 特征后,时间略有增加,共需要 548.97 s,可见时间大部分被视频解码占用。切变检测 SVM 分类器用时 15.37 s,后续的渐变分类是在前期切变检测基础上进行的,需要提取颜色矩特征的帧很少,这部分占用时间不到 10 s。在类似实验环境下,文献[33]报道的双阈值镜头检测方法所用时间与视频持续时间长度相当。后续的广告片段定位只对镜头边界进行,一段 50 min 左右完整新闻视频一般有 900~1 000 个镜头,提取滑动窗口特征后参与粗定位聚类分析只有 300 个左右样本,用时不到 0.1 s,精确定位阶段需要提取关键帧特征的镜头数一般不超过 50 个,因此整个广告片段定位所需时间和镜头检测时间相比几乎可忽略。相对人工观看监测,本文方法的时间复杂度可以被用户所接受。

5 结论

本文提出了一种新的模式识别方法来定位新闻视频中的广告,主要是基于时间和视觉特征分析,包括镜头的持续时间、颜色、饱和度和图像边缘统计特征。由于这些特征在广告中不容易被改变,因此该方法可以适用于很多类型的新闻节目。实验表明本文提出的方法有较好的应用前景。

在后续研究中,可进一步考虑将本文方法推广至其他类型的电视节目,如电视剧;另外也考虑加入一些其他的特征进一步提高算法的定位精度,例如音频特征和文字特征。

致谢 感谢江苏省广播电视总台张海勇提供本文实验所需视频。

参考文献 (References)

- Jiang Fan, Zhang Yu-jin. News video indexing with scene sectioning and summary generation [J]. Chinese Journal of Computers, 2003, 26(7): 859-866. [姜帆, 章毓晋. 新闻视频的场景分段索引及摘要生成 [J]. 计算机学报, 2003, 26(7): 859-866.]

- 2 Xie Yu-xiang, Luan Xi-dao, Wu Ling-da, *et al.* NVPS: A multimodal news video processing system [J]. *Journal of The China Society for Scientific and Technical Information*, 2004, **23**(4):404-409. [谢毓湘, 栾悉道, 吴玲达等. NVPS: 一个多模态的新闻视频处理系统 [J]. *情报学报*, 2004, **23**(4):404-409.]
- 3 Yu Jun-qing, Tang Yang, Zhou Xiang-dong. Detecting anchorperson frame from the news video using the matching of dominant color template [J]. *Journal of Computer Aided Design & Computer Graphics*, 2005, **17**(3):558-562. [于俊清, 汤畅, 周向东. 利用主色模板匹配检测新闻视频口播帧 [J]. *计算机辅助设计与图形学学报*, 2005, **17**(3):558-562.]
- 4 Lienhart R, Kuhmunch C, Effelsberg W. On the detection and recognition of television commercials [A]. In: *IEEE International Conference on Multimedia Computing and Systems [C]*, Ottawa, Ont, Canada, 1997: 509-516.
- 5 Albiol A, Fulla M J, Albiol A, *et al.* Detection of TV commercials [A]. In: *IEEE International Conference on Acoustics, Speech and Signal Processing [C]*, Montreal, Canada, 2004: 541-544.
- 6 Yeh J, Chen J, Kuo J, *et al.* TV commercial detection in news program videos [A]. In: *IEEE International Symposium on Circuits and Systems [C]*, Kobe, Japan, 2005: 4594-4597.
- 7 Sánchez J M, Binefa X. Local color analysis for scene break detection applied to TV commercials recognition [A]. In: Huijsmans D P, Smeulders A W, eds. *Lecture Notes In Computer Science [C]*, London: Springer-Verlag, 1999, **1614**: 237-244.
- 8 Duygulu P, Chen M Y, Hauptmann A. Comparison and combination of two novel commercial detection methods [A]. In: *Proceedings of IEEE International Conference on Multimedia and Expo [C]*, Taipei, China, 2004: 1267-1270.
- 9 Hauptmann A, Withrock M. Story segmentation and detection of commercials in broadcast news video [A]. In: *Proceedings of the Advances in Digital Libraries Conference [C]*, Santa Barbara, CA, USA, 1998: 168-179.
- 10 Hua X S, Lu L, Zhang H J. Robust learning-based TV commercial detection [A]. In: *Proceedings of IEEE International Conference on Multimedia and Expo [C]*, Amsterdam, Netherlands, 2005: 149-152.
- 11 Duan L, Wang J, Zheng Y, *et al.* Segmentation, categorization, and identification of commercial clips from TV streams using multimodal analysis [A]. In: *Proceedings of the 14th ACM international Conference on Multimedia [C]*, New York: ACM Press, 2006: 201-210.
- 12 Zhang Liang, Zhu Zhen-feng, Zhao Yao, *et al.* Video commercial detection based on the robustness of shot [J]. *CAAI Transactions on Intelligent Systems*, 2007, **2**(2):83-88. [张亮, 朱振峰, 赵耀等. 基于镜头的鲁棒视频广告检测 [J]. *智能系统学报*, 2007, **2**(2): 83-88.]
- 13 Liu Yu-chi, Luan Xi-dao, Xie Yu-xiang, *et al.* Removing non-news parts from news video [J]. *Journal of Chinese Computer Systems*, 2007, **28**(10):1837-1841. [刘宇驰, 栾悉道, 谢毓湘等. 新闻视频中非新闻段的去除 [J]. *小型微型计算机系统*, 2007, **28**(10): 1837-1841.]
- 14 Han Bing, Gao Xin-bo, Ji Hong-bing. A shot boundary detection method for news video based on rough sets and fuzzy clustering [J]. *Journal of Image and Graphics*, 2007, **12**(3):522-528. [韩冰, 高新波, 姬红兵. 基于粗糙集和模糊聚类的新闻视频镜头边界检测方法 [J]. *中国图象图形学报*, 2007, **12**(3):522-528.]
- 15 Wan Chong-wei, Li Wei-ming, Li He-ping. Video shot detection based on scale invariant features [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2007, **19**(9): 1094-1099. [万崇玮, 李炜明, 李和平等. 基于尺度不变特征的视频镜头检测 [J]. *计算机辅助设计与图形学学报*, 2007, **19**(9):1094-1099.]
- 16 Cheng Yong, Xu De. A method for shot boundary detection using adaptive threshold [J]. *Acta Electronica Sinica*, 2004, **32**(3): 508-511. [成勇, 须德. 一种自动选取阈值的视频镜头边界检测算法 [J]. *电子学报*, 2004, **32**(3):508-511.]
- 17 Hanjalic A. Shot-boundary detection: unraveled and resolved? [J]. *IEEE Transactions on Circuits System Video Technology*, 2002, **12**(2): 90-105.
- 18 Zhang W, Lin J, Chen X, *et al.* Video shot detection using hidden Markov models with complementary features [A]. In: *Proceedings of the First International Conference on Innovative Computing, Information and Control [C]*, Beijing, China, 2006: 593-596.
- 19 Ling Jian, Lian Yi-qun, Zhuang Yue-ting. A new method for shot gradual transition detection using support vector machine [A]. In: *Proceedings of International Conference on Machine Learning and Cybernetics [C]*, Guangzhou, China, 2005: 5599-5604.
- 20 Ren Wei, Singh S. Automatic video shot boundary detection using machine learning [A]. In: *Proceedings of 5th International Conference on Intelligent Data Engineering and Automated Learning [C]*, Exeter, UK, 2004: 285-292.
- 21 Cao Jian-rong, Cai An-ni. A robust shot transition detection method based on support vector machine in compressed domain [J]. *Pattern Recognition Letter*, 2007, **28**(12): 1534-1540.
- 22 Ngo C. A robust dissolve detector by support vector machine [A]. In: *Proceedings of the Eleventh ACM International Conference on Multimedia [C]*, New York: ACM Press, 2003: 283-286.
- 23 Lu Hong, Tan Yap-peng. An effective post-refinement method for shot boundary detection [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2005, **15**(11): 1407-1421.
- 24 Bezerra F N, Leite N J. Using string matching to detect video transitions [J]. *Pattern Analysis and Application*, 2007, **10**(1): 45-54.
- 25 Chang Chih-chung, Lin Chih-jen. LIBSVM: a library for support vector machines [CP/OL]. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- 26 Chawla N V, Hall O, Joshi A. Wrapper-based computation and evaluation of sampling methods for imbalanced datasets [A]. In: *Proceedings of the 1st International Workshop on Utility-based Data Mining [C]*, Chicago, Illinois, USA, 2005: 24-33.
- 27 Lienhart R. Reliable dissolve detection [A]. In: *SPIE Storage and Retrieval for Media Databases [C]*, San Jose, CA, USA, 2001:

- 219-230.
- 28 Li Bin, Chi Ming-min, Fan Jian-ping, *et al.* Support cluster machine [A]. In: Proceedings of the 24th International Conference on Machine Learning[C], Corvallis, Oregon, USA, 2007: 505-512.
- 29 Theodoridis S, Koutroumbas K. Pattern Recognition[M]. 2nd ed. Beijing:Academic Press, 2003.
- 30 Ianeva T I, Vries A P, Rohrig H. Detecting cartoons: a case study in automatic video-genre classification [A]. In: Proceedings of IEEE International Conference on Multimedia and Expo [C], Baltimore, Maryland, USA, 2003: 449-452.
- 31 Cutzu F, Riad H, Leykin A. Estimating the photorealism of images: Distinguishing paintings from photographs [A]. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition[C], Madison, WI, USA, 2003: 305-312.
- 32 Li Shi-jin, Lin Lin, Li Xiao-fang. A novel approach to video transition detection based on a two-phase classification strategy [A]. In: Proceedings of 5th IEEE International Conference on Networking, Sensing and Control[C], Sanya, China, 2008: 1343-1347.
- 33 Geng Yu-liang, Xu De, Feng Song-he. A rapid and effective method for shot boundary detection [J]. Acta Electronica Sinica, 2006, **34**(12): 2272-2277. [耿玉亮, 须德, 冯松鹤. 一种快速有效的视频镜头边界检测方法 [J]. 电子学报, 2006, **34**(12): 2272-2277.]