

一种基于 SVM 主动学习的卡通视频检测方法

高新波 田春娜 张娜

(西安电子科技大学工程学院 西安 710071)

摘要: 通过分析卡通与非卡通视频在视觉上的差异,对视频片断提取了 MPEG-7 描述子等 8 组视觉特征来构造卡通视频的特征空间;并将主动相关反馈技术引入到支撑向量机(SVM)算法中,设计了一种基于主动学习的卡通视频检测分类方法。利用大量实际视频片断所做的测试实验结果表明,该文选取的特征对卡通和非卡通视频有较好的区分能力;且与单纯的 SVM 算法以及传统相关反馈和 SVM 算法结合的方法相比,该文算法在检测性能上有较大的优势。

关键词: 视频检测; MPEG-7; 支撑向量机; 相关反馈; 主动相关反馈

中图分类号: TN919.8

文献标识码: A

文章编号: 1009-5896(2007)06-1338-05

A Cartoon Video Detection Method Based on Active SVM Learning

Gao Xin-bo Tian Chun-na Zhang Na

(School of Electronic Engineering, Xidian University, Xi'an 710071, China)

Abstract: Through analyzing visual differences between cartoon and other videos, 8 groups of typical features including MPEG-7 descriptors are extracted to construct the feature space of cartoon videos. Then, a content-based video classifier is designed by introducing the active relevance feedback technique into Support Vector Machine (SVM) for the cartoon video detection. Experimental results on a great many real video clips illustrate that the constructed feature space can represent the cartoon videos effectively. In addition, compared with the classifier based on SVM and that based on the traditional relevance feedback technique and SVM, the proposed classifier has a higher performance for cartoon video detection.

Key words: Video detection; MPEG-7; SVM; Relevance feedback; Active relevance feedback

1 引言

随着网络多媒体及视频压缩技术的飞速发展,各种数字视频信息不断涌现。为便于从视频数据库中迅速获取感兴趣的视频类型,基于内容的自动视频分类技术引起人们的极大关注。除了对海量的数字视频进行自动分类整理以减少人工工作量之外,人们更希望视频播放器具有自动识别播放内容、视频过滤等多种智能功能。这些功能将在视频检索、点播(Video On Demand, VOD)、在线验证等应用中发挥重要作用。自动视频分类领域的研究内容广泛,关于视频类型的分类始于 1995 年,由 Fischer 首先提出,他将广播视频分为新闻、广告、卡通等类型^[1]。随后基于类型的视频分类研究逐步发展起来。很多研究者针对不同的应用领域对不同类型的视频进行分类研究。其中对于儿童的培养和教育来说,卡通片的自动检测和检索是一项很有意义的工作。

目前,基于内容的卡通视频检测的研究尚处于起步阶段。Ianeva 等虽然专门针对卡通视频提出了一种检测方法^[2],但所使用的视频饱和度等简单特征不足以区分目前种类日

益繁多的卡通视频;Roach 等基于运动特征对卡通视频进行分析^[3],实现实时检测,但是由于卡通视频的帧间运动与判作方法有很大关系,所以该方法仅局限于传统卡通视频的判别,对于目前制作方法越来越丰富的卡通片很难获得广泛的应用。近年来提出的示例学习方法支撑向量机(SVM)在统计模式识别领域获得了成功的应用。因此,本文将采用 SVM 来设计卡通视频分类器。在 SVM 学习系统中,有两个重要概念:示例空间和特征空间。示例空间是向系统提供的训练例集合。特征空间是示例空间中潜在的某种事物规律的集合。学习系统的任务就是从大量的训练例中自行总结出这些规律。为克服上述卡通检测和分类方法在特征空间描述不足的问题,本文对现有的卡通类型在视觉特征方面做了全面分析和分类,选取了 MPEG-7 视觉描述子等 8 组特征构成高维卡通视频特征空间。并引入独立成分分析(Independent Component Analysis, ICA)对产生的特征降维^[4]。理论上说,训练例越多,提取的规则越准确。然而在实际应用中,由于计算量对硬件的要求及多样性示例获取的代价等因素使得训练数据集的规模受到极大的限制^[5]。这样学习的性能很大程度上依赖于训练示例对概念的代表能力。通常人们通过引入主动相关反馈来让算法选择更典型的训练示例^[6],从而在示例有限的情况下提高分类器的性能。这就是基于 SVM 主动学习的卡通视

2005-09-19 收到, 2006-11-22 改回

新世纪优秀人才支持计划(NCET-04-0948),教育部重点项目(104173),国家自然科学基金(60202004)和教育部留学归国人员实验室资助课题

频检测方法的基本思想。

下文结构安排如下:第2节包括卡通视频的分类,本文选取的8组视觉特征构成的卡通视频特征空间及特征降维;第3节介绍了SVM主动学习算法的基本原理及分类器设计;第4节是实验设计、结果及分析;第5节是结束语。

2 特征空间描述

目前卡通视频的种类繁多,制作方法也千变万化,为给出有效的特征空间,本文先分析常见的卡通视频的视觉特点,并据此将现有卡通视频进行类别划分,并提出能将这些子类与非卡通有效区分的视觉特征。

2.1 卡通视频及其分类

简单地说,卡通视频就是用各种方法制作出来的非自然视频,从画面构成上看一般具有以下视觉特点中的几个:色彩饱和度强;颜色呈大块分布;有明显的黑色边缘;纹理简单,很少有细小的纹理;黑、白和灰色在视频中比重较小,颜色有多样性。

卡通视频的多样性导致所有的卡通不能同时具有以上全部特性,因此无法简单地将卡通片归为一类而非卡通区分开来。据此,将卡通片分为3个子类:一类是传统卡通,包括中国传统卡通、日本传统漫画等,这类卡通片颜色饱和度和强,人物的形态及背景简单;第2类是迪斯尼电影卡通,此类卡通片颜色的饱和度介于传统卡通和自然视频之间,线条比传统卡通丰富,但是人物的形态比较简单;第3类是近年兴起的日本宫崎骏导演的电影卡通片,它们的特点是颜色的饱和度低于一般自然视频颜色的饱和度,但是图像中色彩的对比度很高,人物及背景的制作也较精致。本文通过把现有卡通分为上述3种类型进行分类器设计,实现卡通视频的检测。下面针对这些卡通的特点寻求有效的视觉特征。

2.2 卡通视频的特征选择

通过对卡通与非卡通视频以及卡通视频子类间的特征分析,本文选择MPEG-7中的视觉描述子和图像质量评估测度等,得到8组视觉特征来改进卡通视频的特征描述,使其能够更加准确地区分不同视频的类型。

(1) HSV颜色直方图 HSV颜色直方图描述的是图像的整体颜色情况。且比较符合人类的视觉感受。为了更好地模拟人类对图像色彩的感受,很多研究者对HSV直方图进行了不同方式的量化。本文采用文献[7]中的非均匀量化方法得到一个36bin的颜色直方图来描述卡通视频的颜色分布。

(2) 彩色区域占整个图像的比重 因为卡通片中,彩色使用较多,而灰、白和黑3种颜色所占的比重较小,所以彩色像素点的个数占一帧图像中总像素点个数的百分比可作为一个特征区分卡通和非卡通图像。

(3) 颜色结构特征直方图 颜色结构特征是MPEG-7中的一个视觉特征描述子^[8]。它不仅能描述图像的整体颜色特征,还可以体现图像的局部颜色结构特征。颜色结构描述子

是在HMMD (Hue-Min-Max-Difference)空间上进行操作的。HMMD中H指色调(Hue);两个M分别指像素点R, G, B 3个值中的最大值(Max)和最小值(Min),一般应用中采用两者之和S(Sum)来表示;D代表上述最大值与最小值的差。针对卡通视频的特点,本文对HMMD空间进行如表1的非均匀量化,即首先将D空间非均匀量化为5个子空间,然后在每个子空间分别对H和S进行量化,最终形成一个128bin的直方图。为描述图像的局部颜色结构,在空间量化后,用8×8的栅格对整幅图像进行逐像素扫描。当栅格移动到图像的某一个位置时,栅格中包含的颜色在相对应的bin上频度加1。

表1 卡通视频的HMMD空间量化方法

D(Difference)	0-6	6-20	20-60	60-100	100-225
H(Hue)	1	4	8	8	8
S(Sum)	16	4	4	4	4

(4) 图像边缘方向直方图 图像边缘方向直方图表达了图像边缘方向的空间分布^[8]。因为在卡通视频中,人物或景物的形态抽象,线条简单,边缘大多具有规则的方向,而自然视频中则因很多细小的纹理,造成边缘不规则。所以使用该特征能有效地区别卡通和非卡通视频。可参照文献[8]中的方法构成一个80bin的边缘方向直方图。

(5) 图像梯度直方图 相对于自然视频来说,卡通视频帧中有明显的黑色边缘,所以可选用图像梯度直方图作为卡通视频的测度。该特征是在RGB颜色空间用式(1)直接计算图像中像素点与其八邻域点的梯度值,并选取最大值为该点的梯度值。然后将这个区域以8为间隔均匀量化成96bin。

$$\text{gradient} = |r - r'| + |g - g'| + |b - b'| \quad (1)$$

(6) 边缘点密度 因为卡通视频中纹理较少且线条简单。所以本文提出用边缘像素点个数占图像总像素点个数的比例作为图像的测度。利用式(1)来计算相邻像素点的梯度,当梯度gradient值大于某个给定的阈值时,我们认为它是边缘点;反之则认为这个差值是由噪声造成的。

(7) 图像纹理能量特征 针对卡通视频中纹理能量较低的特点,本文通过在频域用图像纹理能量特征算子计算能量和能量方差来提供纹理的量化描述。采用文献[9]中的方法在归一化的频域中,将频域空间分成30个频道。在30个频道上分别计算图像的纹理能量和能量偏差,另外计算图像纹理信息的亮度值 f_{DC} 和整个图像像素点的亮度标准差 f_{SC} ,由此生成62bin的特征矢量。

(8) 图像的彩色程度 这个特征是模仿人对图像的整体色彩感觉来描述一幅图像整体上的彩色程度^[10]。通过大量的实验,我们总结出式(5)来计算图像的彩色程度。M的值越大说明图像的彩色程度越高。相对非卡通视频,卡通视频的M值较高。该特征是在RGB空间实现的,对于图像中的任一像素点p,定义: $p_{rg} = R - G$, $p_{yb} = 0.5(R + G) - B$ 。对于一幅图像I,定义如下参量:

$$\mu_{rg} = \frac{1}{|I|} \sum_{p \in I} p_{rg}, \quad \mu_{yb} = \frac{1}{|I|} \sum_{p \in I} p_{yb} \quad (2)$$

$$\sigma_{rg}^2 = \frac{1}{|I|} \sum_{p \in I} (p_{rg}^2 - \mu_{rg}^2), \quad \sigma_{yb}^2 = \frac{1}{|I|} \sum_{p \in I} (p_{yb}^2 - \mu_{yb}^2) \quad (3)$$

$$\sigma_{rygb} = \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2}, \quad \mu_{rygb} = \sqrt{\mu_{rg}^2 + \mu_{yb}^2} \quad (4)$$

图像的色彩指标可用式(5)表示:

$$M = \sigma_{rygb} + 0.3\mu_{rygb} \quad (5)$$

计算视频中某一帧的上述特征,产生一个405维的特征矢量。但是这仅仅能够反映视频中某帧图像的特点。而视频片断是由连续的若干帧产生。要体现整个视频的特征,最简单的方法是将每个特征各帧取均值。但是这易受噪声的影响。为此,本文对各帧图像产生的同一特征的中值表示该视频的这一特征。对每一维特征做上述运算,最终得到整个视频的特征直方图。然而高维特征空间在训练时会产生“维数灾”和特征冗余。为此,本文使用ICA不仅能够实现维数约减,而且可使特征保持高阶相互独立,得到更有效的低维特征。本文采用ICA的快速定点算法^[4](Fast ICA),根据实验经验将高维特征降到20维。

3 SVM 主动学习算法原理及分类器设计

SVM算法是 90 年代中期在统计学习理论上发展起来的一种适合小示例分类问题的机器学习方法,并被广泛地应用到模式识别与人工智能等领域^[1]。针对本文中训练示例有限的情况及SVM的优点,我们选用SVM作为基学习器,且结合主动相关反馈技术实现训练过程中的示例优选。并针对卡通视频这个多类分类问题,设计二叉树型的分类器。

3.1 SVM 主动学习算法原理

SVM是在原空间或投影后的高维空间中寻找最优分类面将两类示例分开的一种学习算法。所谓最优分类面就是要求分类面不但能将两类示例无错误的分开,而且要使两类的分类空隙最大^[11]。传统的SVM被动地接受给定的训练示例,并从中归纳出一般性的知识。通常这些示例被假定是由相关领域的专家给出,事实上这样的专家未必存在或者专家标记示例要付出很大代价;且过多的训练示例对计算条件提出更高的要求。近年来研究者试图采用相关反馈算法来解决这个问题^[6],它的实现是使用设计好的分类器对扩展示例集进行首次检测,随机选取检测结果中的一部分供用户标记,然后将用户标记为分类错误或不相关的检测结果返回给训练集进行训练。一方面它能修改查询向量,使其向相关示例的中心移动;另外一方面利用反馈信息调整距离公式中各分量的权值,改善分类器的性能。经过若干次的反馈以得到用户满意的结果。但相关反馈算法给出的示例是主观意义上的最优数据,从算法的角度却未必如此。而且这种方式忽略了学习者对新示例学习的积极性。结合了主动学习的相关反馈不再随机选取检索或分类结果中的一部分数据^[6],而是选取结果中类别不确定性较大的示例给用户进行标记反馈。这样就

会使系统得到较大的信息增益。从而可以通过较少次数的反馈达到好的检索效果。

对于SVM来说,不确定性较大的数据大多分布在两个支撑向量面之间,因为这些数据处于分类面的边缘,分类面较小的移动,就会使它们的分类结果发生改变。关键问题是如何度量这种不确定性。本文使用测试示例 I_i 相对于两类训练示例的距离和 D_i 作为度量标准:

$$D_i = D_i^c + D_i^{nc} = \frac{1}{N_c} \sum_{j=1}^{N_c} \|I_i - I_j\|_2 + \frac{1}{N_{nc}} \sum_{j=1}^{N_{nc}} \|I_i - I_j\|_2 \quad (6)$$

其中 D^c 为测试示例与卡通训练示例的平均距离, D^{nc} 为测试示例与非卡通训练示例的平均距离, N_c 表示卡通示例的个数, N_{nc} 表示非卡通示例的个数, I 表示示例的特征向量。式(6)中 D 值越小,示例的分类不确定性越大,假定在这些被反馈的测试示例中不确定度的最大值为 D^M 。在使用生成的分类器对测试示例进行检测的同时计算每个测试示例的不确定度。将 D 值较小的前若干个测试示例反馈给用户进行标记,将分类错误的测试示例返回给系统。在SVM中,训练示例中只有支撑向量才对分类器的性能起作用。根据这个性质,我们同时计算非支撑向量的分类标签和不确定度,不确定度小于或等于 D^M 且分类错误的非支撑向量仍然留在训练示例集中,而其它的非支撑向量由用户标记过的错误测试示例替换掉。得到新的训练示例集后,重新进行训练,并且进行新一轮的分类、反馈,直至达到满意的分类性能。

3.2 卡通视频分类器设计

SVM是为解决二分类问题而提出的,然而卡通视频存在类间可分离性,如果修改SVM二次规划的形式直接进行多分类,会极大地增大二次规划的维数,从而增加问题的复杂度,且分类效果远不如二分类。另一种方法将多类分类分解为多个二分类问题,分别求解,而后再按一定的方式重构起来成为多类分类的解。后者常用的方法有一对多(One Against All, OAA)、一对一(One Against One, OAO)和二叉树SVM(Binary Tree SVM, BT-SVM)等。本文中数据集是由卡通视频和非卡通视频两个大类组成:卡通视频类又由几个相关的子类组成,这刚好符合BT-SVM的分级分解形式的要求。据此,本文设计了如图1所示的分类流程,通过使用3个层叠的二分类分类器实现卡通视频的检测和分类。

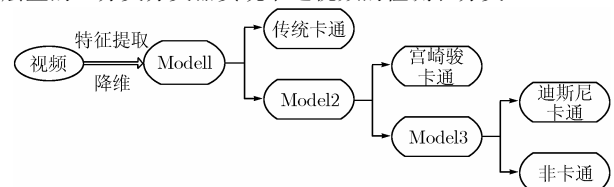


图 1 卡通视频的检测和分类流程图

使用上述流程得到了一个初始的分类树,因为SVM是使用小示例训练大集合,当随机选取示例的时候,通过少量的示例不一定能够得到一个适用的分类面。所以再采用上节所述的主动相关反馈方法对产生的初始分类面做进一步的修

正。当一个待判决的示例向量进入这个分类树,逐级地用各个判决函数进行判定,得到最终的分类结果。

4 实验结果及分析

为了验证本文提出的卡通视频检测方法的可行性和有效性,特设计了如下的测试实验,测试视频材料来自随机选取的日常电视节目。其中卡通视频来自于 81 部卡通片,从中随机截取若干 1min 的片断;非卡通视频来自于新闻、访谈、电影、电视剧、体育和广告等节目,同样截取若干 1min 的片断。按照第 3 节中设计的分类器,在训练阶段,生成 Model1 需要传统卡通片以及非卡通片示例各 40 个作为训练示例;同理,生成 Model2 和 Model3 也分别需要相应分类类别的卡通片和非卡通片示例各 40 个作为训练示例。在测试阶段,共选用卡通测试视频 237 个,非卡通测试视频 100 个。实验中,选择径向基函数(Radial Basis Function, RBF)为 SVM 的核类型;参数 $\gamma=0.25$, $C=1$, $\text{cache_size}=40$ 。

针对相同的训练和测试数据库,按照图 1 的分类流程,分别使用以下 3 种算法做了卡通视频检测实验:(E1) 不加反馈只用 SVM 算法;(E2) SVM 与传统相关反馈相结合的方法;(E3) SVM 与主动相关反馈相结合的方法。E1 中 SVM 的查全率为 84.0%,查准率为 86.9%。E2、E3 的实验结果见表 2。其中,查全率和查准率的定义可参见文献[12]。

表 2 3 种算法的卡通视频检测性能比较(%)

	SVM+相关反馈 (E2)		SVM+主动 相关反馈(E3)	
	查全率	查准率	查全率	查准率
第 1 轮反馈	84.0	88.5	84.2	90.1
第 2 轮反馈	87.5	87.2	84.2	93.0
第 3 轮反馈	84.9	91.4	87.8	93.3

实验结果对比分析如下:

(1)从 E1 结果可知,本文选取的视频特征对多种视频材料有较好的区分能力;

(2)为提高 SVM 分类器的检测性能,加入了传统相关反馈和主动相关反馈算法后得到的查全率和查准率都远高于单纯使用 SVM 算法得到的结果;

(3)主动相关反馈经过 3 次反馈后查全率和查准率都处于平稳上升的状态,而传统相关反馈在第 2 轮反馈时出现了波动,其查全率的升高是以查准率的降低为代价的。总的来说,加入主动相关反馈后检测性能要高于加入传统相关反馈后分类器的性能。

实验结果与第 3 节中的理论分析完全吻合。因为 SVM 分类器是小示例分类器,随机选取的示例不一定能很好地确定一个分类面来区分大量的测试示例;并且少量的几个示例就有可能改变整个分类面的位置。当反馈的示例不能给系统提供很多有价值的信息时,新的分类面与原来的分类面差别

不大。如果反馈的示例不当,就会造成分类面很大程度的偏向一方,第 2 轮反馈结果证明了这一点。所以最终主动相关反馈的查全率和查准率都高于传统相关反馈的结果。虽然引入反馈增加了分类器的训练时间,但是所选的支撑向量的个数变化不大,所以相对于检测性能的提高,引入主动相关反馈对系统检测时间的影响较小。

5 结束语

本文通过分析卡通视频与非卡通视频在视觉上的差异,提出了 8 组视觉特征来构造区分两类视频的特征空间,并基于 SVM 和主动相关反馈相结合的方法设计了一种二叉树型的卡通视频分类器。利用实际视频片断所做的测试实验结果表明本文所提出算法具有检测率高、性能稳定等优点。当然本算法还有待于进一步的改进和增强,比如,由于提出的特征空间较为复杂,针对卡通视频这种大数据量的媒体,算法尚不能实现实时检测,这也将是下一步的研究重点。此外,今后还需要对其它类型的视频检测和分类问题进行深入的研究,如:体育视频的分类,广告的自动检测过滤等,最终形成一套完整的视频分析系统。

参考文献

- [1] Fischer S, Lienhart R, and Effelsberg W. Automatic recognition of film genres. The 3rd ACM International Multimedia Conference and Exhibition. San Francisco, California, USA, Nov. 5-9, 1995, 1: 295-304.
- [2] Ianeva T I, de Vries A P, and Rohrig H. Detecting cartoons: A case study in automatic video-genre classification. Proc. IEEE International Conference on Multimedia and Expo. Baltimore, Maryland, Jul. 6-9, 2003, 1: 449-452.
- [3] Roach M, Mason J S, and Pawlewski M. Motion-based classification of cartoons. International Conference on Intelligent Multimedia, Video and Speech Processing, Hong Kong, May 2-4, 2001: 146-149.
- [4] Hyvarinen A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 1999, 10(3): 626-634.
- [5] 田春娜, 高新波, 李洁. 基于嵌入式 Bootstrap 的主动学习示例选择方法. *计算机研究与发展*, 2006, 43(10): 1706-1712.
Tian C N, Gao X B, and Li J. An example selection method for active learning based on embedded Bootstrap algorithm. *Journal of Computer Research and Development*, 2006, 43(10): 1706-1712.
- [6] 陈可佳, 姜远, 周志华. 基于主动相关反馈的图像检索方法. *模式识别与人工智能*, 2005, 18(4): 480-485.
Chen K J, Jiang Y, and Zhou Z H. An image retrieval method based on active relevance feedback. *Pattern Recognition and Artificial Intelligence*, 2005, 18(4): 480-485.
- [7] Zhang L, Lin F Z, and Zhang B. A CBIR method based on

- color-spatial feature. Proceedings of the IEEE Region 10 Conference on TENCN. Korea. Sept. 15–17, 1999, 1: 166–169.
- [8] Manjunath B S, Ohm J R, and Vasudevan V V, *et al.* Color and texture descriptors. *IEEE Trans. on Circuits and Systems for Video Technology*, 2001, 11(6): 703–715.
- [9] Ro Y M, Kim M, and Kang K, *et al.* MPEG-7 homogeneous texture descriptor. *ETRI Journal*, 2001, 23(2): 41–51.
- [10] Hasler D and Susstrunk S. Measuring colourfulness in natural images. Proc. SPIE/IS&T Human Vision and Electronic Imaging. Santa Clara, California, USA, Jan. 20–24, 2003: 87–95.
- [11] Burges C J C. A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining*, 1998, 2(2): 121–167.
- [12] Gao X B and Tang X. Unsupervised video shot segmentation and model-free anchorperson detection for news video story parsing. *IEEE Trans. on Circuits Systems for Video Technology*, 2002, 12(9): 765–776.
- 高新波: 男, 1972年生, 博士, 教授, 博士生导师, 研究方向为图像/视频信息处理、模式识别与智能系统.
- 田春娜: 女, 1980年生, 博士生, 研究方向为智能信息处理、图像及视频语义分割等领域的研究.
- 张娜: 女, 1980年生, 硕士, 研究方向为模式识别与智能信息处理.