

# 视频主对象特征抽取与分类挖掘研究<sup>\*</sup>

陈 芬<sup>1</sup> 苏新宁<sup>2</sup>

<sup>1</sup>(南京理工大学经济管理学院 南京 210094)

<sup>2</sup>(南京大学信息管理学院 南京 210093)

**【摘要】**尝试在区域分割的基础上,针对视频帧中用户最为关注的主要角色,提出基于主对象的颜色特征抽取新方法,并在视频分类实验中验证其效果。结果显示,基于主对象的特征抽取能够取得更好的挖掘效果,显示该方法的有效性。

**【关键词】**主对象特征 视频分类 视频挖掘

**【分类号】**G350

## Video Classification Based on Main Object Feature Extraction

Chen Fen<sup>1</sup> Su Xinning<sup>2</sup>

<sup>1</sup>(School of Economics & Management, Nanjing University of Science and Technology, Nanjing 210094, China)

<sup>2</sup>(School of Information Management, Nanjing University, Nanjing 210093, China)

**【Abstract】** This article focuses on the main object that users are most concerned in a video frame, proposes a visual feature extraction of main object based on the region segmentation, and validates the mining effect using the proposed feature. The result shows the better mining accuracy using the new feature, and indicates the validity of the method.

**【Keywords】** Main object feature Video classification Video mining

## 1 引 言

计算机和网络技术的飞速发展促进了视频资源的快速增长,视频信息的组织、检索与挖掘日益成为国内外研究者广泛关注的研究课题。

视频具有不同于其他媒体形式的显著特征,包括海量性、内容的丰富性、结构的复杂性等。视频数据的特点以及视频资源的快速增长,使得传统的手工处理方式已不可行,如何借助于计算机、采用自动化手段进行视频信息的组织、检索和挖掘,是当今计算机检索领域面临的一个重要问题。此外,视频信息是典型的大数据,该领域研究也迎合了当今大数据处理与挖掘的发展趋势。

分类是视频挖掘中最常见的研究内容,其目标是按类存放视频各级组成,便于管理、浏览和访问。分类是视频查询和浏览的有效手段,便利了资源方对视频的有效组织以及用户对视频信息的有效使用。

近年来,国外视频分类的典型研究包括 Bagheri - Khaligh 等<sup>[1]</sup>、Azhar 等<sup>[2]</sup>、Kafai 等<sup>[3]</sup>、Ekenel 等<sup>[4]</sup>、Connolly 等<sup>[5]</sup>、Wang 等<sup>[6]</sup>、Mithun 等<sup>[7]</sup>的研究。例如,Bagheri - Khaligh 等<sup>[1]</sup>针对足球视频进行分类,提出最大球衣颜色比例和球衣色彩比例均值等特征。Azhar等<sup>[2]</sup>提出了监控视频中不完整的、背景重叠等物体对象的二元分类框架,

收稿日期: 2013 - 05 - 27

收修改稿日期: 2013 - 08 - 27

<sup>\*</sup> 本文系教育部人文社会科学研究青年基金项目“基于云平台的视频信息组织和挖掘研究”(项目编号:10YJC870001)、中国博士后科学基金项目“基于多模式的视频信息分类挖掘研究”(项目编号:2012M521061)和中央高校基本科研业务费专项资金项目“基于多模式的视频信息结构和模式挖掘”(项目编号:30920130132010)的研究成果之一。

通过混沌序列仿真,将 MPEG-7 视觉描述子结合起来,类别涉及“包含人”、“包含一群人”、“包含车辆”、“包含未知物体”等。Kafai 等<sup>[3]</sup>提出了随机多类车辆分类系统,将车辆分成轿车、小型载货卡车、SUV、未知类型 4 类,从汽车尾灯和车辆尺寸角度抽取特征集,使用动态贝叶斯网络进行分类。Ekenel 等<sup>[4]</sup>抽取底层音频特征、视觉特征、感知和结构信息及网络视频标签对电视节目和 YouTube 视频进行分类,使用 SVM 分类算法;其中,视觉描述子包括颜色、纹理,音频描述子包括信号能量、过零率、基本频率,感知信息对应于通过人脸识别抽取的信息,结构信息与镜头编辑相关,标签描述子基于 TF-IDF 公式对 YouTube 视频进行分类。

国内的代表性研究包括胡颖俊等<sup>[8]</sup>、赵士伟等<sup>[9]</sup>、林彬等<sup>[10]</sup>、李真超等<sup>[11]</sup>的研究。例如,胡颖俊等<sup>[8]</sup>使用包括 RGB 通道平均值、颜色直方图、Gabor 纹理等在内的特征。赵士伟等<sup>[9]</sup>抽取了关键帧颜色数目、颜色直方图均值、颜色直方图方差、主颜色、主色调、最分散颜色、最分散色调、最紧凑颜色、最紧凑色调等特征,并利用决策树算法进行分类。林彬等<sup>[10]</sup>通过 logo 模板匹配检测并定位出视频中的慢镜头,对其余的正常比赛部分做镜头边界检测完成视频切分;并基于分块的思想,对正常比赛镜头帧计算其各块的场地像素比率值作为特征,利用 SVM 分类器将正常比赛镜头分为远镜头、中镜头、球员特写或场外镜头三类。

从以上文献可以看出,新特征提取算法的实现是视频分类研究的重点,分类方法上,几乎所有的主流分类挖掘算法都已经被应用到视频处理中。总的来说,目前视频分类挖掘使用的技术多样化、没有固定的模式,各种研究涉及不同的特征及抽取方式、分类方法,以及视频类别的不同选择。

本文主要进行视频新特征的抽取。与以往方法相比,本文更加关注视频的主要对象区域,这也是普通用户在浏览视频时最为关注的部分。本文针对主要对象所代表的局部空间,选择最能够代表视频内容的主要对象区域进行视频表征,并基于新特征进行视频分类挖掘实验。

## 2 研究框架与方法

本文重点进行颜色新特征的抽取。颜色特征是在视频检索中应用最为广泛的视觉特征,颜色往往和视

频帧中所包含的物体或场景十分相关。

到目前为止,研究者已经提出了多种颜色特征抽取方法,最著名的包括颜色直方图、颜色矩、颜色聚合向量和颜色相关图等。其中,颜色直方图<sup>[12]</sup>是视频挖掘使用最广泛的视觉特征,颜色矩<sup>[13]</sup>是对颜色直方图的一种改进,无需对特征进行向量化。然而,无论颜色直方图还是颜色矩,都无法表达视频帧的空间信息。针对它们的缺点,研究者提出了颜色聚合向量等<sup>[14,15]</sup>特征,包含了颜色分布的空间信息,能够获得比颜色直方图更好的检索效果。

与颜色聚合向量类似,本文也引入局部空间信息,在对象分割的基础上,基于主对象进行视觉特征抽取。主对象即视频帧中用户最关心的主要角色,因此笔者将其命名为“主对象”。与颜色聚合向量针对视频帧所有像素进行处理不同,本文方法针对主对象所在的局部空间,不涉及视频帧的所有像素。

在新特征抽取的基础上,进行视频分类挖掘实验,并将基于新特征的挖掘效果与颜色直方图和颜色聚合向量进行对比,分析本文特征抽取的有效性。

### 2.1 新特征抽取的主要步骤

#### (1) 视频预处理

预处理包含两个主要步骤:镜头边界检测和关键帧提取。其中,镜头边界检测基于颜色直方图进行,当连续两帧直方图超过预定的阈值时,认为镜头边界存在;关键帧提取则采用笔者提出的最大最小直方图识别方法,利用信息论的观点,即不同或相关性较小的帧图像比相似的帧图像携带更多的信息,具体算法将在另一篇文章中详细论述。

#### (2) 区域分割

采用分水岭的区域分割方法<sup>[16]</sup>来获取不同的对象。分水岭算法对微弱边缘具有良好的识别能力,能够得到封闭连续的边缘。另外,该算法得到的汇水盆地,为分析视频帧的区域特征提供了良好的基础。

分割的结果形成视频帧的不同逻辑组成。帧中每一个像素被赋予一个标识,表征图像区域号。区域分割结果如图 1 所示。

#### (3) 主对象新特征提取

如前所述,颜色直方图、颜色矩等方法无法表达视频帧的空间分布信息,因此,引入空间分布的颜色特征提取显得十分必要。



图 1 区域分割结果

本文基于区域分割,引入局部空间信息,进行主对象的识别与颜色特征提取,详细算法见 2.2 节。

#### (4) 挖掘实验

在新特征抽取的基础上,基于新特征进行视频分类挖掘实验,并与颜色直方图和颜色聚合向量进行对比,从而验证本文提取方法的有效性。

### 2.2 基于主对象的视觉新特征提取

#### (1) 算法思想

对象代表了视频最重要的语义信息。在对象分割的基础上,笔者对视频帧进行分析发现,视频帧中的大部分内容为背景信息,最重要的目标对象在视频帧中所占的比例可以认为大致在一半以内。因此,笔者提出了如下基本算法思想:

- ①根据区域标识码,计算不同对象区域面积;
- ②统计视频帧一半面积的区域,用于表征主对象;
- ③结合视频原帧与选定区域,计算主对象颜色新特征。

由于 HSV 颜色空间更接近人们对颜色的主观认识,因此本文采用该颜色空间。HSV 空间的三个分量分别代表色彩 (Hue)、饱和度 (Saturation)、亮度 (Value),其中,受光照条件影响最大的是 V 分量,因此在识别目标时,笔者减少了亮度 V 的权值,以 H 和 S 作为判定的主要特征。本文中,H、S、V 分别取量化权重为 8、3、3,从而形成了 72 维特征。

#### (2) 算法流程

主对象特征具体计算流程伪码如下:

```
for (i:视频帧像素的区域标识矩阵的列数) //计算区域面积
for(j:视频帧像素的区域标识矩阵的行数)
    取得当前处理像素(i,j)的区域号 Flag(i,j);
    针对区域变量 nRegion(包括 id,number),查找 id 为
        Flag(i,j) 的元素
    if(找到)
        Flag(i,j) 区域对应的像素数量累加;
    else
        新的区域,增加新区域的标记;
        对 nRegion 变量进行排序,根据 Flag(i,j) 值升序排列;

for (m:区域个数标识) //主对象区域计算
```

```
读取nRegion 中的当前数据,计算当前区域占整个区域的面积
比例 Rate(m);
将 Rate(m)累加到变量 s(s 用于计算当前区域的累加面积占
视频帧的比例);
if (s>0.50)
    确定主对象的具体区域并退出;
```

for (n:主对象所属区域的像素) //主对象新特征计算

根据公式(1)和公式(2)进行 H、S、V 的量化与特征值计算;

$$H = \begin{cases} 0 & \text{if } h \in [0, 44] \\ 1 & \text{if } h \in [45, 89] \\ 2 & \text{if } h \in [90, 134] \\ 3 & \text{if } h \in [135, 179] \\ 4 & \text{if } h \in [180, 224] \\ 5 & \text{if } h \in [225, 269] \\ 6 & \text{if } h \in [270, 314] \\ 7 & \text{if } h \in [315, 360] \end{cases} \quad S = \begin{cases} 0 & \text{if } s \in [0, 1/3] \\ 1 & \text{if } s \in [1/3, 2/3] \\ 2 & \text{if } s \in [2/3, 1] \end{cases}$$

$$V = \begin{cases} 0 & \text{if } v \in [0, 1/3] \\ 1 & \text{if } v \in [1/3, 2/3] \\ 2 & \text{if } v \in [2/3, 1] \end{cases} \quad (1)$$

$$G = HQ_s Q_v + SQ_v + V \quad G = 9H + 3S + V \quad (2)$$

其中,公式(1)如前所述,H、S、V 分别量化为 8、3、3;公式(2)中, $Q_s$  和  $Q_v$  分别是分量 s 和 v 的量化级数,这里, $Q_s = 3, Q_v = 3$ 。

除了上述的主对象颜色特征,本文还结合 8 维灰度共现纹理矩阵,用于特征的联合挖掘。这样,最终的维度共计 80 维。

### 3 挖掘实验及讨论

#### 3.1 数据集

实验数据集为自录视频,共包含 141 个视频片段,如表 1 所示:

表 1 实验数据集

类型	数量	备注
电视	38	各种电视剧、电影节目(古装、现代;国内、国外;言情、武打、刑侦等各种类别)
其他视频	103	包括运动视频(足球、篮球、自行车、摔跤等);纪录片;新闻(央视新闻与地方新闻);卡通、广告等

其中,手工选择 60 个视频片段作为训练集,其他作为测试集。本文以电视识别为目标,这方面的研究目前比较少见。

#### 3.2 评估方法

本文使用如下指标表征识别效果:

##### (1) 准确率

$$\text{Accuracy} = \frac{\sum_{i=1}^k C_i}{n} \quad (3)$$

其中,k 为类的数目, $C_i$  为分类结果中与 Gold

Standard(每一个视频所属的正确集合类别)重合的视频数目, $n$ 是测试集的视频数目。

## (2) Cohen's Kappa 系数

Cohen's Kappa 系数是一个可靠性的统计度量,公式如下<sup>[17]</sup>:

$$k = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)} \quad (4)$$

其中, $\text{Pr}(a)$ 为相对观测一致性系数, $\text{Pr}(e)$ 为一致性发生的概率。当 $k$ 值大于0.41时,可以认为结果比较可靠<sup>[17]</sup>。

## 3.3 分类器

多种分类算法已被应用于视频挖掘,包括决策树(Decision Tree)、贝叶斯(Naïve Bayes)、SVM等。鉴于SVM方法在多维特征挖掘方面的优良效果,本实验采用SVM算法。

## 3.4 主对象特征结果示例

不同类型视频的主对象颜色特征分布如图2所示:

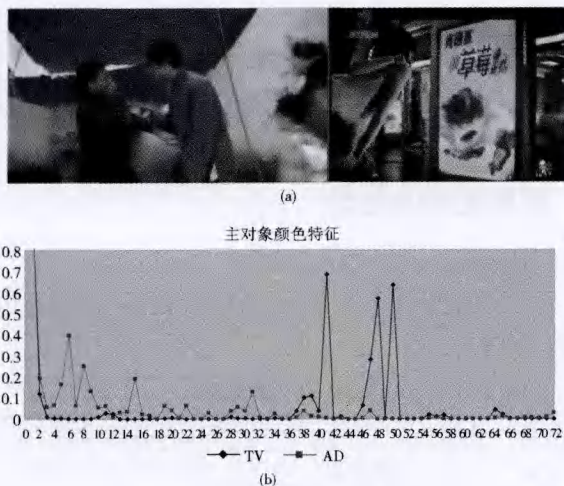


图2 不同类型视频的主对象颜色特征分布示例

其中,图2(a)为两个视频帧分别属于电视类及广告类,图2(b)展示了这两个不同帧的主对象颜色标准化分布。从图2中可以看出,它们在颜色分布上存在着较为明显的差别。

## 3.5 挖掘结果对比

以电视类为识别目标,针对表1数据进行挖掘后得到的结果如图3所示,同时为了验证本文提取方法的效果,将提取的基于主对象的视觉特征与著名的颜色直方图<sup>[12]</sup>和颜色聚合向量<sup>[14]</sup>进行对比。

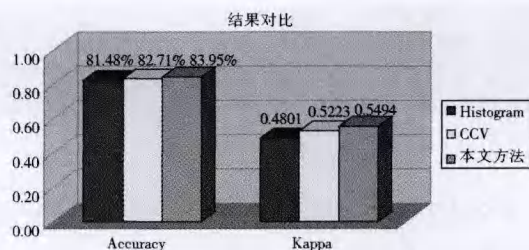


图3 挖掘结果对比

图3中,Histogram代表颜色直方图,CCV代表颜色聚合向量特征。可以看出,无论从准确率还是Kappa值,基于本文方法提取的主对象颜色特征都取得了较好的、同时比颜色直方图和颜色聚合向量更好的效果,从而说明本文提出的特征提取方法的有效性,也说明了本文方法对于空间信息以及视频帧的最重要的语义组成——主要对象的关注是合理的。

根据已有经验,笔者发现,多角度不同特征的联合挖掘能够取得更高的识别准确率。后续的工作中,笔者将融合更多特征进一步提高挖掘效果。

## 4 结 语

本文主要进行基于主对象的视觉特征提取研究,增加颜色特征提取中对于空间信息的考量,在对象分割的基础上重点考察视频中用户更为关注的主要对象,体现视频帧的局部和细节信息。

在特征抽取的基础上,针对141个视频片段进行分类实验。结果表明,基于本文的抽取方案能够获得较好的挖掘效果,证明本文抽取方法的有效性,同时也为视频信息的浏览、分类导航与检索提供了借鉴。

## 参考文献:

- [1] Bagheri - Khaligh A, Raziperchikolaei R, Moghaddam M E. A New Method for Shot Classification in Soccer Sports Video Based on SVM Classifier [C]. In: *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*. 2012; 109-112.
- [2] Azhar H, Amer A. Classification of Surveillance Video Objects Using Chaotic Series[J]. *IET Image Processing*, 2012, 6(7): 919-931.
- [3] Kafai M, Bhanu B. Dynamic Bayesian Networks for Vehicle Classification in Video[J]. *IEEE Transactions on Industrial Informatics*, 2012, 8(1): 100-109.



- [ 4 ] Ekenel H K, Semela T. Multimodal Genre Classification of TV Programs and YouTube Videos[ J]. *Multimedia Tools and Applications*, 2013, 63(2): 547 - 567.
- [ 5 ] Connolly J F, Granger E, Sabourin R. An Adaptive Classification System for Video - based Face Recognition[ J]. *Information Sciences*, 2012, 192: 50 - 70.
- [ 6 ] Wang X F, Zhang X P. An ICA Mixture Hidden Conditional Random Field Model for Video Event Classification[ J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2013, 23(1): 46 - 59.
- [ 7 ] Mithun N C, Rashid N U, Rahman S M M. Detection and Classification of Vehicles from Video Using Multiple Time - Spatial Images[ J]. *IEEE Transactions on Intelligent Transportation Systems*, 2012, 13(3): 1215 - 1225.
- [ 8 ] 胡颖俊, 沈航. 基于语义的视频镜头的分类技术[ J]. *计算机应用与软件*, 2010, 27(7): 230 - 232. (Hu Yingjun, Shen Hang. Semantic - based Classification Technique of Video Shots[ J]. *Computer Applications and Software*, 2010, 27(7): 230 - 232. )
- [ 9 ] 赵士伟, 卓力, 孙少卿, 等. 基于数据挖掘的视频镜头分类方法[ J]. *北京工业大学学报*, 2012, 38(5): 722 - 726. (Zhao Shiwei, Zhuo Li, Sun Shaoqing, et al. Data Mining - based Video Shot Classification Method[ J]. *Journal of Beijing University of Technology*, 2012, 38(5): 722 - 726. )
- [ 10 ] 林彬, 刘群, 王群, 等. 足球视频镜头分类方法[ J]. *计算机工程与设计*, 2012, 33(4): 1468 - 1471. (Lin Bin, Liu Qun, Wang Qun, et al. Method of Shot Classification for Soccer Video[ J]. *Computer Engineering and Design*, 2012, 33(4): 1468 - 1471. )
- [ 11 ] 李真超, 纪传俊, 林哲, 等. 基于多特征距离学习的视频分类[ J]. *计算机应用与软件*, 2012, 29(12): 10 - 12, 26. (Li Zhenchao, Ji Chuanjun, Lin Zhe, et al. Video Classification with Multiple Feature Distance Learning[ J]. *Computer Applications and Software*, 2012, 29(12): 10 - 12, 26. )
- [ 12 ] Swain M J, Ballard D H. Color Indexing[ J]. *International Journal of Computer Vision*, 1991, 7(1): 11 - 32.
- [ 13 ] Stricker M A, Orengo M. Similarity of Color Images[ C]. In: *Proceedings of SPIE's Symposium on Electronic Imaging: Science & Technology. Storage and Retrieval for Image and Video Databases*, 1995, 2420: 381 - 392.
- [ 14 ] Pass G, Zabih R, Miller J. Comparing Images Using Color Coherence Vectors[ C]. In: *Proceedings of the 4th ACM International Conference on Multimedia*. New York, NY, USA: ACM, 1996: 65 - 73.
- [ 15 ] 黄诚, 王国营. 一种基于颜色聚合向量的图像检索方法[ J]. *计算机工程*, 2006, 32(2): 194 - 196, 199. (Huang Cheng, Wang Guoying. A Method of Image Retrieval Based on Color Coherence Vector[ J]. *Computer Engineering*, 2006, 32(2): 194 - 196, 199. )
- [ 16 ] Gonzalez R C, Woods R E, Eddins S L. *Digital Image Processing* [ M]. Prentice Hall, 2005: 501 - 502.
- [ 17 ] Cohen's Kappa Coefficient[ OL]. [ 2013 - 03 - 29 ]. [http://en.wikipedia.org/wiki/Cohen%27s\\_kappa](http://en.wikipedia.org/wiki/Cohen%27s_kappa).  
(作者 E-mail: pku.lanyan@gmail.com)