基于视觉注意特征和SVM的镜头边界检测算法

陈 萍[1],李秀强[2],肖国强[2],江健民[2]
CHEN Ping[1],LI Xiu-qiang[2],XIAO Guo-qiang[2],JIANG Jian-min[2]

1.黄淮学院 计算机科学系,河南 驻马店 463000
2.西南大学 计算机与信息科学学院,重庆 400715
1.Department of Computer Science,Huanghuai University,Zhumadian,Henan 463000,China
2.College of Computer and Information Science,Southwest University,Chongqing 400715,China
E-mail:qiangxiuli@163.com

CHEN Ping,LI Xiu-qiang,XIAO Guo-qiang,et al.Shot boundary detection using SVM on visual attention features.Computer Engineering and Applications,2010,46(7):184-186.

Abstract: Shot boundary detection is the basis of video analysis.This paper proposes an algorithm of shot boundary detection, which employs Support Vector Machine (SVM) to classify visual attention features based on the research results of psychology. Extensive experiments carried out on TRECVID2007 database show that the proposed approach works well in detecting shot boundary measured by both recall and precision.

Key words: shot boundary detection;visual attention feature;Support Vector Machine(SVM)

摘 要:镜头边界检测是视频分析的基础。借鉴心理学中有关视觉注意的研究成果,提出了一种采用符合人类视觉注意的特征,并利用支持矢量机进行视频镜头边界检测的算法。通过对TRECVID2007数据库进行实验的结果表明,该算法在查全率和查准率方面都获得了较好的性能。

关键词:镜头边界检测;视觉注意特征;支持矢量机
DOI:10.3778/j.issn.1002-8331.2010.07.056 **文章编号:**1002-8331(2010)07-0184-03 **文献标识码:**A **中图分类号:**TP391.41

1 引言

镜头边界检测是进行视频内容分析的首要步骤,是视频语义、内容分析的基础。镜头是一组连续的相互关联的帧,是摄像机的一次连续拍摄,代表时间或空间上连续的一组动作。镜头边界的形成是两个镜头进行切换的结果,对于观察者来说,是视频镜头的内容发生了某种意义上的变化,即边界是由于视频内容的不连续造成的。镜头边界的类型一般可以分为突变和渐变两种。从检测的过程来看,整个过程可以归结为3个阶段:(1)特征的提取;(2)帧间差值的构造;(3)镜头变换类型的检测和分类。

在镜头边界的检测中,特征的提取,主要集中在压缩域和像素域中进行。像素域中的方法主要有:基于像素比较方法;基于像素块的比较方法;全局或局部直方图的比较;基于模型分割等。在压缩域中主要有:基于DCT系数的分割;基于DC分量的分割;基于DCT系数、宏块类型和运动矢量的分割等[1]。边界分类的方法,在早期主要采用固定阈值的方法,但固定阈值检测对于阈值的设定过于敏感,除非采用好的特征计算出来的不连续值对于边界和镜头内的帧间差别很大,一般采取自适应阈值。近年来,有许多学者和机构用机器学习的方法检测镜头边界。如AT&T采用了像素域中特征,并结合有限状态机和支持矢量机进行镜头边界检测的算法[2];Lee等人提出利用压缩域中特征,并结合神经网络进行镜头边界检测的方法[3],用机器学习的方法进行镜头边界检测是目前的一种发展趋势。

文章提出了一种基于视觉注意特征和支持矢量机(SVM)的镜头边界检测算法,该算法通过提取符合人类视觉注意的特征作为特征数据,同时采用支持矢量机进行镜头边界检测。该算法的实验结果与TREVID 2007公布的评测结果[4]进行对比后,说明该算法具有较好的性能。首先概要介绍了视觉注意特征的提取,然后分别介绍了突变和渐变的检测,最后给出了TREVID2007数据集上的测试结果。

2 算法描述

2.1 特征的提取

基于视觉注意的显著区域检测对于图像分析过程有着非常重要的意义。注意是人类信息加工过程中的一项重要的心理调节机制,它能够对有限的信息加工资源进行分配,使感知具备选择能力。如果能够将这种机制引入图像分析领域,将计算资源优先分配给那些容易引起观察者注意的特征,那么必将极

基金项目:重庆市自然科学基金(the Natural Science Foundation of Chongqing City of China under Grant No.CSTC-2008BB2252)。
作者简介:李秀强(1983-),男,硕士研究生,主要研究方向:数字媒体处理;肖国强(1965-),男,博士后,教授,主要研究方向:信号与信息处理、无线通信;江健民(1956-),男,博士,教授,主要研究方向:图形图像处理。
收稿日期:2008-09-08 **修回日期:**2009-06-29

大地提高现有图像分析方法的工作效率。将视觉注意特征应用到镜头边界检测中，也正是在这种思想的基础上提出的。

根据显著性度量方法的不同，目前的显著区域检测算法分为两大类：基于局部特征的算法和基于视觉反差的算法。后者主要根据视觉感知过程提出的，通用性强，是当前的主要研究方向，其中又以Itti的算法最具代表性[5-7]。算法以一幅图像的显著性度量结果合成为一副显著图，文章就在该算法思想的基础上提取得到视觉注意特征的。

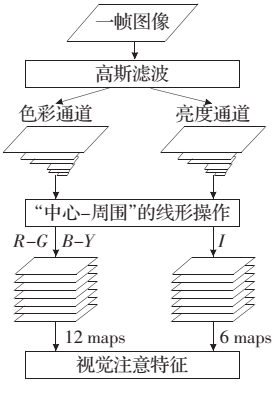

图1　视觉特征提取流程图

在MPEG-2解码平台的基础得到一段视频的帧序列$(1,2,\cdots,n)$。然后再对每一帧图像进行特征提取的操作，具体的视觉特征提取流程如1图所示。首先采用一个5×5高斯滤波操作把每一帧图像变换成一个7层的金字塔模型，每帧图像都按照下面的公式进行模型的构造：

$$g_{l+1}(i,j)=\sum_{m=-2}^{2}\sum_{n=-2}^{2}w(m,n)g_l(2i+m,2j+n) \tag{1}$$

其中，$g_l(i,j)$代表金字塔中第L层中像素点的位置，$w(j)$为高斯滤波的核函数[8]。经过计算得到一个高斯金字塔模型，金字塔的最底层为原始图像，上面诸层的水平和垂直像素点数目分别与原始层的像素点数目比例为：1:1(原始层0)到1:64(最高层6)。

人类的视觉神经细胞总是对一个场景的一些显著地方很敏感，因此特征提取来自于"中心—周围"的线形操作。中心点是指位于金字塔中第$c\in\{0,1,2\}$层中的像素点，周围点是指位于金字塔中第$s=c+d$，$d\in\{3,4\}$层上对应的像素点。

按照Ewald Hering's颜色对立理论[6]，在提取到每个像素点的三基色r、g、b值后，做以下变换：

$$R=r-\frac{g+b}{2} \tag{2}$$

$$G=g-\frac{r+b}{2} \tag{3}$$

$$B=b-\frac{r+g}{2} \tag{4}$$

$$Y=r+g-2(|r-g|+b) \tag{5}$$

$$I=\frac{r+g+b}{3} \tag{6}$$

经过变换得到红(R)、绿(G)、蓝(B)、黄(Y)4种颜色。

在此基础上，再计算"中心—周围"的差值，也就是中心层和周围层对应像素点的差值。

$$RG(c,s)=|(R(c)-G(c))\Theta(R(s)-G(s))| \tag{7}$$

$$BY(c,s)=|(B(c)-Y(c))\Theta(B(s)-Y(s))| \tag{8}$$

$$I(c,s)=|I(c)\Theta I(s)| \tag{9}$$

公式中c和s表示在上文中已经说明，这样就得到6个对应层的差值，对应层的差值是一个很庞大的数据，在求得差值后，可以把每个对应层作为一幅图片对待。为了达到降维的目的，对每张对应图片进行了区域划分，将其分成为2×2的块，这样从每帧图像提取一个72维的特征矢量$\boldsymbol{T}=\{t_1,t_2,\cdots,t_{72}\}$。

2.2　镜头边界检测

镜头是视像序列的基本元素，如前所述，镜头的边界可分为两类：突变和渐变。基于上面的分析，提出了一种检测镜头边界的方法，该方法分别对突变和渐变进行检测，利用上面分析得到的特征数据，通过支持矢量机学习镜头边界在视频帧序列中的特征变化规律。

支持向量机(Support Vector Machine，SVM)是由Vapnik的等人基于统计学习理论，采用结构风险最小化原理提出的一种新的机器学习算法。通过调整判别函数使得它最好地利用边界样本点的分类信息，构造最佳分类超平面。其主要优点有：(1)是专门针对有限样本情况的，其目标是得到现有信息下的最优解而不仅仅是样本趋于无穷大时的最优值；(2)算法最终将转化成为一个二次型寻优问题，从理论上说，得到的将是全局最优点，解决了在神经网络方法中无法避免的局部极值问题；(3)算法将通过非线形变换将原数据空间转换到高维的特征空间，在高维空间中构造线形判别函数来实现原空间中的非线形判别函数，这种特殊性质能保证机器有较好的推广能力，同时它巧妙地解决了维数问题。

若给定的样本集为(x_i,y_i)，$i=1,2,\cdots,n$，$x_i\in R^d$，$y_i\in\{-1,+1\}$是类别标号，则支持向量机的判别函数为：

$$f(x)=\mathrm{sgn}\Big(\sum_{i=1}^{sv}\alpha_i^* y_i K(x_i,x)+b^*\Big) \tag{10}$$

式中sv是支持向量的个数，$K(x_i,x)$为核函数，支持向量机核函数选用RBF函数：

$$K(x_i,x_j)=\exp(-\gamma\|x_i-x_j\|^2),\gamma>0 \tag{11}$$

则由RBF核函数构建的支持向量机分类器，只含有参数C和γ。对于最优参数的选择采用台湾大学林智仁在文献[9]中介绍的方法，利用grid.py工具[9]自动进行数据的缩放和最优参数的选择。

最后基于学习得到的模板进行镜头边界的分类。在确定了用支持矢量机进行镜头边界检测后，进一步的工作就是选取合适的特征矢量描述不同边界的特征。下面分别对其进行介绍。

2.2.1　突变边界的检测

突变是镜头间的突然变化，在两帧图像间完成，因此对它检测只需要考虑选择合适的特征矢量，在得到一段视频的帧序列$(1,2,\cdots,n)$的连续特征矢量后，位于突变边界的两帧图像的特征值也会有明显的变化。采用上面提取的视觉注意特征，对相邻两帧图像的特征值进行差分，突变点处的差值矢量作为正样本进行训练，而非突变镜头间的差值矢量作为负样本。

2.2.2　渐变边界的检测

现行算法的突变的检测效果比较好，而渐变边界的检测效果并不理想[10]。相对于突变边界，渐变边界具有更加复杂的特点，首先渐变的长度从几帧到上百帧不等，这就要求检测算法能够考察整个变化的过程；其次渐变的变化比较平缓，相邻的帧间差比较小，甚至和正常帧间差没有明显差异，这就要求检测算法能容忍渐变过程中少量平缓变化的帧[4]。常用的方法是

跨 n 帧作差。在一个镜头内部,镜头场景基本没有变化,跨 n 帧间的差会比较小,而在镜头变换处会明显地增大。

鉴于上面的分析,每帧图像都提取到一个 72 维的特征矢量 $\boldsymbol{T}=\{t_1,t_2,\cdots,t_{72}\}$,因为渐变的特征值变化比较平缓,在这里采取跨 n 帧间的差值作为特征矢量。公式如下:

$$F(t)=\sum_{i=1}^{72}(T_i(t)-T_i(t-n)) \tag{12}$$

其中,$T_i(t)$代表视频帧序列中第 t 帧在第 i 维的视觉注意特征数据。

通过对实验结果的分析,视觉注意的特征对于物体的快速移动比较敏感,因此,在这里也引入了绝对亮度帧差 I 的特征来辅助检测渐变边界。由于亮度帧间差对于运动不敏感,它对运动和噪音具有鲁棒性[1],在提取视觉注意特征的同时,也提取亮度的绝对帧间差 I:

$$I(t)=\sum^{X}\sum^{Y}abs(f(x,y,t)-f(x,y,t-n)) \tag{13}$$

其中,X 和 Y 代表视频帧的宽和高;$f(x,y,t)$是视频帧序列中第 t 帧在(x,y)坐标处像素点的亮度值;$f(x,y,t-n)$是视频序列中第 $t-n$ 帧在(x,y)坐标处像素点的亮度值。在实验中,选取 n=3 和 n=7 时的差值作为特征矢量。处于渐变位置的特征矢量作为正训练样本,不处于渐变位置的特征矢量作为负训练样本。

3 实验结果和分析

该文的算法在 TRECVID2007 提供的部分视频数据测试集上进行了测试。用 TRECVID2007 的 10 段视频数据作为训练样本,用其他的 7 个视频作为测试数据,测试的结果主要用查全率(Recall)和查准率(Precision)两个指标衡量检测性能,结果如表 1 所示。

表 1 镜头边界检测的结果

TRECVID 视频	突变检测结果		渐变检测结果	
	查全率(R)	查准率(P)	查全率(R)	查准率(P)
BG_37417.mpg	0.950	0.961	0.700	0.729
BG_35050.mpg	0.979	0.969	0.667	0.712
BG_35187.mpg	0.909	0.845	0.714	0.692
BG_37879.mpg	0.800	0.863	0	0
BG_37359.mpg	0.920	0.943	0.714	0.677
BG_37822.mpg	0.915	0.900	0.500	0.600
BG_12413.mpg	0.943	0.877	0.563	0.625

在 BG_37879.mpg 中没有渐变镜头转换。在对比 TRECVID2007 公布的前 15 名的结果后[11],给出了图 2 的镜头边界检测统计结果比较图。

从图 2 中可以看出,该算法在 TRECVID2007 镜头边界检测测试中取得了满意的检测性能,但还是存在不足,如乱判、误判和漏判现象,主要的原因还是存在视频特效、渐变跨度太大等,具体情况不再详述。

4 结束语

文章提出了一种基于视觉注意原理的镜头边界检测新算法,利用符合人类视觉信息处理机制的视觉注意特征,采用支持矢量机分类学习方法对镜头边界特征变化规律进行检测。在与 TRECVID2007 公布的结果的对比中,该算法在保证查全率和查准率的情况下,取得了令人满意的效果。

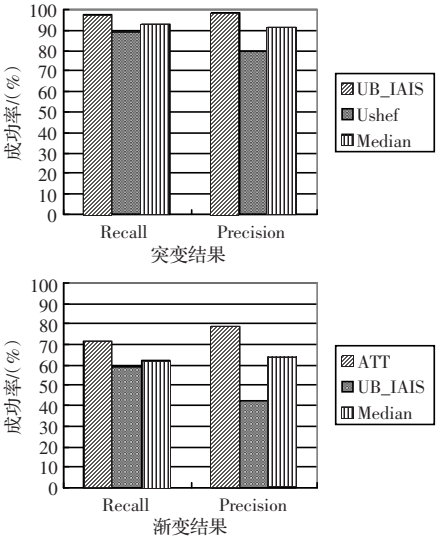

图 2 镜头边界检测结果比较图

(其中 Median ▥ 为该文算法的实验统计结果;▨为所公布结果中最好的;■为所公布结果中表现最差的)

参考文献:

[1] Koprinska I,Carrato S.Temporal video segmentation:A survey[J].Signal Processing Image Communication,2001,16(5):277-500.

[2] Liu Z,Gibbon D,Zavesky E,et al.AT&T Research at TRECVID 2006[C]//Proc of TRECVID 2006 Workshop,2006.

[3] Lee M H,Yoo H W,Jang D S.Video scene change detection using neural network:Improved ART2[J].Expert Systems with Applications,2006,31:13-25.

[4] Tong Zijian,Yuan Jinhui.A new approach for gradual transition detection based on finite-site automata[J].Computer Science,2006,33(1).

[5] Itti L,Koch C,Niebur E.A Model of Saliency-Based Visual Attention for Rapid Scene Analysis[J].IEEE Trans Pattern Analysis and Machine Intelligence,1998,20(11):1254-1259.

[6] Itti L,Koch C.A saliency-based search mechanism for overt and covert shifts of visual attention[J].Vision Research,2000,40(10/12):1489-1506.

[7] Itti L,Koch C.Computational modeling of visual attention[J].Nature ReViews Neuroscience,2001,2(3):194-203.

[8] Greenspan H,Belongie S,Goodman R,et al.Overcomplete steerable pyramid filters and rotation invariance[C]//Proc of IEEE Computer Vision and Pattern Recognition,Seattle,Wash,June,1994:222-228.

[9] Hsu C W,Chang C C,Lin C J.A practical guide to support vector classification[R].Department of Computer Science,National Taiwan University,2003.

[10] Hanjalic A.Shot boundary detection:Unraveled and resolved?[J].IEEE Transactions on Circuits and Systems for Video Technology,2002,12:90-105.

[11] Guidelines for the TRECVID 2007 evaluation[EB/OL].http://www-nlpir.nist.gov/projects/tv2007/#5.4.