基于模糊粗糙集的新闻视频镜头边界检测方法

韩 冰, 高新波, 姬红兵
(西安电子科技大学电子工程学院, 陕西西安 710071)

摘 要: 镜头边界检测是实现基于内容的视频检索的一个重要步骤. 为了将视频分割成镜头, 现有的方法大都是首先提取大量的特征然后构造相异性测度函数. 然而, 太多的特征就会降低算法的效率. 因此, 有必要对镜头边界检测的规则进行特征约简. 本文将粗糙集中的属性重要性和模糊粗糙集中的分类精度相结合定义了模糊粗糙算子, 并构造了相异度检测函数. 最后给出了镜头边界检测的一般性规则. 由于本文检测方案的自适应性, 因此适合于各种类型的新闻视频. 用来自中央电视台的3个多小时的新闻视频所做的镜头边界检测实验获得了95.4%的查全率和96.1%的准确率.

关键词: 镜头边界检测; 模糊粗糙集; 模糊粗糙算子; 特征约简; 相异度函数
中图分类号: TP391 文献标识码: A 文章编号: 0372-2112 (2006) 06-1085-05

A Shot Boundary Detection Method for News Video Based on Rough-Fuzzy Sets

HAN Bing, GAO Xin-bo, JI Hong-bing
(School of Electronic Engineering, Xidian Univ., Xi'an, Shaanxi 710071, China)

Abstract: As a crucial step in content-based news video indexing and retrieval system, shot boundary detection attracts much more research interests in recent years. To partition news video into shots, many metrics were constructed to measure the similarity among video frames based on all the available video features. However, too many features will reduce the efficiency of the shot boundary detection. Therefore, it is necessary to perform feature reduction for the decision rules of shot boundary. For this purpose, the classification method based on rough-fuzzy sets for feature reduction and the dissimilarity function is proposed. Finally, for each given class, the fuzzy if-then rule is generated for decision with fuzzy inference. The efficacy of the proposed method is extensively tested on more than 3 hour of CCTV news programs and 95.4% recall with 96.1% precision have been achieved.

Key words: shot boundary detection; Rough-Fuzzy set; Rough-Fuzzy operator; feature reduction; dissimilarity function

1 引言

随着多媒体技术的发展, 数字视频以其丰富的信息量而受到高度重视, 但视频数据的增多也带来了不容忽视的问题, 即如何对这些数据进行有效的管理. 视频是由一个个镜头串连而成的. 因此, 在各种视频分析方法中, 首要任务是视频分段, 即镜头分割. 镜头分割的正确与否将直接影响到后续分析步骤的准确性.

镜头分割就是要检测镜头间的连接方式. 根据镜头连接方式不同可将镜头切换分为突变过渡(Abrupt Transition)和渐变过渡(Gradual Transition)两类. 现有算法大都是针对某一种过渡类型检测效果比较好, 对其他类型的镜头过渡性能就比较差. 另外, 这些检测标准中大都是使用象素值、亮度或者彩色直方图、边界信息或者压缩域的数据(如DCT系数、宏块类型、运动矢量等), 如果测量的相异性函数大于某一阈值的话, 就认为发生了镜头过渡[1~3]. 但是很多方法都未给出究竟何种特征对镜头边界检测更为有效. 另外, 如何利用已有的检测特征对所有镜头间的变换进行联合识别和分类已成为当今镜头边界检测的一个研究热点. 在保持检测准确率的前提下, 为提高检测效率, 有必要对镜头边界的特征集进行特征约简, 然后根据选择的特征再建立相异度函数. 基于此, 本文提出了一种基于模糊粗糙集的特征约简和相异度检测函数生成算法.

收稿日期: 2005-05-10; 修回日期: 2006-01-20
基金项目: 国家自然科学基金(No. 60202004); 教育部重点项目(No. 104173); 新世纪优秀人才支持计划

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

首先, 提取用作条件属性的视频序列特征, 同时根据文献[4]的方法给出初始决策属性. 通过计算粗糙集条件属性之间的相关性, 分析特征的重要性程度. 由于实验得到的特征属性值是一个模糊值, 同样计算模糊粗糙集的分类精度. 最后, 我们定义了一个新的属性重要性测度——模糊粗糙算子. 根据各个特征所占的比例, 获得最终的主要特征(前 k 个特征), 将这些特征根据其所占的比例进行加权构造出相异性检测函数. 用真实的新闻视频所做的实验结果表明本文方法的有效性和鲁棒性.

2 粗糙集基本概念

粗糙集(RS)理论[5]是由波兰科学家 Pawlak 在 1982 年提出的一种处理含糊和不确定问题的有效的数学工具. 它的优点是不需要预先知道某些特征或属性的数量描述, 而是直接从给定问题的描述集合出发, 通过不可分辨关系和不可分辨类确定给定问题的近似域(上、下逼近)来描述、揭示出条件属性对于决策属性的重要性, 从而明确有用信息和冗余信息, 发现隐含在大量数据背后的内在规律. 该理论已经渗透到人工智能的各个分支, 在机器学习、知识获取、决策分析、知识发现、专家系统、模式识别等方面也都已有成功的应用. 本节首先介绍有关 RS 的几个基本概念.

定义 1[6] 设(U, R)是 Pawlak 近似空间, 即 R 是论域 U 上的一个等价关系. 若 A 是 U 上的一个模糊集合, 则 A 关于(U, R)的一对下近似 $\underline{A}$ 和上近似 $\overline{A}$ 定义为 U 上的一对模糊集合, 其隶属函数分别定义为

$$\underline{A}(x) = \inf\{A(y) \mid y \in [x]_R\},\ x \in U \tag{1}$$

$$\overline{A}(x) = \sup\{A(y) \mid y \in [x]_R\},\ x \in U \tag{2}$$

$\underline{A}$ 可以理解为对象 x 肯定属于模糊集 A 的隶属程度; $\overline{A}$ 可以理解为 x 可能属于模糊集 A 的隶属程度.

定义 2 设(U, R)是近似空间, $A \in F(U)$, 定义 A 关于(U, R)的近似分类精度 $\eta_R(A)$ 为

$$\eta_R(A) = |\underline{A}| / |\overline{A}| \tag{3}$$

其中模糊集的基数 $|A|$ 定义为 $|A| = \sum_{x \in U} A(x)$, $A(x)$ 表示元素 x 属于模糊集合 A 的程度; 当 $|\underline{A}| = 0$, 则 $\eta_R(A) = 1$.

3 基于模糊粗糙集的镜头过渡检测

3.1 属性选择

在现有的算法中, 最常用的是直方图特征以及统计特征. 根据大量的文献[1~4,7,8], 我们总结出了以下 5 类共 12 种常用特征. 从这些特征中找到镜头检测的重要特征, 根据这些重要特征来构造检测测度函数, 从而达到有效的检测. 具体特征选取和表示为: (1) 图像的红色(R)、绿色(G)和蓝色分量(B); (2) 图像的色度(H)、饱和度(S)和亮度分量(V); (3) 灰度直方图(G-H); (4) 颜色直方图: RGB 颜色直方图(Rgb-H)和 HSV 颜色直方图(Hsv-H); (5) 统计特征: 均值(M)、方差(St)和偏态(P).

3.2 基于模糊粗糙集的镜头边界检测

在本文中, 将视频的各种特征作为条件属性集合 $P = \{c_1, c_2, \cdots, c_n\}$, 镜头的过渡类型作为决策属性 $Q = d$, n 为总的属性个数(即特征的个数). 在此, 我们定义一个新的度量属性重要性的测度, 即模糊粗糙算子:

$$\kappa_i(Q) = \eta_{P_i}(Q)^* \gamma_{P_i}(Q),\ i = 1, 2, \cdots, n \tag{4}$$

由于从视频序列提取的各个特征集合可以看作是模糊集合, 即 P 为模糊集合. 其中, $\gamma_P(Q) = |POS_P(Q)| / |U|$, $POS_P(Q)$ 是 Q 的 P 正域[5], 则 $\gamma_{P_i}(Q)$ 是第 i 个条件属性特征 P_i 对于分类决策 Q 的分类精度, 是对应于普通的粗糙集模型, 因此在计算 $\gamma_{P_i}(Q)$ 的值时需要将条件属性集合进行离散化, 在此我们利用图 1 所示的特征值隶属度函数来进行(关于隶属度函数的选择在下面将作进一步地解释). $\eta_{P_i}(Q)$ 是第 i 个条件属性特征 P_i 对于分类决策 Q 的模糊粗糙精度, 其定义如式(3)所示, 它是针对模糊粗糙集模型而言的. 为了避免在普通粗糙集模型中数据离散化对结果造成的影响, 我们将模糊粗糙集中的分类精度引入, 定义了上述衡量属性重要性的模糊粗糙算子 $\kappa_i(Q)$, 它表示第 i 个属性特征对于决策 Q 的重要性程度.

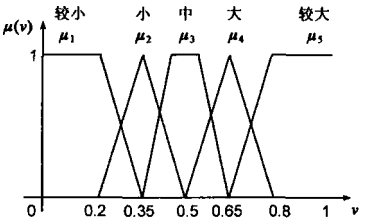

图 1 特征值的隶属度函数定义

将属性的重要性按照由大至小的顺序进行排列. 若前 k 个属性的重要性占整个属性重要性的百分比超过一个预先设定的门限 T, 即 $\sum_{j=1}^{k} \kappa_j / \sum_{j=1}^{n} \kappa_j \geqslant T$, 则取前 k ($j = 1, 2, \cdots, k$) 个属性作为重要属性. 在此, 我们定义一个自适应门限 $T = c \cdot \sum_{j=1}^{n} \kappa_j$, c 为一常数, 后面将讨论 c 的取值.

根据以上的分析, 我们可以定义镜头边界检测的相异度函数:

$$S = \sum_{j=1}^{m} \omega_j (I_j(t+1) - I_j(t)) \tag{5}$$

其中 m 为所选的条件属性的个数. $I_j(t)$ 为第 t 帧第 j 个属性的值, $\omega_j = \kappa_j / \sum_{j=1}^{k} \kappa_j$, $j = 1, 2, \cdots, k$.

视频数据是海量的, 计算机不可能同时处理大量的数据, 所以, 首先将视频进行分段处理. 大量的观察发现, 一个小的新闻单元一般持续时间不会超过 150s, 并且一个镜头过渡不可能持续太长时间, 即一般不超过 5s, 因此选取的视频段长度为 150 帧, 每次处理 300 段视频. 由于在新闻视频中存在的过渡类型有突变、溶解、划变、摄像机摇镜头

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

以及无场景变化五类; 决策集包括突变、渐变、无场景变化三类. 具体检测流程如图 2 所示. 首先计算过渡过程中各种特征的变化值, 并做归一化处理. 另外, 构造各个特征的隶属度, 首先考虑各个过渡形式出现的机率几乎相等且为了便于计算, 因而将归一化后的属性值大致 5 等分. 而对于一些渐变过渡具有较大的变化范围, 因此采用等腰梯形隶属度函数; 对于介于突变和渐变以及无镜头变化和渐变之间的过渡不是很明显, 因而采用三角形隶属度函数; 针对无镜头过渡和 突变过渡具有明显的变化这种特点特征, 因而采用直角梯形的隶属度函数. 据此建立了如图 1 所示的各个特征的隶属度. 根据图 1 以及最大隶属度原则将其转化为相应的模糊语言量. 然后建立信息表和决策表.

4　实验结果与分析

用中央电视台新闻联播节目的 5 段视频对算法进行了测试, 详细信息如表 1 所示. 其中的无场景变化指的是新闻视频中出现的摄像机的一些运动, 如摇镜头, 伸缩等.

为了验证所提出新闻镜头分割算法的有效性, 采用两个标准: 查全率(Recall) 和准确率(Precision) , 其定义如下:

$$\text{Recall}= n_c/(n_c+ n_m),\ \text{Precision}= n_c/(n_c+ n_f) \quad (6)$$

其中 n_c, n_m, n_f 分别表示正确、漏检和错检的镜头个数.

首先用人工方法对视频进行切分, 结果作为检测到的正确镜头个数, 由于视频序列的数据量很大, 一般认为误差在 10 帧以内都是允许的. 为了证明本文方法的鲁棒性, 采用 Twin Comparison 方法[4] 对相同的视频片段进行了检测比较, 结果列于表 2.

表 1　新闻节目的详细信息

新闻视频	1	2	3	4	5	6	7	总共
长度	34′55″	30′2″	36′26″	24′47″	15′44″	20′38″	28′26″	190′58″
镜头边界数	374	323	396	247	187	245	340	2112
突变镜头数	339	315	390	237	178	225	315	1999
渐变镜头数	35	8	6	10	9	20	25	113
无场景变化	18	9	12	10	8	13	20	90

表 2　本文方法与 Twin Comparison 方法[4] 对比检测结果

测试视频	本文方法					Twin Comparison[4] 方法				
	检出数	漏检数	误检数	查全率	准确率	检出数	漏检数	误检数	查全率	准确率
News 1	372	12	10	96.8%	97.3%	380	30	36	92.0%	90.5%
News 2	320	12	9	96.3%	97.2%	350	15	42	95.4%	88.0%
News 3	392	18	14	95.5%	96.4%	400	46	50	88.4%	87.5%
News 4	245	12	10	95.1%	95.9%	250	22	25	91.1%	90.0%
News 5	185	12	10	93.6%	94.6%	190	17	20	90.9%	89.5%
News 6	243	15	13	94.7%	93.9%	255	18	28	92.7%	89.0%
News 7	340	16	16	95.3%	95.3%	358	22	40	93.5%	88.9%
总共	2097	97	82	95.4%	96.1%	2183	170	241	92.0%	89.0%

其中, 正确检出数= 检出数- 误检数.

7 段新闻视频共有 2112 个镜头边界, 本文的方法检测到 2097 个镜头边界, 其中包括 97 个漏检和 82 个误检. 准确率和查全率分别达到 95.4% 和 96.1%. 检测过程中的漏检主要是由于在一些镜头边界中帧对间小的内容变化, 而大多数错检则是由于在渐变过程中不规则的摄像机运动造成的. 实验发现, 在粗检测阶段的误检将影响后续的特征提取和镜头边界检测; 而其中的漏检对后续的检测结果影响不大, 因为模糊粗糙算子在一定程度上削弱了在粗检测阶段的错误. 因此, 由模糊粗糙算子构造的检测相异度函数更适合于各种视频.

表 3　基于模糊粗糙集的三种过渡类型分类结果

过渡类型	突变过渡	渐变过渡	无场景变化
突变过渡	1999	0	0
渐变过渡	0	108	7
无场景变化	0	5	83
准确率	100%	95.6%	92.2%

表 3 给出了分辨矩阵形式的 3 种镜头变化类型分类结果, 表中最左边一列表示真实的类别而最上面的一行为由分类策略得到的分类结果.

表中在渐变镜头过渡的检测中有 5 处误检为无场景变化. 主要原因一是由于渐变过渡前后两个镜头具有相似的检测特征, 因此被认为只是由于摄像机的摇动所带来的影响, 即无场景变化. 另一误检是渐变中的溶解变化前后过渡时间非常短造成的; 另外, 对于无场景变化的检测中有 7 处被检测为渐变过渡, 主要是由于在镜头摇动的过程中速度过快以及摇动过程中有物体的运动造成的.

我们选择 News3 作为例子, 来具体分析一下本文算法的性能. 使用 Twin Comparison 方法[4] 对前 300 段视频进行预检测的时候出现了误检, 即导致了决策属性中有错误的分类. 如在 News3 中, No. 8 为一处突变过渡, 由于过渡前后帧的直方图具有相似的结构而被漏检, 即被错判成无

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

场景变化; No. 14 为一处溶解过渡被误检成了无场景变化; No. 25 原本为摄像机镜头的拉伸即无场景变化被检测为渐变过渡. 因此, 直接计算属性的重要性会有很大的误差, 从而造成检测结果明显降低. 以下就来分析一下 Twin Comparison 对本文算法的影响. 首先, 根据属性值所对应的隶属度, 我们可以计算得到各个条件属性的重要性 l_{P_i}, $i = 1, 2, \cdots, 12$. 则根据门限 $T = 1/2 \cdot \sum l_{P_i}$ 可得重要属性为 P, H, Hsv-H 和 S. 如果仅仅根据这四个特征进行检测的话, 查全率和准确率分别为 93.8% 和 92.2%. 如果假设 Twin Comparison 方法能够得到正确的结果, 即可以得到正确的初始决策, 则可以计算得到各个条件属性的重要性 l_{P_i}, $i = 1, 2, \cdots, 12$. 得到的重要属性为 P, HSV, RGB, V, H, 因为 V 和 H 都有相同的分类能力, 就人类视觉观点来说, V 较之 H 更为敏感, 所以我们最终选择的特征为 P, Hsv-H, Rgb-H, V. 用这四个特征做检测所得的查全率和准确率分别为 95.2% 和 93.4%, 可见使用不好的检测特征会使检测效率降低. 这是由于 Twin Comparison 方法造成的错误决策最后得出的重要属性中有特征 S, 即对比度特征. 这显然不太符合我们的经验判定, 即当镜头发生过渡时, 饱和度的变化并不大. 而当镜头发生过渡是, 一般情况亮度特征会变化明显. 因此对于由 Twin Comparison 方法作出的初始决策, 我们不能直接应用粗糙集的属性重要性来直接进行属性约简. 因此, 进一步定义了模糊粗糙算子来对属性进行约简. 如上面所述, 根据模糊隶属度的关系得到模糊分类精确度为 η_{P_i}, $i = 1, 2, \cdots, 12$. 则可以计算出对应的模糊粗糙算子 κ_i, $i = 1, 2, \cdots, 12$. 最终得到的重要属性为 P, H, Hsv-H, V, 由这四个属性构造的相异性函数最终可以得到 97.0% 的查全率和 95.4% 的准确率. 由此可见, 本文所定义的模糊粗糙算子以及构造的检测相异性函数是行之有效的.

对于以上的七段视频, 分别得出了他们的重要属性, 如表 4 所示, 并据此可以进一步得到属性的规则.

表 4 7 段新闻视频特征选择结果

视频片段	News 1	News 2	News 3	News 4	News 5	News 6	News 7
重要属性	P, Hsv-H, Rgb-H, V	P, Hsv-H, G-H	P, H, Hsv-H, V	P, V, Rgb-H	P, St, G-H, B	P, Hsv-H, V Rgb-H,	P, St, Hsv-H, V

由表 3 可以看出, 对于上述的视频片段, 特征 P, Hsv-H 和 V 是比较重要的. 而这也是非常符合我们平常经验的. 在 News5 中, 特征 B(即蓝色调) 也成为主要特征, 这是因为在该段视频的开始部分是新闻联播的开始部分, 其场景主要都是以蓝色调为主, 因此将蓝色作为主要颜色之一, 这对后面的检测稍有影响. 但是他的加权系数 ω_j 较小, 即所占的比例较低, 因此影响是可以接受的. 经过以上的分析我们可以得出镜头边界检测的一般性规则, 即

If $V = \mu_5$ or $V = \mu_4$ and $Hsv - H = \mu_5$ or $Hsv - H = \mu_4$ and $P = \mu_5$ or $P = \mu_4$, Then N;

If $V = \mu_1$ or $V = \mu_2$ and $Hsv - H = \mu_1$ or $Hsv - H = \mu_2$ and $P = \mu_1$ or $P = \mu_2$, Then A;

If $V = \mu_3$ and $Hsv - H = \mu_3$ and $P = \mu_3$, Then G

其中 N 表示无镜头过渡; A 表示突变镜头过渡; G 表示渐变镜头过渡. μ 表示各个特征值所对应的隶属度表示. 当然这只是一般的决策, 由于视频的多样性, 有时具体的情况还得根据图 1 所示的流程进行处理, 得到更为精确的结果.

对于门限 T 的选择, 本文采用的是基于自适应的方法, 即 $T = 1/2 \cdot \sum_{j=1}^{n} \kappa_j$. 在这里, 取 $c = 1/2$, 并称此时的 T 为临界门限, 即前 k 个特征的分类能力占所有特征分类能力的一半以上. 仍以 News 3 为例, 讨论 c 的取值. 首先增大门限, 即增大 c 的值. 若令 $c = 3/5$, 则需加入特征 S, St 和 Rgb-H 中的两个, 就经验而言, S 对人眼视觉来讲不敏感, 因此我们选择 St 和 Rgb-H 特征. 此时的查全率和准确率分别上升 0.2 和 0.3 个百分点, 而在检测相异度函数中增加了两维特征, 从而使得运算复杂度大大增加. 若降低门限的话, 即 $c = 2/5$. 选择的特征为 P, H, Hsv-H, 所选择的特征减少了, 则查全率和准确率分别下降 2 个百分点和 3 个百分点. 相比较而言, 选择 $T = 1/2 \cdot \sum_{j=1}^{n} \kappa_j$ 是非常合适的, 即它为临界门限.

5 结论

本文提出基于模糊粗糙集的镜头边界检测特征约简方法, 根据约简结果给出了镜头边界检测相异度函数. 通过计算粗糙集条件属性的相关性, 可以得到其重要性. 由于, 特征值集合是模糊集合, 进一步计算得到模糊粗糙集的分类精度. 从而, 我们定义了一个新的条件属性重要性的判定标准 ——模糊-粗糙算子. 根据每一个特征所占重要性的大小, 选取前 k 个特征作为重要特征, 并且通过对这 k 个特征的加权构造镜头边界检测的相异性函数. 实验证明, 该方法不但简单有效而且减少数据维数的同时又最大程度的保留了原始视频图像的信息, 并能获得很高的检测效率.

在进一步的工作中, 我们将本文的方法与其他信息(如音频、字幕等) 相结合, 以期达到更高的检测效果.

参考文献:

[1] BORECZKY J S, ROWE L A. Comparison of video shot boundary detection techniques[A]. Sethi I K. SPIE Conference on Storage & Retrieval for Image & Video Databases[C]. San Jose: SPIE Press, 1996. 170- 179.

[2] GARGI U, et al. Performance characterization of video-shot-change detection methods[J]. IEEE Trans, 2000, CSVT-10(1): 1- 13.

[3] FORD R M, et al. Metrics for shot boundary detection in

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

digital video sequences[J]. Multimedia System, 2000, 8(1): 37- 46.

[4] ZHANG H J, et al. Automatic partitioning of full motion video[J]. Multimedia Systems, 1993, 1(1): 10- 28.

[5] PAWLAK Z. Rough sets[J]. International Journal of Computer and Information Science, 1982, 11(5): 341-356.

[6] DUBOIS D, et al. Rough fuzzy sets and fuzzy rough sets [J]. International Journal of General Systems, 1990, 17(2): 191- 209.

[7] GAO Xin-bo, et al. Unsupervised model-free news video segmentation[J]. IEEE Trans, 2002, CSVT-12(9): 765-776.

[8] HAN Bing, et al. An efficient algorithm of gradual transition for shot boundary segmentation[A]. LU H Q. SPIE Conference on Multispectral Image Processing and Pattern Recognition[C]. Beijing: SPIE Press, 2003. 956-961.

作者简介:

韩　冰　女, 1978 年 6 月出生于陕西省西安市, 现为西安电子科技大学电子工程学院博士研究生, 主要研究方向为: 视频检索、模式识别等. 发表论文 10 余篇.

E-mail: hanbing@lab202.xidian.edu.cn

高新波　男, 1972 年 8 月出生于山东省莱芜市, 教授, 博士生导师, 主要研究方向为: 视频分析与检索、模式识别、人工智能等.

E-mail: xbgao@lab202.xidian.edu.cn

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net