

压缩域中基于支持向量机的镜头边界检测算法

曹建荣¹, 蔡安妮²

(1. 山东建筑大学信息与电气工程学院, 山东济南 250101; 2. 北京邮电大学电信工程学院, 北京 100876)

摘要: 针对如何进一步提高镜头边界检测精度问题, 本文提出了一个基于支持向量机 SVM (Support Vector Machine) 的镜头边界检测算法. 该算法利用视频压缩域中特征, 如宏块类型, 帧间对应宏块 DC 系数差和帧类型将视频帧分为发生切变的帧、发生渐变的帧和非镜头变换帧三类, 从而实现视频的镜头分割. 实验结果表明该算法对摄像机的运动和大物体的进入具有很好的鲁棒性, 且没有大多数算法中阈值选择的困难, 将我们的算法与 2001 TREC 评估中最佳指标进行了比较, 在综合度量查全率和查准率的性能指标 F1 上, 比 2001 TREC 评估中最佳指标高约 8%.

关键词: 支持向量机; 镜头边界检测; 压缩域; 视频

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2008) 01-0203-06

Algorithm for Shot Boundary Detection Based on Support Vector Machine in Compressed Domain

CAO Jianrong¹, CAI Anni²

(1. Shandong Jianzhu University, Jinan, Shandong 250101, China; 2. Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Improving the precision of shot boundary detection is very important. This paper presents an algorithm for shot boundary detection based on SVM (support vector machine) in compressed domain. It uses the features, such as the type of macroblock, the difference between DC coefficients of two co-located blocks in successive frames and the type of frame, to segment a video into the shots by classifying the frames into three classes, namely, the frames of cut change, gradual change and non change. No thresholds, which are often hard to select in most shot detection methods, are involved in our algorithm. Experiments have shown that our algorithm is robust for motion of camera and large objects in video, and the experimental result of our algorithm on TREC-2001 video data set is 8% higher than the best result of 2001 TREC evaluation in F1 comparison when cut and gradual changes are both considered.

Key words: SVM (support vector machine); shot boundary detection; compressed domain; video

1 引言

随着互联网和多媒体技术的不断发展, 基于内容的视频检索技术越来越受到重视, 成为近几年信息检索技术研究的热点课题^[1-4]. 要对视频进行内容检索, 首先要将视频序列分割为镜头, 所谓镜头是指由一个摄像机连续拍摄得到的在时间上连续的若干帧图, 它是基于内容视频检索的一个基本单元. 考虑到视频的大数据量的特点, 实际中分割主要采用基于边界的方法, 即主要通过检测镜头间的分界来确定镜头的转换位置^[5].

一般认为, 镜头之间的转换方式主要有两大类: 突变(切变)和渐变. 突变是镜头间的突然变化, 常在两帧图像间完成. 渐变是许多镜头切换方式的总称, 其特点是从一个镜头缓慢变化到另一个镜头, 常延续十几或几

十帧, 常见的渐变种类有淡入淡出(fade in/fade out)、叠化(dissolve)和擦除(wipe). 对镜头分割方法的研究已有很多, 主要是在非压缩域和压缩域中进行. 在非压缩域中主要有像素比较方法, 基于块的比较方法, 全局或局部直方图的比较, 基于聚类的分割, 基于特征的分割, 基于模型分割, 基于熵的度量方法等. 如今绝大多数视频是以压缩的形式存储, 因此研究在压缩域中如何对视频进行分割成为目前研究的重点. 在压缩域中对视频分割的方法主要有基于 DCT 系数的分割, 基于 DC 分量的分割, 基于 DC 分量和宏块类型的分割, 基于 DCT 系数、宏块类型和运动矢量的分割, 基于宏块类型和比特率信息的分割, 统计序列法等. 在大多数算法中都要涉及到阈值选择问题, 阈值选择的好坏将直接影响到检测的精度. 为此一些研究者提出了一些自动阈值选择的方法适

应不同的视频素材,来提高边界检测的精度^[6,7].上述方法一个详细的综述见文献[8].

近几年,基于有监督和无监督学习分类的方法也应用到镜头边界的检测上.文献[9]利用帧的颜色直方图作为特征,利用模糊k均值聚类的方法将视频帧分为有场景变化帧和无场景变化类帧,对场景变化类帧又进一步分成切变帧和渐变帧,从而完成镜头的分割.文献[10]利用连续帧之间颜色直方图的差作为特征,采用k均值的方法将视频帧分为场景变化帧和非场景变化帧,对场景变化帧进一步处理,得到切变帧和渐变帧,实现镜头的分割.文献[11]采用有监督学习分类的方法得到镜头边界.它选取颜色直方图的帧间差、当前帧镜头运动的概率和当前帧是黑帧的概率作为特征,分别采用k最近邻域分类、Naïve Bayes 概率分类和支持向量机将视频帧分成切变帧和非切变帧,对非切变帧,又采用第二层二类分类器分类出渐变帧,从而完成镜头的分割.文献[12]也利用SVM对视频帧分类来检测镜头的转变.它选择了P帧和B帧的运动矢量作为特征,对每一帧分成6块,对每一块计算x轴和y轴的运动矢量均值和方差,再加上帧内编码的数量,每帧可以得到一个25维的特征矢量.利用这些特征矢量将镜头分割为有镜头操作和物体运动一类以及发生叠化(dissolve)渐变一类,从而将叠化渐变的镜头边界检测出来.

本文利用SVM进行分类的思想,将视频序列帧看成是由发生切变的帧、发生渐变的帧和非镜头变换帧三类帧组成,利用压缩域中各帧的特征,如宏块类型,帧间对应宏块DC系数差和帧类型作为特征,根据选定的滑动窗口,将窗口内的帧特征组成一个多维的特征矢量做为窗口中心帧的特征矢量,对SVM进行训练,将训练好的SVM对视频序列帧进行分类,从而得到镜头的边界,完成对视频的分割.

我们的方法属于有监督学习分类的方法,与文献[11]的方法相比,首先所使用的特征不同;其次我们采用了基于SVM的3类分类器,而文献[11]分两层,每层采用2类分类器,其中只有一层采用基于SVM的分类器.与文献[12]方法相比,一方面选用的特征也不相同,其次我们的方法既检测切变,又检测渐变,而文献[12]的方法只检测渐变中的一种类型叠化,由于视频中存在切变和不同类型的渐变,因此文献[12]的方法有很大的局限性.与文献[11,12]不同的是我们的特征组成考虑了当前帧的前面和后面帧的特征,这是因为镜头分割时只考虑当前帧的特征是不够的,比如若P帧主要包含帧内编码的宏块,显然不是从前面参考帧预测的,很有可能在前面I/P帧和当前P帧之间发生镜头转变,显然当前帧发生镜头转变,其后面P帧的帧内编码数量可能会很大,因此用多帧的特征形成特征矢量进行

分类是必要的.

2 支持向量机基本原理

设一组训练数据 $(x_i, y_i)_{1 \leq i \leq N}$,其中 $x_i \in R^n$, $y_i \in \{-1, 1\}$ 是 x_i 所属类的标号.设训练数据可以被一个超平面

$$y_i(w \cdot x_i + b) > 0, i = 1, \dots, N \quad (1)$$

分开,则训练数据是线性可分的.此时,总能调节向量 w 和标量 b ,使下式成立

$$\min_{1 \leq i \leq N} y_i(w \cdot x_i + b) \geq 1, i = 1, \dots, N \quad (2)$$

这样,离超平面最近的点到超平面的距离是 $1/\|w\|$.式(1)就变成

$$y_i(w \cdot x_i + b) \geq 1 \quad (3)$$

如果满足式(3)的向量集合能被超平面没有错误的分开,并且离超平面最近的向量与超平面之间的距离是最大的,则这个超平面被称为最优超平面.由于超平面到最近点的距离是 $1/\|w\|$,寻找最优超平面就等价于在式(3)的条件下最小化 $\|w\|$.

由于 $\|w\|^2$ 是凸的,在线性约束(3)下最小化 $\|w\|$ 可以通过Lagrange乘子得到.记 $\alpha = (\alpha_1, \dots, \alpha_N)$ 是 N 个非负的与约束(3)有关的Lagrange乘子,最优优化问题归为最大化下式

$$W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (4)$$

约束条件为: $\alpha_i \geq 0, \sum_{i=1}^N y_i \alpha_i = 0$

上式可以用标准的二次规划方法得到.

一旦求出式(4)的解 $\alpha^0 = (\alpha_1^0, \dots, \alpha_N^0)$,最优超平面满足下式

$$w_0 = \sum_{i=1}^N \alpha_i^0 y_i x_i \quad (5)$$

对 $\alpha_i^0 > 0$ 且式(3)等号成立的那些点称为支持向量.超平面的判决函数可以写为

$$f(x) = \text{sgn} \left[\sum_{i=1}^N \alpha_i^0 y_i x_i \cdot x + b_0 \right] \quad (6)$$

当数据是线性不可分时,引入松弛变量 (ξ_1, \dots, ξ_N) 且 $\xi_i \geq 0$,这样就有

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, i = 1, \dots, N \quad (7)$$

亦即允许有部分错误分类的点.求解最优超平面归结为求下列问题的解

$$\min_{\xi, w, b} \frac{1}{2} w \cdot w + C \sum_{i=1}^N \xi_i \quad (8)$$

约束条件是式(7)且 $\xi_i \geq 0$. C 是常数.

对非线性支持向量机,首先通过一些非线性映射将数据映射到高维特征空间,在这个特征空间中构造最优超平面.设将 x 通过函数 $\phi(x)$ 映射到特征空间,则

式(4)变为

$$W(\mathbf{a}) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \quad (9)$$

令 $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$, 称为核函数, 对称且满足 Mercer 条件, 则有

$$W(\mathbf{a}) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (10)$$

判决函数为

$$f(\mathbf{x}) = \text{sgn} \left[\sum_{i=1}^N \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b_0 \right] \quad (11)$$

支持向量机最初是为两类分类问题设计的, 对于多类问题已经研究了不同的多类分类方法. 主要有修改 SVM 的设计, 直接用于多类分类^[14] 和采用几个二元分类器组合的方法^[14, 15]. 根据比较研究, 这些方法的精度差不多^[14]. 多类分类的问题不再详述, 可参阅相关的文献^[13-16].

3 基于 SVM 的镜头边界检测算法

3.1 特征的选取

本文算法是在压缩域进行的, 首先提取压缩域的特征, 组成训练和测试数据. 选取的压缩域特征如下所述.

(1) 帧内编码块的数量: 如图 1(a) 例所示. 对 I 帧, 所有的宏块都是帧内编码的; 对 P 或 B 帧, 这个参数的

值从几乎等于零到图像宏块最大值变化. 若 P 帧主要包含帧内编码的宏块, 显然不是从前面参考帧预测的, 很有可能在前面 I/P 帧和当前 P 帧之间发生镜头转变.

(2) 前向预测宏块的数量: 如图 1(b) 例所示. 对 I 帧, 该参数为零; 对 B 帧, 若其大多数宏块是从前面 I/P 帧得到的前向预测宏块, 很有可能在当前的 B 帧和下一个 I/P 帧之间有镜头转换.

(3) 后向预测宏块的数量: 如图 1(c) 例所示. 对 I 和 P 帧, 该参数为零; 对 B 帧, 若大多数宏块是后向预测的, 则很有可能前面的 I/P 帧和当前 B 帧之间有镜头的转换.

(4) 双向预测宏块的数量: 如图 1(d) 例所示. 对 I 和 P 帧, 该参数为零.

(5) 相邻两帧对应位置 8×8 块 DC 系数之差大于阈值的数量: 如图 1(e) 例所示. 这个参数表明一个块是“变化块”还是“不变化块”, 差值大于阈值是变化块, 否则是不变化块. 若当前帧变化块的数量占绝大多数, 则很可能发生了镜头变化. 对所有视频片段该阈值都固定为 150, 设置阈值的目的是克服视频帧图像的噪声对检测的影响.

(6) 帧类型: 对视频中所有 I 帧标记为 0.2; B 帧标记为 0.5; P 帧标记为 0.8.

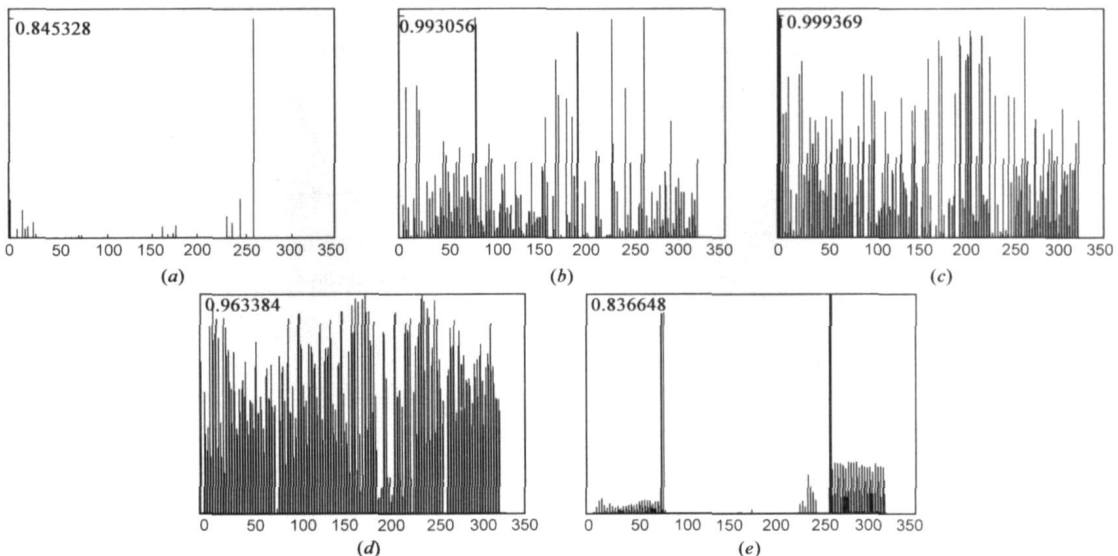


图 1 归一化后 P 帧帧内编码 (a), B 帧前向预测 (b), B 帧后向预测 (c), B 帧双向预测帧内 (d), DC 差 (e) 的直方图举例

3.2 训练和测试数据的建立

前面分析过只用单个帧的特征数据对帧进行分类是不够的, 因为每帧是否是镜头转换帧是与其前后帧有关的. 为此, 考虑建立一个滑动窗口, 窗口中间位置对应的帧的特征数据由其窗口内前后帧的特征数据和本身的特征数据组成, 设窗口的宽度为 n , 而每个帧有 6 个特征, 则组合后每帧的特征总数为 $6 \times n$ 个. 改变窗口的宽度, 就可以改变每帧特征的总数, 即改变每帧特征的维

数. 对用于训练和测试的视频片段半解码, 提取上述的特征, 然后根据滑动窗将训练数据和测试数据准备好. 通过不同窗口下边界检测结果的对比实验, 窗口宽度为 5 时, 边界检测的精度最高, 因此选取滑动窗宽度为 5.

3.3 支持向量机的构造和训练

支持向量机的基函数采用径向基函数 RBF (Radial Basis Function), 其表达式是

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp \left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2 \right), \quad \gamma > 0 \quad (12)$$

对多类分类采用了一对一的方法 (one against one)^[15]来构造 $k(k-1)/2$ 个分类器, 其中 k 是类别的个数, 这里只有切变、渐变和非镜头变换帧 3 类. 文献 [17] 首先将此方法用于 SVM. 每个分类器训练来自不同类的数据, 对来自第 i 类和第 j 类的训练数据, 解决了下列二元分类问题:

$$\min_{w^i, b^i, \xi^i} \frac{1}{2} (w^i)^T w^i + C \left(\sum_i (\xi^i)_i \right) \quad (13)$$

约束条件:

$$\begin{cases} (w^i)^T \phi(x_i) + b^i \geq 1 - \xi^i, & \text{如果 } x_i \text{ 属于 } i \text{ 类,} \\ (w^j)^T \phi(x_i) + b^j \leq -1 + \xi^j, & \text{如果 } x_i \text{ 属于 } j \text{ 类,} \\ \xi^i \geq 0 \end{cases}$$

在分类时采用了投票 (Voting) 策略, 即每个二元分类器输出可看成一票, 数据点根据得到最大票数决定它是属于那一类. 若大多数分类器认为该数据点属于 i 类, 则该数据点划为 i 类.

SVM 的实现是基于 LIBSVM^[18], 我们使用了 C-SVM^[13]. 我们分别采集了 18 段视频片段, 这些片段中有镜头的淡入淡出、叠化、翻页等渐变类型和扫视和变焦以及物体的进入等情况. 我们随机将其分成两组, 一组由 10 段视频片段组成, 用于训练; 一组由 8 段视频片段组成, 用于测试. 对采集的视频片段在压缩域中提取前面所述的每帧的特征, 归一化处理后, 选取窗口宽度为 5, 形成训练和测试的数据. 用训练数据对 SVM 进行交叉对比训练, 得到最优的参数 $C=90$ 和 $\gamma=0.28$.

4 实验结果

根据得到的最优参数 C 和 γ 构造好 SVM, 然后对测试数据进行测试, 每段视频分类结果首先进行校正, 方法是: 对检测到的切变帧如果其前后的帧都是渐变帧且其前或后的渐变帧数量大于 2, 则说明该帧判别有误, 改为渐变帧. 对检测到的渐变帧, 从起始帧开始依次检查后面连续的帧是否也是渐变帧, 若不是则是误检, 改为非镜头变换帧; 若是继续检查后续的帧, 若连续的渐变帧中间有不多于两个的非镜头变换帧, 说明这几个帧检测有误, 将其改为渐变帧. 若两个渐变段之间是黑帧, 则合并这两个渐变段.

镜头检测算法的评价指标通常选用查全率 (Recall) 和查准率 (Precision) 这两个参数, 它们的定义如下:

$$\text{查全率} = \frac{\text{正确检出数}}{\text{正确检出数} + \text{漏检数}}$$

$$\text{查准率} = \frac{\text{正确检出数}}{\text{正确检出数} + \text{误检数}}$$

查全率和查准率越高说明检测方法越好. 一个好的镜头边界检测器应该同时具有较高的查全率和查准率, 常用 $F1$ 来综合度量查全率和查准率, 即

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

只有当 precision 和 recall 都高时, $F1$ 才高.

我们采用 TREC-2001 视频数据集^[19]来评估我们的镜头边界检测算法. 这个视频数据集主要由记录片风格的视频组成, 而且视频年代、风格和质量很不相同, 且视频中有很多镜头的操作 (如缩放、水平和上下移动等) 和物体的运动. 这个视频集由 26 个 MPEG-1 视频片段组成, 是美国国家标准和技术协会 NIST (National Institute of Standards and Technology) 提供的, 供世界各个研究机构 and 高校等对其镜头分割、语义提取等算法进行测试的一个视频数据集, 每年都会举行一次竞赛, 每年选用的视频集也不同. 利用我们前面训练的 SVM 分类器来检测这个视频集合的镜头边界并与 NIST 提供的边界结果比较, 得到我们算法的性能. 另外, 我们还将我们的算法与文献 [9] 方法、文献 [10] 方法在同一视频集合检测的结果比较, 总共进行了 3 轮检测. 实验结果如表 1 所示.

表 1 镜头边界检测结果

RUNS	切 变			渐 变			SUM F1
	Recall	Precision	切变 F1	Recall	Precision	渐变 F1	
RUN-1	0.7748	0.5122	0.6178	0.3949	0.3924	0.3936	0.5057
RUN-2	0.6121	0.8992	0.7278	0.7261	0.5000	0.5922	0.6600
RUN-3	0.9332	0.9836	0.9577	0.7626	0.9330	0.8392	0.8985

从表 1 可以看出, 与上述 2 种方法相比, 我们的算法的性能最好. Run-1 和 Run-2 方法只使用帧的直方图差来检测镜头边界, 对于有较多物体和摄像机运动的视频, 由于物体和摄像机的操作也会引起较大的帧差, 因此检测性能较差. 我们的算法不仅考虑了帧差, 还采用了帧的宏块类型等特征, 而且采用滑动窗口选取帧特征组成特征矢量, 得到了较高的检测精度. 我们的算法对切变产生漏检的原因主要是在很短时间内对类似背景切变检测产生的; 对某些渐变的误检有时是由于大物体以很快的速度运动引起的. 我们的算法对相对较慢的物体运动能较好防止误检. 图 2 给出了这 3 种方法的 precision vs. recall 曲线.

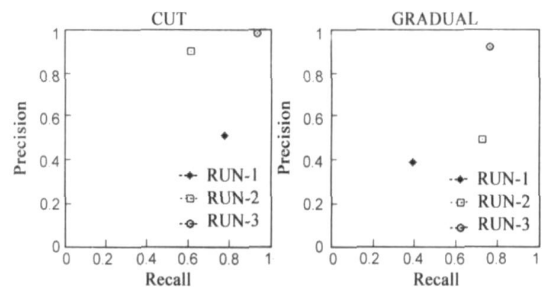


图 2 precision vs. recall 曲线

我们将实验结果还与 2001 TREC 评估^[20]中最佳指标以及文献 [11] 方法 (也采用了 TREC 2001 视频数据集) 进行了比较. 表 2 给出了这 4 种方法 $F1$ 指标比较的

结果.

表 2 $F1$ 指标性能比较结果

RUNS	切变_ $F1$	渐变_ $F1$	SUM_ $F1$
我们方法	0.9577	0.8392	0.8985
TREC 中最佳切变	0.9659	0.6706	0.8182
TREC 中最佳渐变	0.8572	0.7297	0.7934
文献[11]方法	0.9424	0.6983	0.8204

从表 2 中可以看出在切变检测上我们的方法与最佳切变检测的 $F1$ 很接近, 且好于文献[11]的方法, 但渐变要远远好于文献[11]和最佳渐变检测结果, 而且我们的算法在综合性能上最佳.

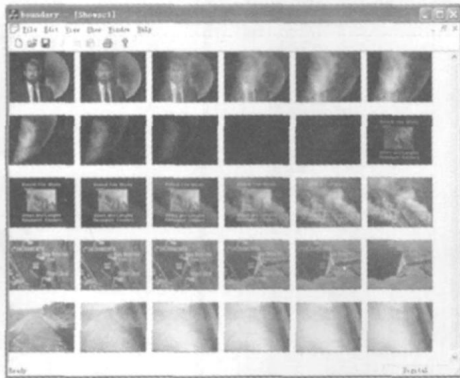


图 3 正确检测到的渐变例子

5 结论

本文利用支持向量机 SVM 通过在视频压缩域中提取宏块类型、帧间对应宏块 DC 系数差和帧类型作为特征对 SVM 进行训练, 利用 SVM 对视频帧分类, 从而实现镜头的转换检测. 实验结果表明该算法没有大多数算法中阈值选择的困难, 对镜头的运动和大物体的进入具有很好的鲁棒性, 通过与其它三种相关的算法进行比较, 我们算法的查全率和查准率最高, 另外我们算法的综合性能指标 SUM_ $F1$ 比 2001 TREC 评估中获得 SUM_ $F1$ 最佳指标的算法高约 8%.

参考文献:

[1] Shih-Fu Chang, William Chen, Horace J Meng, et al. VideoQ: An automated content based video search system using visual cues [A]. In Proc. the 5th ACM Multimedia [C]. Seattle, USA, 1997. 313–324.

[2] Yong Rui, T S Huang, Sharad Mehrotra. Constructing table of content for videos [J]. Multimedia System, 1999, 17(5): 359–368.

[3] Vasileios Mezaris, Ioannis Kompatsiaris, Nikolaos V. Boulgouris, et al. Real-time compressed domain spatiotemporal segmentation and ontologies for video indexing and retrieval [J].

IEEE Trans. on Circuits and Systems for Video Technology, 2004, 14(5): 606–621.

- [4] Jianping Fan, Hangzai Luo, Elmagarmid, A. K. Concept oriented indexing of video databases: toward semantic sensitive retrieval and browsing [J]. IEEE Trans. on Image Processing, 2004, 13(7): 974–992.
- [5] R Lienhart. Reliable transition detection in videos: A survey and practitioner's guide [J]. International Journal of. Image and Graphics (IJIG), 2001, 1(3): 469–486.
- [6] G Boccignone, M De Santo, G Percannella. Automated Threshold Selection for the Detection of Dissolves in MPEG Videos [A]. In Proc. IEEE ICME 2000 (C). New York, 2000, 1535–1538.
- [7] 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, Vol. 32(3): 508–511.
Cheng Yong, Xu De. A method for shot boundary detection using adaptive threshold [J]. Acta Electronica Sinica, 2004, 32(3): 508–511. (in Chinese)
- [8] Koprinska I, Carrato S. Temporal video segmentation: a survey [J]. Signal Processing Image Communication, Elsevier Science, 2001, 16(5): 477–500.
- [9] X Gao and X Tang. Unsupervised video shot segmentation and model-free anchorperson detection for news video story parsing [J]. IEEE Trans. Circuits and System for Video Technology, 2002, 12(9): 765–776.
- [10] B Günsel, A M Feman and A M Tekalp. Temporal video segmentation using unsupervised clustering and semantic object tracking [J]. Journal of Electronic Imaging 1998, 7(3): 592–604.
- [11] Y Qi, Hauptmann A T Liu. Supervised classification for video shot segmentation [A]. Proc. IEEE ICME03 [C]. Baltimore, MD, USA, 2003, vol. 2, 689–692.
- [12] Guo Lihua, Yang Shutang, Li Jianhua, et al. Video shot boundary detection in MPEG compressed sequences using SVM learning [J]. Journal of Electronic Science and Technology of China. 2003, 1(1): 15–17.
- [13] V Vpanik. The Nature of Statistical Learning Theory [M]. New York: Springer Verlag, 1995.
- [14] J Weston, C Watkins. Multiclass support vector machines [R]. Univ. London, U. K., Tech. Rep. CSD-TR-98-04, 1998.
- [15] M Pontil, A Verri. Support vector machines for 3d object recognition [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1998, 20(6): 637–646.
- [16] V Blanz, B Scholkopf, H Bulthoff, et al. Comparison of view-based object recognition algorithms using realistic 3d models [A]. In Proc. ICANN'96 [C]. Berlin, Germany, 1996, 251–256.
- [17] J Friedman. Another approach to polychotomous classification [R]. Technical report, Department of Statistics, Stanford Uni-

versity, 1996.

[18] C C Chang, C J Lin. LIBSVM: a library for support vector machine[OL] . <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.

[19] Videos in the 2001 TREC video retrieval test collection[OL] .

<http://www.operrvideo.org/>, 2006.

[20] A F Smeaton, P Over, R Taban. The TREC-2001 video track framework[A] . In Proc. ITREC-2001[C] . Gaithersburg, Maryland, 2001. 13- 16.

作者简介:



曹建荣 男, 1965年出生于山东省济南市, 济南山东建筑大学信息与电气工程学院副教授、博士. 主要研究方向为数字图像处理、视频分析与检索. E-mail: jrcao@sdjzu.edu.cn



蔡安妮 女, 1943年出生于北京市, 留美博士, 北京邮电大学电信工程学院教授、博导. 主要研究方向为多媒体通信、图像处理和视频分析与检索. E-mail: annicai@bupt.edu.cn