第36卷 第15期 计 算 机 工 程 2010年8月
Vol.36 No.15 Computer Engineering August 2010
·多媒体技术及应用· 文章编号：1000—3428(2010)15—0206—02 文献标识码：A 中图分类号：TN911.73

基于语义的体育视频场景分割方法

华 漫
(中国民航飞行学院计算机学院，广汉 618307)

摘 要：以网球视频为例，提出一种基于语义的体育视频场景分割方法。基于网球视频的先验知识设计一个具有6个语义场景的分类器，并根据各个场景的视觉特点提取球场地标线连接点、球场颜色、相机运动模式和人物等可感知特征作为特征。利用支持向量机技术对视频镜头进行语义分类，并给出一种利用聚类提取示例的主动学习算法。对大量网球视频进行实验，结果表明该方法能够得到比传统方法更好的效果。

关键词：视频检索；主动学习；语义

Semantic-based Scene Annotation Method of Sport Video

HUA Man
(School of Computer, Civil Aviation Flight University of China, Guanghan 618307)

【Abstract】This paper presents techniques and results on semantic annotation of sport video with active learning. It takes tennis video as examples and defines 6 scenes containing semantic concept. The mid-level features such as court line, court color percentage, domain motion, camera move pattern etc are extracted through analyzing the visual difference of semantic scenes. The Support Vector Machine(SVM) is used to classify the shots. The proposed method employs a clustering based active learning scheme. It performs most representative samples selection as well as to avoid repeatedly labeling samples in the same cluster. Experimental results show encouraging result compared with the traditional methods.

【Key words】video indexing; active learning; semantic

1 概述

体育视频的场景分割一般是指对视频中不同语义的镜头进行区分，便于构建视频索引。由于体育视频场景的分割往往是内容分析、精彩片段检测等应用的一个重要步骤，因此近年来越来越多的研究者投入到该研究领域中。目前，利用无监督的聚类技术和有监督的示例学习技术是对视频场景进行分类的2种常见思路[1-2]。

在场景分割中，有2个要点：特征选取和示例选取。本文提出一种基于语义的体育视频场景分割方法，从在特征选取和示例选取方面进行改进以获得更好的分类效果。

2 网球场景分类器设计

根据网球比赛的特点，将网球视频预先定义成6类语义类别，如图1所示。

(a)CB (b)CM (c)CC
(d)NM (e)NC (f)NA

图1 语义场景分类示意图

首先分为场地镜头(CS)和非场地镜头(NS)。场地镜头根据摄像机的位置分为远(CB)、中(CM)、近(CC)三类，非场地镜头分为人物近景镜头(NM)、人物特写(NC)和观众席镜头(NA)，这样划分基本上覆盖了网球比赛的所有场景，而且使得每个类别具有较高的语义。图1为场景分类示意图。

下面针对这些场景的特点寻求有效的视觉特征：

由于不同的网球赛事，其场地和观众席的颜色等都有不同，因此采用传统的颜色直方图等低级特征难以满足要求[2]，本文提取的特征分别是：地标线的连接点个数记为 L_a、L_c，球场颜色为 H_c，运动显著性为 M_L，镜头运动能量为 M_g，人物特征为 P_f。

2.1 地标线特征

球场地标线是体育视频区分于其他视频的一个重要特征。利用地标线可以有效地判定球场出现的比例以及摄像机的位置。如何有效地检测和利用地标线是网球视频内容分析的一个关键，与已往的工作不同[3]，本文先检测出所有的线条，并在检测结果中筛选出白色的线作为地标线候选，并用一种修复和检测其连接点的方法确定地标线，具体做法是首先将彩色图像转换成灰度图像并进行Canny转换，在Canny转换后的边缘图中利用Hough算法检测线，对于检测到的线获取其颜色，如果颜色大于一个白色的阈值(例如RGB颜色空间中的(200,200,200))，则为地标线候选。在检测线过程中，

基金项目：国家自然科学基金资助项目(60879023)
作者简介：华 漫(1976－)，男，讲师、硕士，主研方向：图形图像处理，信息安全
收稿日期：2010-02-15 **E-mail：**hua.man@163.com

常常会遇到遮盖等问题，导致地标线不连续，从而导致提取的同一根地标线由不连续的几根线组成，为了解决该问题，给出一种遮盖地标线修补方法，该法将被遮蔽的地标线恢复出来，具体做法如下：

(1)读取任意 2 个线段 L1((x11,y11), (x12,y12)), L2((x21,y21), (x22,y22)):

(2)IF |atan(y12-y11,x12-x11)-atan(y22-y21,x22-x21)|< δ

THEN IF |atan(y21-y11,x21-x11)- atan(y12-y11, x12-x11) < δ

DrawLine(x12,y12,x21,y21);

(3)Goto step 1,until 所有线段比较完毕

其中，atan 为反正切函数；δ 为一个容忍阈值，设置成 0.02。再将地标线连接点检测出来，即一条线段的端点和另一条线段重合，则认为检测到一个连接点。

2.2 场地颜色特征

在网球视频中，大多数镜头为场地镜头，但是网球比赛有草地、塑胶、红土等不同场地类型，其颜色各不相同，而且观众席常常在该类镜头中比例也较大，因此利用固定的颜色或者采用包含最多颜色作为场地颜色并不适用于网球比赛。本文用球场线来确定场地颜色，由地标线相交点所围成的最小区域作为球场区域。设球场帧的颜色直方图为 H_f，地标线围成的区域颜色直方图为 H_r，则球场颜色大致分布区域如下式所示：

$$H_c = 2H_r - H_f$$

为了减少遮挡或者地标线落在屏幕外等干扰，对于每帧进行场地颜色检测，并采用投票(vote)的方式取大多数帧检测到的场地颜色作为镜头场地颜色。

2.3 运动特征

运动特征一般分为全局运动和局部运动。本文分别提取全局和局部运动特征作为运动特征。具体做法如下：对于局部运动，人眼对于运动的反差比较敏感，因此，可以用一种基于反差的运动显著图特征作为局部运动特征[4]。

运动特征中另一个重要特征是镜头运动，在此采用文献[5]中的 6 参数相机运动的仿射变换模型，计算平均每帧的相机运动能量。则镜头的全局运动特征为 $M_g = \frac{1}{F}\sum_{k=1}^{F}\sum_{i,j\in R} E_{i,j}$，其中，$E_{i,j}$ 为帧中每块的相机运动能量；F 为镜头的帧数目；R 为帧区域。

2.4 人物特征

在非场地镜头中常常有人物特写镜头，例如运动员、裁判或者观众，该类镜头中常常包含大量的人物。由于拍摄位置的关系，并非所有人物镜头都能检测到正面的人脸，因此采用文献[6]中的方法。本文用 P_f 作为人物特征，定义为 $P_f = \frac{N_f}{N_{all}}$，其中，$N_f$ 为镜头中包含人物的帧数；N_{all} 为镜头所有的帧数。P_f 可以大致表示成镜头中人物出现的比例，计算视频中每个镜头的上述特征，并将所有特征形成一个特征向量 V，用于进一步的样本提取和场景分类。

3 基于主动学习的优化示例选取

主动学习的基本思想是在未标注的示例中选择最有价值的示例交给专家进行标注[7]。为了提高最佳示例查找效率，本文提出一种基于聚类的主动学习优化示例选取方法：首先将训练样本根据其相似性进行大致分类，分类后相似或重复的样本被基本归类，并提取聚类中部分样本的标注结果作为对于该聚类的预测；当分类器对于该聚类中新样本的预测和实际一致时，则表明该样本是和大多数样本预测结果一致的，予以保留；当新样本与聚类的预测不一致时，则认为与大多数样本预测不一致，将其淘汰。

4 实验和结果分析

为了检验算法性能，本文对于实际网球视频进行了实验。实验数据包括各类网球录像视频，总共时长为 140 min。表 1 中 R1 为人工随机标注方法，R2 为本文的方法，R3 为主动学习样本提取法[2]。表 1 给出了 3 种方法的查全率和准确率，其中，查全率和准确率分别定义为

查全率=正确分类的场景数/此类场景个数

准确率=正确分类的场景数/被分为此类场景的个数

表 1 实验结果对比表 (%)

视频名称	R1		R2		R3	
	查全率	查准率	查全率	查准率	查全率	查准率
鸟瞰镜头	81.8	85.7	81.8	90.0	68.1	78.9
场地全景	91.6	90.0	94.3	91.5	83.3	86.7
场地近景	79.6	80.3	83.9	85.5	54.3	69.5
人物特写	90.4	86.7	90.4	83.1	73.1	69.0
人物近景	84.1	87.6	90.3	93.4	74.0	77.2
观众席	77.3	73.4	86.7	89.0	54.6	66.1
平均	84.1	83.9	87.9	88.8	67.8	67.8

从表 1 的数据可以看出，方法 R2 的查全率和查准率在大多数分类中都明显高于采用随机提取示例的方法 R1，这表明整个主动学习策略获得了预期的效果。但在人物特写镜头中，其准确度不如 R1 方法，这可以用增加聚类中初始样本个数来解决。

通过表 1 中 R2 与 R3、R1 的实现结果对比，可以知道本文选取的视觉特征对于不同网球场景具有较好的区分能力。当采用 R3 时，其分类性能下降明显。从实验结果来看，本文提出的基于主动学习的网球视频场景语义分割方法是有效的。

5 结束语

本文通过分析不同体育视频场景的视觉差异，提取了多个高级视觉特征，并基于主动学习思想设计了一个利用聚类将待标注示例进行预分类和提取优化样本的场景分割方法。实验表明，该方法具有稳定性好、分类效率高等特点，能够在大量示例中选取优化示例。但对于提取的高级视觉特征，其耗时较多，难以应用到实时应用中，这也是下一步优化的目标。

参考文献

[1] Duan Lingyu, Xu Min, Tian Qi. A Unified Framework for Semantic Shot Classification in Sports Video[J]. IEEE Transactions on Multimedia, 2005, 7(6): 127-134.

[2] Peng Wang, Liu Zhiqiang. Investigation on Unsupervised Clustering Algorithms for Video Shot Categorization[J]. Soft Computing——A Fusion of Foundations, Methodologies and Applications, 2007, 11(2): 335-360.

[3] Rea N, Dahyot R. Classification and Representation of Semantic Content in Broadcast Tennis Videos[C]//Proc. of IEEE International Conference on Image Processing. Genova, Italy: [s. n.], 2005: 1204-1207.

（下转第 210 页）