一种体育视频中广告牌商标的实时识别算法

卜 江 [1] 老松杨 [1] 白 亮 [1] TOLLARI Sabrina [2] MARSALA Christophe [2]

摘 要 近期, 体育视频分析中的广告牌商标的探测和识别方法已经广泛应用于许多其他领域, 比如商业电视. 基于此, 提出了一种能在不同体育视频 (如足球、篮球和 F1 赛车等) 中进行广告牌商标实时识别的算法, 该算法主要包括两个步骤, 首先, 利用基于模糊决策树的方法进行广告牌图像帧的探测; 其次, 利用颜色特征和局部 SIFT (Scale-invariant feature transform) 特征来描述不同商标的外观, 并最终通过基于潜在语义分析 (Latent semantic analysis, LSA) 的 SIFT 词汇匹配来识别所给定的商标模板. 初步的实验表明了本文算法的有效性, 并且该算法能在实时情况下运行.

关键词 商标识别, 模糊决策树, SIFT 词汇, 潜在语义分析

DOI 10.3724/SP.J.1004.2011.00418

A Real-time Billboard Trademark Recognition Algorithm in Sports Video

BU Jiang[1] LAO Song-Yang[1] BAI Liang[1] TOLLARI Sabrina[2] MARSALA Christophe[2]

Abstract Recently, in the sports video analysis domain, applications like TV commercials can be developed by detecting and recognizing the trademark on the billboard. In this paper, we propose a method for real-time trademark recognition in different sports video such as soccer, basketball, and Formula 1. There are two stages in this algorithm. Fuzzy decision tree based method is used to detect billboard frame in the first stage, while in the second stage, color and regional SIFT (scale-invariant feature transform) features are combined to describe the appearance of trademarks, and latent semantic analysis (LSA) based SIFT vocabulary matching is performed to recognize the given template trademark. The preliminary experiments demonstrate the effectiveness and efficiency of our algorithm.

Key words Trademark recognition, fuzzy decision tree, scale-invariant feature transform (SIFT) vocabulary, latent semantic analysis (LSA)

现今, 随着高速宽带网和数字视频技术 (包括存储、压缩和处理技术) 的发展, 我们已经步入了一个能随时获取大量 (视频) 数据的后古腾堡时代[1]. 视频数据中的文本部分 (例如商标、电视台标和字幕) 承载了大量的语义信息, 因而在探测和识别这些文本内容的基础上能进行许多不同的应用, 例如视频索引、基于关键字的视频搜索和商业电视等. 一般来说, 在商业电视相关领域主要有以下三方面需求: 1) 对于广告商来说, 他们迫切地想知道他们的广告商标在视频中的可视程度是否与他们的广告花费相匹配; 2) 对于消费者来说, 如果能在体育视频中商标位置的附近提供相应的超链接, 将极大地方便消费者浏览相关的内容; 3) 对于广播电视公司来说, 有的情况下广告商要将其商标展现给不同国家的消费者, 这就需要广播电视公司能够根据不同的转播国家及语种对商标的内容进行自适应调整.

商标主要由文本、图形和具有特定含义的图像组成, 它们一般位于放置在体育比赛场地周围的广告牌、旗帜或者其他物理载体之上. 其中, 文本中的内容是广告商的名称, 图形和图像则是有关广告产品和公司的形象符号[2]. 在本文中, 我们仅仅考虑体育比赛转播中常见的场地广告牌上的商标. 在对多个视频观察的基础上, 我们总结了广告牌商标的一些视觉特性如下 (见图 1): 1) 广告牌商标在视频帧中具有较高的对比度; 2) 广告牌商标的前景和背景都具有各自相同的颜色; 3) 广告牌商标的前景通常由许多非连通的线段组成; 4) 相邻商标的背景具有不同的颜色; 5) 广告牌商标一般是刚性并且平坦的对象. 广告牌商标的探测和识别可以归纳为更为一般的对象识别问题, 但是在本质上其具有更大的难度, 主要是因为 (见图 1): 1) 由视频下采样技术、压缩技术限制所导致的图像帧的低质量; 2) 缺乏与之相关联的信息, 例如没有相关字幕同时出现; 3) 由摄像机在比赛场地中的摆放位置和角度所引起的透视失真; 4) 由于运动员或其他障碍物位于摄像机和商标之间所引发的遮挡现象; 5) 由摄像机运动所导致的运动模糊.

在不同的体育比赛类型中进行商标的精确识别

收稿日期 2010-06-30 录用日期 2010-11-23
Manuscript received June 30, 2010; accepted November 23, 2010
国家自然科学基金 (60902094) 资助
Supported by National Natural Science Foundation of China (60902094)
1. 国防科学技术大学信息系统工程实验室 长沙 410073 中国 2. 巴黎第六大学 LIP6 实验室 巴黎 75005 法国
1. Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, P. R. China 2. UMR CNRS 7606 LIP6 Laboratory, Université Pierre et Marie Curie – Paris 6, Paris 75005, France

是一个极具挑战性的工作, 并且一般很难满足实时性的需求. 文献 [1] 为了在足球视频中叠加不同的广告, 作者研究了广告牌的探测和追踪, 该方法虽然考虑了传感器噪声, 但是并没有进行商标的识别. 文献 [2] 描述了一个在体育视频中探测和识别商标的系统, 商标由 SIFT (Scale-invariant feature transform) 特征点进行表示, 并通过匹配商标模版和视频帧商标中的一系列 SIFT 特征描述符来进行识别, 但是该系统没有考虑颜色特征, 并且它同样有许多缺点, 例如时间消耗大和 "远景镜头" 中 SIFT 特征点不足所导致的匹配错误等. 文献 [3] 提出了一种在广播足球视频中结合颜色特征和边缘特征进行广告牌探测的方法, 该方法可以在同一镜头的图像帧中进行感兴趣区域的定位, 但该方法的缺点在于时间消耗大以及对体育类型有限制. 文献 [4] 提出了一个在视频序列中进行商标探测和识别的实时系统, 每个商标的视觉外观由彩色高斯感受域中归一化的多维直方图来进行描述, 并且对不同的识别算法进行了比较, 该系统的缺点是需要有监督者的参与, 并且在候选区域不包含商标的情况下识别效果很差. 当然, 商标识别的相关研究中所面临的挑战还远远不止这些[5−7]. 但是, 我们相信已经讨论了其中很多的关键问题, 如: 利用多种特征来改进识别精确度; 在不同的体育类型中识别给定商标所存在的困难; 实时识别处理的难度; 识别不同大小和有遮挡情况下商标的重要性.

图 1　不同体育视频中的广告牌商标实例 (白色矩形代表广告牌商标, 黑色矩形代表噪声或者障碍物)

Fig. 1　Examples of billboard trademarks in different sports video (White rectangles represent billboard trademarks and black rectangles represent noises or obstacles.)

本文提出了一个两步骤的算法来解决在不同体育类型中进行广告牌商标实时识别所面临的许多重要挑战. 算法主要包括两个步骤, 即广告牌图像帧的探测和商标的识别. 具体来说, 该方法首先利用基于模糊决策树的算法来探测广告牌图像帧, 并在训练过程中通过平衡处理来保证每个类别是均匀分布的, 然后在得到分类概率的基础上进行后处理以改进分类结果. 在探测到广告牌图像帧后, 首先利用颜色特征在广告牌图像帧中定位候选区域, 然后通过基于潜在语义分析 (Latent semantic analysis, LSA) 的局部 SIFT 词汇匹配在候选区域中识别给定的商标. 通过这种方式, 我们得以实现颜色特征与 SIFT 特征在商标识别中的结合. 两步骤的商标识别算法流程图见图 2.

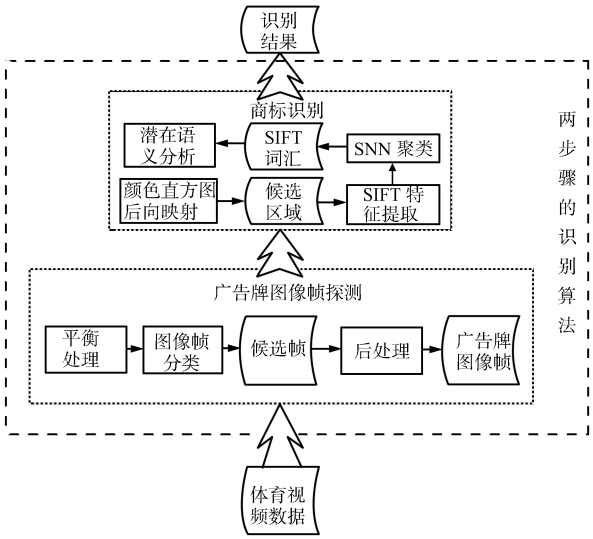


图 2　两步骤的商标识别流程图

Fig. 2　Two-step solution of trademark recognition

这个分步算法的一个新颖之处在于它能够显著地提高商标识别的速度. 首先, 模糊决策树模型通过离线进行训练, 并且仅从关键帧中提取简单和易处理的特征来进行分类; 其次, 它仅对广告牌图像帧进行商标识别, 并且只在广告牌图像帧的候选区域中提取 SIFT 特征点, 图像帧和区域的减少将极大地降低时间的消耗; 最后, 对 SIFT 特征点聚类得到 "SIFT 视觉词", 并且利用其取代 SIFT 特征点来进行匹配工作. 另外一个创新之处是该算法能够极大地改进商标识别的精确度. 首先, 在广告牌图像帧的探测中, 监督式的机器学习方法 (即模糊决策树) 能够消除预先定义阈值所带来的影响, 并且当广告牌出现时, 它能够对图像帧中内在的模式进行建模; 其次, 在商标识别中, 我们利用颜色直方图后向映射来寻找可能包含给定商标的候选区域, 这与传统的 SIFT 特征匹配方法相比, 这种方式能将颜色特征引入到商标识别过程中, 因而能改进商标识别的结果; 最后, 由于 SIFT 特征点是高维数据, 我们利用基于距离的共享最近邻聚类 (Shared nearest neighbor clustering, SNN) 取代传统的 K 均值聚类来搜索不同大小的类, 这使得类的数目参数可以自适应确定.

同时, 与欧氏距离相比, 由于距离考虑了点之间的相对差别, 因而能更好地进行距离的测算.

本文内容的结构如下: 在第 1 节和第 2 节中, 分别对广告牌图像帧探测算法和商标识别算法进行了讨论, 这是本文的重点. 第 3 节则主要介绍在足球、橄榄球、篮球和 F1 赛车比赛视频中的相关实验结果. 最后, 第 4 节将会对全文进行小结并讨论今后的工作方向.

1 广告牌图像帧探测

在广告牌图像帧探测中, 为了消除阈值的影响, 本文提出一种基于模糊决策树的方法在不同的体育视频图像帧中探测可见的广告牌. 同时, 为了加快处理速度, 我们利用一些简单的和具有代表性的特征 (即图像帧不同空间结构中的颜色特征和边缘特征) 作为属性来离线训练分类器. 由于模糊决策树的构造是建立在各个类别平均分布的假设之上, 因此在训练过程中我们加入了平衡处理来保证各个类别的实例是平均分布的. 同时, 在得到分类概率之后, 我们利用一个包括体育类型加权、镜头类型加权、时间连续性和依赖性分析的后处理过程来改进分类结果. 由于篇幅问题, 本文对模糊决策树算法和平衡处理不做具体介绍, 详见我们在文献 [8] 中的相关工作. 下面将主要介绍广告牌图像帧探测中所选用的特征描述符及后处理过程.

1.1 特征描述符

1.1.1 颜色描述符

由于人眼对于高对比度的对象更为敏感, 并且广告商花费巨资想提高其商标在体育视频中的可见度, 因而广播电视公司在播放体育比赛时往往利用颜色来吸引观众的注意力. 为了使广告牌具有更高的对比度, 广告商通常会将广告牌的颜色设置成与背景和场地不同的颜色, 如黑色、蓝色或白色等. 因此, 在广告牌图像帧的探测中, 颜色提供了一种计算快速并且可靠的特征来探测潜在的广告牌商标区域. 本文中我们利用两种类型的颜色特征: 1) 归一化的 RGB 值; 2) HSI 空间中的色调与饱和度分量.

已有相关的研究利用归一化的 RGB 值在变化的成像环境下进行对象的鲁棒表示, 这种对象表示方法在照明强度和对象几何外观变化的情况下能保持不变. 归一化的 RGB 值 r, g, b 由式 (1) 所定义:

$$r = \frac{R}{R+G+B}$$
$$g = \frac{G}{R+G+B}$$
$$b = \frac{B}{R+G+B} \tag{1}$$

HSI 分别指的是色调 (Hue)、饱和度 (Saturation) 和亮度 (Intensity). HSI 模型在描述与人眼感知有关的颜色时比 RGB 模型更为精确, 且其计算仍然非常简单. 由于广告牌的亮度会随着天气和照明条件的变化而变化, 因此在本文中不考虑亮度分量.

在模糊决策树构造和分类过程中, 本文将提取图像帧的 r, g, b 以及 h, s 值作为每个图像帧实例 (Instances) 的五个属性 (Attributes) 用于训练和测试.

1.1.2 Sobel 边缘描述符

边缘一般用来分离图像中不同的内容, 它在颜色或者亮度有变化的情况下是鲁棒的. 由于体育视频中的商标具有高对比度, 并且根据上述对广告牌商标的视觉特性分析可知, 广告牌商标的前景通常由许多非连通的线段组成, 因而广告牌商标区域通常包含着大量的边缘. Sobel 边缘描述符是一个用于计算图像灰度级函数近似梯度的离散微分算子, 它利用一个小的整型模板在水平和竖直方向分别对图像进行卷积, 因此时间消耗比较少.

本文将在图像帧的水平和竖直方向运用 Sobel 边缘检测算法以提取梯度值较大的边缘点, 其中, 敏感度阈值 $sensitivitythreshold$ 设为 0.04. 最终, 我们将所检测到的边缘点数目作为每个图像帧实例的一个属性用于模糊决策树的训练和测试.

1.1.3 空间结构描述符

图像帧的空间结构即为图像帧中各个单独元素的布局. 由于广告牌一般出现在比赛场地的周围, 并且视频帧序列中的广告牌一般占据了图像帧的整个宽度, 这使得图像帧中的不同位置和形状区域 (即不同的空间结构) 中出现广告牌商标的概率是不同的, 因而我们要根据广告牌的外形和位置特性来设定相应的空间结构描述符. 本文将空间描述符设为条状子区域, 如图 3 所示, 其中条状子区域的宽度设为 $height/10$ (其中, $height$ 为图像帧的高度) 个像素.

在模糊决策树构造和分类过程中, 本文将在条状子区域中进行第 1.1.1 节和第 1.1.2 节中所提到的颜色和边缘特征提取, 即首先提取每个图像帧的 r, g, b 和 h, s 值, 然后计算其在每个条状子区域中的均值和方差, 并同时统计每个条状子区域中的边缘点个数, 最后将这三个统计值作为三个属性用于训练和测试.

1.2 后处理

从模糊决策树模型中所返回的分类结果是测试帧属于肯定类别的概率值. 为了改进分类结果, 我们

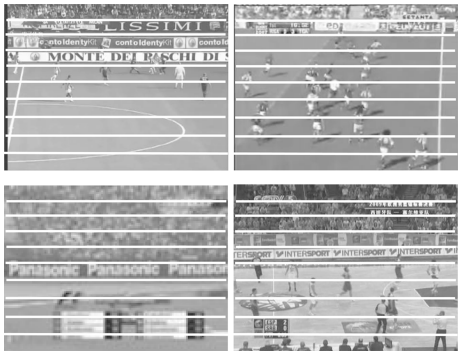
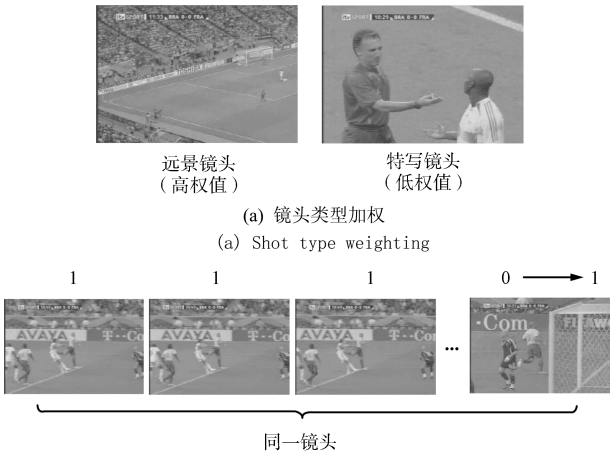


图 3　广告牌图像帧中的条状子区域

Fig. 3　Bar-shape subregions in billboard frames

通过对分类概率值进行体育类型加权、镜头类型加权以及时间连续性和依赖性分析来实现后处理 (见图 4). 体育类型加权和镜头类型加权需要根据不同的镜头类型和体育类型对视频中的图像帧赋予相应的权值, 例如, 图 4 (a) 中, 属于足球比赛视频 “远景镜头” 的图像帧应当被赋予较高的权值, 而属于足球比赛视频 “特写镜头” 的图像帧应当被赋予较低的权值. 另外, 当某个图像帧的分类结果在连续图像帧中出现异常时, 根据时间连续性和依赖性分析则可以对其分类结果进行修改. 例如, 图 4 (b) 中, 除了最后一个图像帧, 其余属于同一镜头中的所有图像帧中都探测到了广告牌的存在, 那么根据时间连续性和依赖性, 该异常的图像帧分类结果将被改变 (由 0 变为 1).



(a) 镜头类型加权

(a) Shot type weighting

(b) 时间连续性与依赖性分析

(b) Temporal continuity and dependeney

图 4　后处理的典型实例

Fig. 4　Typical cases of post processing

2　商标识别

2.1　候选区域搜索

为了降低计算的复杂度, 本文利用候选区域搜索在广告牌图像帧中找到可能包含商标的区域. 一旦用户指定了需要识别的商标, 算法就会从商标数据库中检索相关的示例图像帧和商标模版 (见图 5), 然后完成从示例图像帧到广告牌图像帧的直方图后向映射.

图 5　商标识别的实验数据 (第 1～3 行中的图片分别是商标模板、示例图像帧及其中的商标区域)

Fig. 5　Some experimental data in trademark recognition (Figures from the first row to the bottom row are template trademark, example frames and corresponding trademark areas in the example frames, respectively.)

直方图后向映射是在没有任何空间信息的情况下, 使直方图数据与图像空间域发生联系的一种方法, 它能够给图像中的每个像素点分配概率值, 用以表明其属于需要识别的商标的可能性. 它的实现过程是首先计算示例图像帧中所需识别商标像素的颜色直方图, 并且通过对许多帧的平均来得到较为鲁棒的直方图模型, 然后直方图比率由式 (2) 计算得到 (在 HSI 空间中的情况, 与第 1.1.1 节一样, 亮度分量不作考虑):

$$H_{\text{ratio}}^{col} = \frac{\sum_{i=1}^{n} H_{TA}^{col,i}}{\sum_{i=1}^{n} H_{EF}^{col,i}}, \quad col \in \{h, s\} \tag{2}$$

其中, $H_{TA}^{col,i}$ 和 $H_{EF}^{col,i}$ 分别是在颜色空间 col 中, 第 i 个示例图像帧商标区域像素点的颜色直方图和所有区域像素点的颜色直方图. 这个直方图比率赋予了 HSI 空间 h, s 分量中每个颜色值属于要识别商标的概率. n 是示例图像帧的数量. 最后, 所得到的直方图比率可以反向映射到每个待测试的广告牌图像帧中, 广告牌图像帧中每个点 p 的概率之和可以通过式 (3) 进行计算:

$$p_{i,j}^{\text{sum}} = p_{i,j}^{h} + p_{i,j}^{s}, \quad 1 \le i \le a, \ 1 \le j \le b \tag{3}$$

其中, $p_{i,j}^{h}$, $p_{i,j}^{s}$ 分别是广告牌图像帧中位于坐标 (i, j) 的点在 h, s 分量中属于要识别商标的概率, a 和 b 分别是广告牌图像帧的高度和宽度.

通过赋予广告牌图像帧中每个像素相应的概率值, 并结合适当的阈值, 我们可以得到广告牌图像帧的二值化表示 (见图 6).

随后, 在二值图像上进行形态学闭操作和连通性分析就能得到所搜索的候选区域 (见图 7). 一般来说, 在广告牌图像帧中只能探测到几个候选区域 (最多 6 个). 这也证明了我们利用搜索候选区域来实现实时商标识别的思想是正确的.

图 6 直方图后向映射的实验结果 (第 1 行和第 2 行分别是测试广告牌图像帧和二值化后的图像帧)

Fig. 6 Results of histogram back-projection (The first row is the test billboard frame and the second row is the binary frame.)

图 7 候选区域搜索 (第 1 行) 和局部 SIFT 特征点提取 (第 2 行) 的实验结果

Fig. 7 Results of candidate region search (the first row) and regional SIFT points extraction (the second row)

2.2 SIFT 词汇匹配

从广告牌图像帧中搜索到候选区域后, 我们分别从模板商标和这些候选区域中提取 SIFT 特征点. SIFT 特征由 Lowe 于 1999 年提出并在 2004 年[9] 对其进行了完善, 该特征对于图像尺度变换和旋转变换保持不变, 并且在视角及亮度变化的条件下是鲁棒的.

一个商标模板 i 可以由一个 SIFT 特征点集 T_k^t 所表示:

$$T_k^t = \{x_k^t, y_k^t, s_k^t, d_k^t, O_k^t\},\ \ k \in \{1, 2, \cdots, N_i\} \quad (4)$$

其中, x_k^t, y_k^t, s_k^t 和 d_k^t 分别表示第 k 个特征点在 x 轴上的位置、y 轴上的位置、尺度和主方向. 元素 O_k^t 是有关特征点的 128 维的局部边缘方向直方图. N_i 是商标模板 i 中所包含的 SIFT 特征点的数量. 上标 t 主要用于区分商标模板和广告牌图像帧的候选区域.

类似地, 广告牌图像帧中的每个候选区域 j 也可以由 SIFT 特征点集 T_k^c 所表示:

$$T_k^c = \{x_k^c, y_k^c, s_k^c, d_k^c, O_k^c\},\ \ k \in \{1, 2, \cdots, M_j\} \quad (5)$$

与传统方法不同的是, 本文引入 "SIFT 视觉词" 取代 SIFT 特征点进行商标模板和候选区域的匹配. 对商标模板中的 SIFT 特征点进行基于 x^2 距离的 SNN 聚类[10] 就可以得到 SIFT 词汇. 为了体现 SIFT 特征点在几何变形和尺度变换下保持不变, 我们仅仅对 SIFT 特征点的局部方向直方图 (即 O_k^t 和 O_k^c) 进行聚类. 对于同一个商标模板 i 中的所有 SIFT 特征点 T_k^t $(k = 1, 2, \cdots, N_i)$, 每一对特征点的 x^2 距离[11] 可以根据式 (6) (见本页下方) 进行计算.

一般来说, SIFT 特征点数据在高维空间中是稀疏的, 从而导致特征点之间的相似性比较小, 因而在高维空间中进行相似度的直接测量并不可靠. 本文改进了 SIFT 特征点之间的相似度量, 即根据 x^2 距离计算得到相似矩阵后, 通过分析其共享最近邻来得到共享最近邻曲线图, 从而消除噪声点并确定中心点, 同时根据这些中心点建立类. 并且, 该算法中聚类的数目是自适应的, 不需要预先设定.

在进行聚类处理之后, 每一个类被看作是一个 "SIFT 视觉词", 并且所有的 "SIFT 视觉词" 构成了 SIFT 词汇. 最终, 我们根据 SIFT 词汇并利用 LSA 来匹配商标模板和广告牌图像帧的候选区域. LSA[12] 在信息检索领域也叫做潜在语义索引, 它被用于分析文档和词之间的联系. 该方法通过建立一个 "词 — 文档" 矩阵, 并利用奇异值分解来寻找该矩阵在低维空间的近似, 最后在低维的潜在语义空间中比较其相似度. 在传统 LSA 的 "词 — 文档" 矩阵 X (见式 (7)) 中, 元素 x_{T_i,D_j} 描述了词 T_i 在文档 D_j 中出现的频率或次数, X 中的每一行对应于一个词, 每一列对应于一个文档.

$$X = \begin{bmatrix} x_{T1,D1} & \cdots & x_{T1,Dn} \\ \vdots & \ddots & \vdots \\ x_{Tm,D1} & \cdots & x_{Tm,Dn} \end{bmatrix} \quad (7)$$

$$Dis_{T_{k_i}^t, T_{k_j}^t} = \sqrt{\sum_{d=1}^{128} \frac{1}{O_{k_i}^t(d) + O_{k_j}^t(d)} \left(\frac{O_{k_i}^t(d)}{sum1} - \frac{O_{k_j}^t(d)}{sum2} \right)^2},\ \ sum1 = \sum_{d=1}^{128} O_{k_i}^t(d),\ \ sum2 = \sum_{d=1}^{128} O_{k_j}^t(d) \quad (6)$$

在本文中, 为了测量商标模板和广告牌图像帧之间的相似度, 我们将传统 LSA 中的 “词 — 文档” 矩阵 X 转化为如下的 “SIFT 视觉词” — “广告牌图像帧” 矩阵 $A_{m\times n}$

$$A_{m\times n}=\begin{bmatrix} x_{SV1,BF1} & \cdots & x_{SV1,BFn} \\ \vdots & \ddots & \vdots \\ x_{SVm,BF1} & \cdots & x_{SVm,BFn} \end{bmatrix} \tag{8}$$

其中, 元素 x_{SV_i,BF_j} 描述了 “SIFT 视觉词” SV_i 与广告牌图像帧 BF_j 之间的相关性, 其中的每一行 (SV) 代表一个 “SIFT 视觉词”, 它由 SIFT 特征点的聚类中心所表示, 每一列 (BF) 则代表一个广告牌图像帧. 我们通过转换文献 [13] 中所提出的 OkapiBM 25 相关得分规则来最终产生此矩阵的元素 x_{SV_i,BF_j}:

$$x_{SV_i,BF_j}=sb_f\times\frac{\log(\frac{N-n+0.5}{n+0.5})}{k_1\times((1-b)+\frac{b\times bl}{avbl})+sb_f} \tag{9}$$

式中, x_{SV_i,BF_j} 是 “SIFT 视觉词”SV_i 与广告牌图像帧 BF_j 的相关值, sb_f 是 “SIFT 视觉词”SV_i 在广告牌图像帧中 BF_j 出现的频率, $k_1=2.0$, $b=0.75$, N 是广告牌图像帧的总数量, n 是包含至少一个 “SIFT 视觉词”SV_i 的广告牌图像帧的数量, bl 是广告牌图像帧 BF_j 中包含的所有 “SIFT 视觉词” 的数量, $avbl$ 则是所有广告牌图像帧包含的 “SIFT 视觉词” 的平均数量.

为了衡量商标模板和广告牌图像帧中候选区域的匹配程度, 我们采用余弦距离来测算两者的相似度. 通过对 “SIFT 视觉词” — “广告牌图像帧” 矩阵的奇异值分解, 可得到 “SIFT 视觉词” — “广告牌图像帧” 矩阵 $A_{m\times n}$ 在潜在语义空间中的近似矩阵为 $A_{m\times k}$ (k 为 LSA 中所选取的非零奇异值的数目):

$$\hat{A}_{m\times k}=T_{m\times k}S_{k\times k}D^{\mathrm{T}}_{n\times k} \tag{10}$$

其中, T 和 D 是标准正交矩阵, 即满足 $T^{\mathrm{T}}T=D^{\mathrm{T}}D=I$, $S_{k\times k}$ 是对角矩阵 S_t (通过对 $A_{m\times n}$ 进行奇异值分解可以得到 S_t, 其中 $t=\min(m,n)$) 对角线上的前 k ($k<t$) 个值所组成的对角矩阵.

同时, 我们将商标模板 Tt 表示成 “SIFT 视觉词” 的向量形式 $Tt_{m\times 1}$, 那么 $Tt_{m\times 1}$ 与近似矩阵 $\hat{A}_{m\times k}$ 的相似度可由式 (11) 进行表示.

$$Tt^{\mathrm{T}}_{m\times 1}\hat{A}_{m\times k}=Tt^{\mathrm{T}}_{m\times 1}T_{m\times k}S_{k\times k}D^{\mathrm{T}}_{n\times k}=(Tt^{\mathrm{T}}_{m\times 1}T_{m\times k})(S_{k\times k}D^{\mathrm{T}}_{n\times k}) \tag{11}$$

令 $P_t=Tt^{\mathrm{T}}_{m\times 1}T_{m\times k}$, Q_j 为 $S_{k\times k}D^{\mathrm{T}}_{n\times k}$ 的第 j ($0\le j\le n$) 列, Q_j 则代表着第 j 个广告牌图像帧, 两者的相似度最终由余弦距离测算如下:

$$similarity(Q_j,P_t)=\frac{P_t\cdot Q_j}{\|P_t\|\cdot\|Q_j\|} \tag{12}$$

在实验中, 所选取的非零奇异值的数目 (即式 (10) 和 (11) 中的 k) 对于 LSA 的影响是非常大的, 如果 k 值过于大, 则会导致无法去除噪音, 使得匹配失败; 而如果 k 值过于小, 将会导致具有判别力的信息丢失. 在本文后续的实验中, 我们将 k 值设为 60.

3　实验

3.1　数据准备

在预备实验中, 为了体现实验数据的多样性和广泛性, 本文选择了八个具有不同压缩格式 (MPEG, AVI 和 RMVB) 和不同体育类型 (足球、橄榄球、篮球和 F1 赛车) 的体育视频数据. 这些视频中的广告牌商标都具有较大的尺度变化. 视频数据中图像帧的分辨率为 480 像素 × 360 像素或者 352 像素 × 240 像素, 其中广告牌的大小从 480 像素 × 110 像素到 50 像素 × 13 像素不等. 表 1 列举了所选择的视频数据, 所有视频都由本实验室采集得到.

表 1　体育视频数据

Table 1　Sports video data

视频名称	体育类型	比赛类型	视频格式
AC 米兰 vs. 锡耶纳	足球	意甲联赛	RMVB
阿森纳 vs. 赫尔城	足球	英超联赛	RMVB
拜仁慕尼黑 vs. 斯伯丁	足球	欧洲冠军杯	MPEG
巴西 vs. 法国	足球	世界杯	MPEG
南非 vs. 汤加	橄榄球	世界杯	AVI
F1 正赛	1 级方程式	巴西站	AVI
F1 资格赛	1 级方程式	日本站	MPEG
西班牙 vs. 塞尔维亚	篮球	欧洲锦标赛	AVI

在广告牌图像帧探测中, 我们首先将所选择的视频分割成多个片断, 并且将每个体育类型的片段都分为两个部分, 一半用于训练, 一半用于测试. 在训练片断中, 应用文献 [14−15] 中的镜头探测和镜头分类算法后, 可以从不同镜头中得到平衡的关键帧 (即 “肯定类别” 帧 = “否定类别” 帧). 经过平衡处理后,“否定类别” 帧中的 “远景镜头” 帧 = “中景镜头” 帧 = “特写镜头” 帧, 并且 “肯定类别” 帧中的 “远景镜头” 帧 = “中景镜头” 帧. 对于测试片断, 我们每隔 0.2 秒提取一帧测试图像, 这主要是因为广告牌图像帧在视频中的持续时间一般超过 0.2 秒. 最终, 得到了包含 5 282 个视频帧实例 (足球 3 331 个, F1 赛车 1 073 个, 橄榄球 545 个, 篮球 333 个) 155 个属性的训练数据, 以及包含 6 561

个视频帧实例 (足球 2 739 个, F1 赛车 1 085 个, 橄榄球 1 722 个, 篮球 1 015 个) 155 个属性的测试数据. 并且, 我们将所有的训练和测试图像帧都人工标注为 "1" 或者 "0" 来表示广告牌的出现与否. 在商标识别中, 我们从视频数据中人工选择了 900 个 (足球 312 个, F1 赛车 223 个, 橄榄球 182 个, 篮球 183 个) 包含五类商标的广告牌图像帧, 即 "LG"、"BRIDGESTONE"、"TISSOT"、"VISA" 和 "ADIDAS". 其中 300 个作为示例图像帧, 并且为了计算直方图比率, 我们人工标记出其中的商标区域. 剩余的 600 个则用于测试. 另外, 商标模板从相应品牌的官方网站中下载得到.

3.2 实验结果

实验程序的代码由 VC++ 6.0, DirectX 9.0 SDK 以及 OpenCV 来实现, 其中 DirectX 已经被证明是多媒体处理中非常有效的解决方案.

在广告牌图像帧探测中, 我们利用本实验室所开发的 Salammbô 软件[16] 进行模糊决策树的构造和分类, 并测试相关的参数 (主要参数包括: e 熵阈值; i 叶子节点的最小数目; f 判定准则选择等). 表 2 是利用平衡训练数据和非平衡训练数据对相同的测试数据进行分类所得到的结果. 表 3 则是使用平衡训练数据进行训练, 对分类概率进行后处理之前和之后的比较结果. 在这两个表中, 第 2 ∼ 5 行中的实验结果是使用单个体育类型数据作为训练和测试数据而得到. 本文使用查全率和查准率作为衡量广告牌图像帧探测精确度的标准.

在商标识别中, 我们在商标模板和候选区域之间进行基于 LSA 的 SIFT 词汇匹配, LSA 中的余弦相似度设定为 0.5 (其中, −1 表示完全不相同, 而 1 表示完全相同). 为了便于比较, 我们同时实现了文献 [2] 中的基于欧氏距离的 SIFT 特征点匹配方法. 表 4 是本文方法与文献 [2] 方法的比较实验结果, 两种算法的性能同样根据查全率和查准率进行比较. 图 8 为实验中正确匹配的 "SIFT 视觉词".

3.3 实验结果分析

本文实验中, 广告牌视频帧探测的速度为 47.3 Hz (即 2.4 GHz 处理器处理 6 561 帧的时间为 138.6 s), 如果加上后处理则为 32 Hz (即处理 6 561 帧的时间为 205.1 s). 商标识别的处理速度是 29.6 Hz (即处理 600 帧的时间为 20.3 s). 如果使用

表 2 使用平衡训练数据和非平衡训练数据的比较实验结果

Table 2 Experimental results on non-balanced and balanced training data

数据类型	查准率 (非平衡)	查准率 (平衡)	查全率 (非平衡)	查全率 (平衡)
所有比赛视频	0.706	0.771	0.700	0.760
足球比赛视频	0.706	0.771	0.700	0.760
F1 比赛视频	0.784	0.850	0.803	0.855
橄榄球比赛视频	0.663	0.721	0.654	0.709
篮球比赛视频	0.756	0.825	0.739	0.808

表 3 同一训练集上后处理前后的比较实验结果

Table 3 Experimental results before and after post processing with balanced training data

数据类型	查准率 (后处理前)	查准率 (后处理后)	查全率 (后处理前)	查全率 (后处理后)
所有比赛视频	0.771	0.827	0.760	0.961
足球比赛视频	0.751	0.805	0.737	0.951
F1 比赛视频	0.850	0.886	0.855	0.989
橄榄球比赛视频	0.721	0.772	0.709	0.948
篮球比赛视频	0.825	0.914	0.808	0.978

表 4 商标识别方法的比较实验结果

Table 4 Comparative results of trademark recognition methods

商标名称	包含给定商标的图像帧数量	不包含给定商标的图像帧数量	基于欧式距离的 SIFT 特征点匹配[2] (查准率/查全率)	基于 LSA 的 SIFT 词汇匹配 (查准率/查全率)
LG	235	365	0.758 / 0.628	0.885 / 0.842
BRIDGESTONE	156	444	0.854 / 0.648	0.951 / 0.945
TISSOT	112	488	0.659 / 0.538	0.855 / 0.833
VISA	139	461	0.786 / 0.632	0.869 / 0.835
ADIDAS	208	392	0.804 / 0.752	0.913 / 0.892

图 8 商标模板和广告牌图像帧中正确匹配的 “SIFT 视觉词”

Fig. 8 Matched “SIFT visualword” in template trademark and billboard frame

一个处理器同时运行两个算法, 则处理速度为 30.8 Hz, 以上数据表明我们的算法满足实时性要求.

就广告牌图像帧探测的整体性能来说, 在进行平衡处理和后处理之后, 我们得到了 0.827 的查准率和 0.961 的查全率. 同时, 表 2 和表 3 中的结果表明平衡处理和后处理确实改进了广告牌图像帧探测的效果, 有关模糊决策树和平衡处理的更多内容, 参见我们在文献 [8] 中的讨论. 如果考虑以单一体育类型作为训练数据和测试数据, 在 F1 赛车 (查准率 0.886, 查全率 0.989)、篮球 (查准率 0.914, 查全率 0.978) 数据中, 我们得到了更好的探测结果. 这是因为体育视频具有特定的比赛规则和拍摄规则, 利用这些领域特定的知识, 广告牌图像帧探测得到了更好的结果. 另外一个原因是广告牌的出现与镜头类型的联系非常紧密, 而 F1 比赛和篮球比赛中具有较少的镜头类型, 所以在这些体育类型中进行后处理能极大地改进分类结果. 例如, 在篮球视频中, 查准率从后处理前的 0.825 提高到后处理后的 0.914, 而查全率从后处理前的 0.808 提高到后处理后的 0.978.

在商标识别中, 基于欧氏距离的 SIFT 特征点匹配的处理速度为 23.7 Hz (即处理 600 帧的时间为 25.3 秒), 而基于 LSA 的 SIFT 词汇匹配的处理速度是 29.6 Hz. 这主要是因为在候选区域搜索中, 颜色特征的提取和处理速度远远快于 SIFT 特征点的处理速度. 就查全率和查准率来说, 基于 LSA 的 SIFT 词汇匹配不仅提高了查准率, 同时也减少了错判为肯定类别实例的数量, 这是因为颜色特征的引入过滤了不包含给定商标颜色的区域. 同时, 对 SIFT 特征点的聚类消除了许多 SIFT 特征点噪声, 从而使得 “SIFT 视觉词” 的匹配更加精确. 当分析单个商标的识别效果时, 我们发现 “BRIDGESTONE” 和 “ADIDAS” 取得了更好的识别效果 (“BRIDGESTONE” 的查准率是 0.951, 查全率是 0.945,“ADIDAS” 的查准率是 0.913, 查全率是 0.892), 这主要是因为它们相对于别的商标在外观上更为复杂, 从而提供了更多的 SIFT 特征点. 另外一个重要的原因是大部分的 “BRIDGESTONE” 商标出现在 F1 比赛视频中且相对比较清晰, 而大部分的 “VISA” 商标出现在橄榄球和足球视频的 “远景镜头” 中, 它们比较小并且难以辨认, 而 “TISSOT” 商标在篮球视频中经常出现被遮挡的现象.

最后, 与其他方法 (同样包含探测和识别两个步骤) 相比, 本文的分步骤方法 (探测步骤的查准率 0.827, 查全率 0.961, 识别步骤的查准率 0.89, 查全率 0.87) 比文献 [4] 中的基于高斯派生直方图的方法 (探测步骤的查准率 0.74, 查全率 0.863, 识别步骤的查准率 0.863, 查全率 0.779) 更好. 本文的另一个优势是实验数据更为多样和广泛, 而文献 [4] 的实验数据仅包括 F1 赛车视频和足球视频.

4 结论和未来的工作

本文中, 我们提出了一个两步骤的算法来进行体育视频中广告牌商标的实时识别. 第一个步骤主要进行广告牌图像帧的探测, 并且利用平衡处理和后处理来改进结果, 而第二个步骤则主要在广告牌图像帧上进行商标识别. 在已有的研究中, 本文首次结合颜色特征和局部 SIFT 特征点来对给定的商标模板和广告牌图像帧进行匹配, 并且通过 SNN 聚类得到 “SIFT 视觉词” 来减少算法处理的时间. 预备实验表明了本文算法的有效性, 同时本文算法满足实时性要求.

在实验中, 我们发现了许多问题. 例如, 位于图像帧左边或右边的一些广告牌有时候只能看到一部分, 或者在橄榄球和足球视频的 “远景镜头” 中的广告牌太小且难以辨认, 这些情况会导致能探测到的 “SIFT 视觉词” 相对较少. 今后, 我们可以在连续的图像帧中利用基于运动向量的商标追踪来解决这些问题. 我们未来的工作包括利用更具有代表性的特征来改进模糊决策树分类的结果以及通过正确匹配的 “SIFT 视觉词” 来确定商标的确切位置等.

致谢

感谢法国巴黎第六大学 LIP6 实验室所有同事的帮助.

References

1 Aldershoff F, Gevers T. Visual tracking and localization of billboards in streamed soccer matches. In: Proceedings of the International Conference on Storage and Retrieval Methods and Applications for Multimedia. San Jose, USA: SPIE, 2004. 408−416

2 Bagdanov A D, Ballan L, Bertini M, Bimbo A D. Trademark matching and retrieval in sports video databases. In: Proceedings of the International Workshop on Multimedia Information Retrieval. New York, USA: ACM, 2007. 79−86

3 Watve A, Sural S. Soccer video processing for the detection of advertisement billboards. *Pattern Recognition Letters*, 2008, **29**(7): 994−1006

4 Hall D, Pelisson F, Riff O, Crowley L. Brand identification using Gaussian derivative histograms. *Machine Vision and Applications*, 2004, **16**(1): 41−46

5 Phan R, Androutsos D. Content-based retrieval of logo and trademarks in unconstrained color image databases using color edge gradient co-occurrence histograms. *Computer Vision and Image Understanding*, 2010, **114**(1): 66−84

6 Liu W Y, Sun X M, Wang L. Trademark contour extraction based on improved snake model. *Journal of Computational Information Systems*, 2009, **5**(3): 1253−1260

7 Chattopadhyay T, Sinha A. Recognition of trademarks from sports videos for channel hyperlinking in consumer end. In: Proceedings of the 13th IEEE International Symposium on Consumer Electronics. Kyoto, Japan: IEEE, 2009. 943−947

8 Bu J, Lao S Y, Bai L, Tollari S, Marsala C. Goalmouth detection in field-ball game video using fuzzy decision tree. In: Proceedings of the 5th International Conference on Image and Graphics. Xi′an, China: IEEE, 2009. 917−921

9 Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, **60**(2): 91−110

10 Levent E, Michael S, Vipin K. Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data. In: Proceedings of the 3rd Society for Industrial and Applied Mathematics International Conference on Data Mining. San Francisco, USA: SIAM, 2003. 47−58

11 Ma X X, Grimson W E L. Edge-based rich representation for vehicle classification. In: Proceedings of the 10th IEEE International Conference on Computer Vision. Beijing, China: IEEE, 2005. 1185−1192

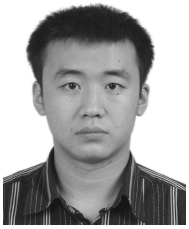
12 Li D D, Kwong C P. Understanding latent semantic indexing: a topological structure analysis using Q-analysis. *Journal of the American Society for Information Science and Technology*, 2010, **61**(3): 592−608

13 Hawking D, Upstill T, Craswell N. Toward better weighting of anchors. In: Proceedings of the 27th Annual International ACM Special Interest Group on Information Retrieval Conference on Research and Development in Information Retrieval. Sheffield, UK: ACM, 2004. 512−513

14 Bai L, Lao S Y, Liu H T, Bu J. Video shot boundary detection using Petri-net. In: Proceedings of the 7th International Conference on Machine Learning and Cybernetics. Kunming, China: IEEE, 2008. 3047−3051

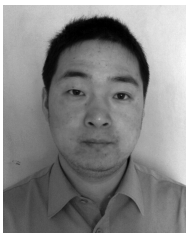
15 Halin A A, Rajeswari M, Ramachandram D. Shot view classification for playfield-based sports video. In: Proceedings of the IEEE International Conference on Signal and Image Processing Applications. Kuala Lumpur, Malaysia: IEEE, 2009. 410−414

16 Marsala C, Bouchon-Meunier B. An adaptable system to construct fuzzy decision trees. In: Proceedings of the 18th International Conference of the North American Fuzzy Information Processing Society. New York, USA: IEEE, 1999. 223−227



卜 江 国防科学技术大学信息系统与管理学院博士研究生. 2004 年获山东大学数学与系统科学学院学士学位. 2006 年获国防科学技术大学理学院硕士学位. 主要研究方向为体育视频处理, 计算机视觉和模式识别. 本文通信作者.
E-mail: veron9@163.com
(**BU Jiang** Ph. D. candidate at the School of Information System and Management, National University of Defense Technology. He received his bachelor degree from Shandong University in 2004 and master degree from National University of Defense Technology in 2006. His research interest covers sports video processing, computer vision, and pattern recognition. Corresponding author of this paper.)

老松杨 国防科学技术大学信息系统与管理学院教授. 主要研究方向为视频基于内容的分析和理解, 视频摘要与可视化. E-mail: laosongyang@vip.sina.com
(**LAO Song-Yang** Professor at the School of Information System and Management, National University of Defense Technology. His research interest covers video content analysis and understanding, video summarization and visualization.)



白 亮 国防科学技术大学信息系统与管理学院讲师. 主要研究方向为多媒体信息系统和智能信息处理.
E-mail: xabpz@163.com
(**BAI Liang** Lecturer at the School of Information System and Management, National University of Defense Technology. His research interest covers multimedia information system and intelligent information processing.)

TOLLARI Sabrina 法国巴黎第六大学副教授. 主要研究方向为信息检索, 图像处理和机器学习.
E-mail: sabrina.tollari@lip6.fr
(**TOLLARI Sabrina** Associate professor at Université Pierre et Marie Curie − Paris 6. Her research interest covers information retrieval, image processing, and machine learning.)

MARSALA Christophe 法国巴黎第六大学副教授. 主要研究方向为模糊学习, 模糊数据挖掘和机器学习.
E-mail: christophe.marsala@lip6.fr
(**MARSALA Christophe** Associate professor at Université Pierre et Marie Curie − Paris 6. His research interest covers fuzzy learning, fuzzy data mining, and machine learning.)