第 36 卷 第 7 期 计 算 机 工 程 2010 年 4 月
Vol.36 No.7 Computer Engineering April 2010
·多媒体技术及应用· 文章编号：1000—3428(2010)07—0227—03 文献标识码：A 中图分类号：TP393

基于内容的视频拷贝检测研究

刘 红[1,2]，文朝晖[2]，王 晔[1,2]
(1. 第二军医大学网络信息中心，上海 200433；
2. 复旦大学计算机科学技术学院，上海 200433)

摘 要：提出基于图的视频拷贝检测方法，该方法将视频序列匹配结果转换为匹配结果图，进而将视频拷贝检测转换成在匹配结果图中查找最长路径的问题。实验结果显示基于图的序列匹配算法拷贝定位准确度高，可弥补图像底层特征描述力不足的缺陷，节约检测时间，批量定位 2 段视频序列中可能存在的多段拷贝。

关键词：视频拷贝检测；聚类；最长路径

Study on Content-based Video Copy Detection

LIU Hong[1,2], WEN Zhao-hui[2], WANG Ye[1,2]
(1. Network Information Center, Second Military Medical University, Shanghai 200433;
2. School of Computer Science, Fudan University, Shanghai 200433)

【Abstract】This paper proposes an efficient graph-based approach for video copy detection. It converts the video sequence matching results to a matching results graph, so the problem of video copy detection becomes a problem of finding the longest path in the matching results graph. Experimental results show that the graph-based video sequence matching algorithm has many advantages, such as high copy locating accuracy, making up for inadequate description of image low-level features, saving detecting time, batch detecting the more than one copy which exist in two video sequences and so on.

【Key words】video copy detection; clustering; longest path

1 概述

随着多媒体硬件和软件技术的快速发展和广泛应用，每天都有数以万计的数字视频产生和发布，这些视频又通过不同的工具进行编辑、转换等操作变成其他的多个版本，最终可能存储在多种不同介质上，如磁带、VCD、DVD，或者通过互联网进行分发。因此，如何检测这些接近拷贝的视频成了一个非常有实际意义的研究课题，因为它有广泛而全新的应用，如版权保护、广播监测/过滤、内容跟踪和管理、商业智能。举一个具体的例子，当视频正在播放时，在线的视频拷贝检测能够警告可能存在的版权侵犯，以便提供即时的知识产权保护。在广播域中，广告客户可以监测网络电视(IPTV)频道或视频流媒体网站，检查其广告节目是否按实际合同以正确的时间与频率播出。当然，视频拷贝检测在技术上也是一个非常具有挑战性的课题。目前，主要有 2 种技术用于视频拷贝检测：数字水印技术和基于内容的方法。本文提出了一种基于图的视频拷贝检测方法。

2 特征选择与视频拷贝检测

图像特征的选择对视频拷贝的检测性能具有至关重要的作用，而且针对不同的拷贝类型，其检测效果也不同。因此，本文对图像特征与视频拷贝类型之间的关系作了大量的研究和实验，总结已有的基于内容的视频拷贝检测方法，大致可以将它们分为 2 类：(1)基于全局描述子的方法(CD_GLF)。文献[1]对不同的距离测度方法和视频序列匹配方法进行了比较，对基于运动方向的特征用卷积度量，对于序数亮度特征使用 L1 距离，直方图相交法应用于颜色直方图。比较结果表明序数特征具有更优的效果。(2)基于局部描述子(CD_LKF)。点、线和形状等局部描述子在图像和视频检索中具有重要的地位，尤其是基于点的局部描述子有广泛的应用。文献[2]利用时空兴趣点分类人的动作并对周期性运动进行检测。文献[3]利用点特征进行重复图片检测和子图检索。基于全局描述子的方法主要使用全局的时空底层特征，如颜色直方图、颜色布局描述子、序数亮度特征(OIS)[1]。其优点是计算简单，能够应对视频在编码、帧分辨率、旋转、缩放等方面的变化。缺点是对于平移等几何(geometric)变化及复杂的编辑等后期处理的视频检测效果不佳。基于局部描述子的方法首先在视频序列上检测局部的时空特征点(特征点也叫兴趣点或者关键点)，然后用特征点周围的信息对特征点进行描述，这一类方法主要有 Harris 角点检测、SIFT 描述子[4]和 PCA-Sift 特征等。文献[5]对主要的局部描述子进行了比较研究，评价结果表明 SIFT 描述子在对象识别方面具有最好的性能。Harris 角点检测和 SIFT 描述子都具有旋转不变性，但 SIFT 描述子具有更好的尺度和亮度不变性，同时对仿射形变、视角改变和噪声等也有一定的鲁棒性。相比于全局描述子方法，局部描述子对光度变化、几何变换、复杂的编辑和后期处理等有较好的效果。图 1 对全局描述子和局部描述子检测方法作了简单对比。

作者简介：刘 红(1975－)，男，工程师、博士研究生，主研方向：图像与视频处理，多媒体信息检索；文朝晖，硕士研究生；王 晔，工程师、博士研究生

收稿日期：2009-09-10 **E-mail：**liuhong2007@fudan.edu.cn

(a)OIS 特征 1　　(b)OIS 特征 2

(c)SIFT 特征匹配结果

图 1　全局描述子和局部描述子对比

图 1(a)表示使用序数亮度特征 OIS(3×3)得到的特征向量(6,7,1,3,5,2,4,8,9)；图 1(b)得到的 OIS 特征向量为(1,4,7,2,9,6,3,8,5)；图 1(c)代表使用 SIFT 特征得到的匹配结果。虽然序数亮度特征 OIS 对亮度变化和小的插入模式有较好的稳定性，但显然全局描述子对一些大的转换操作无能为力，例如对图中的“画中画”转换情形，使用 OIS 特征，图 1(a)和图 1(b)的 L1 距离是 28，远远大于相似判断的阈值，而使用 SIFT 特征得到了很好的匹配结果。因此，局部描述子在描述力和区分度方面都明显优于全局描述子，但在匹配时间消耗方面，全局描述子远远小于局部描述子。图中使用 OIS 匹配的时间消耗小于 1 ms，而 SIFT 匹配的时间消耗约 1 s。

3　基于图的视频拷贝检测方法

在对图像特征与视频拷贝类型之间的关系作重点研究的基础上，本节的目标是设计一个高效的基于视觉信息(即只考虑视频中的视觉信息，不考虑音频信息)的视频拷贝检测系统，对这个问题的研究涉及系统框架设计、视频序列表示和子序列匹配算法等多个方面，其中 2 项非常关键的任务是视频序列的表示和视频子序列的匹配。

3.1　视频序列的表示

连续视频帧的视觉信息通常在时间上有很大的冗余性，因此，视频序列的匹配没有必要在整个视频帧上进行，一种有效的处理方法是通过从视频中抽取一定关键帧来代表视频，匹配时只需要匹配关键帧序列即可，这样能够减少匹配过程中的计算代价。本文的视频帧提取方法如下：(1)提取视频帧，采取的策略是每隔一定时间提取视频帧，经过实验，选择每隔 0.2 s 提取一帧视频。由于通常情况下视频帧有很大的冗余，因此必须对视频帧进行聚类，这里的视频帧聚类与通常所说的镜头分割有一定的区别，对视频帧进行聚类的主要目的是消除冗余帧，减少计算量，提高检测速度。此外，对视频帧进行聚类还带来了另外一个好处，即有利于检测帧率不同的拷贝视频。视频帧聚类方法采用自动双阈值方法。聚类后，每个类取离平均帧最近的实际帧代表该类，并为其分配一个 ID 以方便后期的视频序列匹配，见图 2。(2)通过聚类方法对视频提取代表帧后的工作是提取代表帧的图像特征，通过这种处理方法，最终可以用一组比较稀疏的高维特征向量表示视频序列。根据对比实验，选择 OIS 特征对除“画中画”外的拷贝类型进行检测，选择 SIFT 特征对“画中画”拷贝类型进行检测。

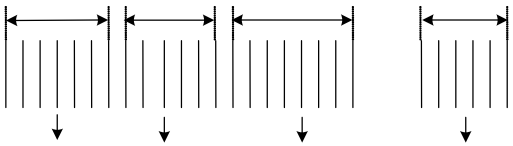
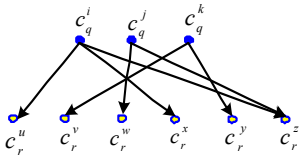


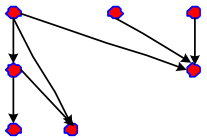
图 2　视频帧聚类和代表帧提取

3.2　视频子序列的匹配

视频子序列的匹配是整个视频拷贝检测过程中的一项核心工作。在本文的方法中，匹配时采取帧帧匹配方式，即用代表帧间的相似度代表相应查询聚类和参考聚类之间的相似度，特征空间距离采用 L1 距离，匹配结果按阈值返回(如图 3(a)所示)，图中 c_q^i 和 c_r^u 间的连线表示第 i 个查询聚类与第 u 个参考聚类之间的相似度大于阈值，这样可以用<i,u>代表匹配结果图中的一个节点。匹配结果中每个查询和参考类对应唯一的 ID，并且记录该类所代表视频对应的起始时刻。



(a)视频子序列匹配结果



(b)匹配结果图和最长路径查找

图 3　基于图的视频序列匹配

为了准确定位拷贝视频在被拷贝视频中的位置，先将匹配结果转换成匹配结果图，进而将该问题转换成在匹配结果图中查找最长路径的问题。为了生成匹配结果图，首先将匹配结果按查询类 ID 排序，然后将排序后的匹配结果按一定的条件生成匹配结果图。其生成方法和条件如下：

利用图 3(a)中的匹配结果，可以得到匹配结果图中的节点，图中节点的方向由代表帧所处的时刻来决定，即在匹配结果图中，查询视频的时间方向与参考视频的时间方向必须一致，这样规定是因为视频序列是一个时间序列，拷贝视频和被拷贝视频的时间序列方向肯定是一致的。而图中 2 个节点之间是否可达则是通过条件控制实现的，经过实验发现，如果采用的条件是：(1)匹配结果图中相邻节点间的时刻相差不大于 20 s；(2)匹配结果图中相邻节点之间的类 ID 相差不大于 5，两者满足一个条件即可。上述条件是通过实验得出的，即在 2 段视频序列的匹配结果中，如果前一个匹配结果与随后的一个匹配结果相差 20 s，则认为这 2 个匹配结果是不相关的。同理，如果相邻 2 个匹配结果跳跃的聚类个数大于 5，也认为这 2 个匹配结果是不相关的。根据上述方法和条件就得到了匹配结果图(如图 3(b)所示)，它是一个有向无环图。在图中，以每个查询类为起点查找其最长路径。查找最长路径有很多经典算法，如 Dijkstra, Bell man_Ford, Floyd_warshall。本文使用了 Floyd_warshall 算法。匹配结果图中通常会存在多条不同起点的最长路径，合并时间层叠的

路径，最终，从匹配图中会得到一些离散的不同起点的最长路径，这些路径就是具有拷贝嫌疑的匹配结果，到底哪些被判定为拷贝还要计算各条路径的相似度，相似度大于某个阈值就判定为拷贝。2 段匹配视频相似度的计算过程如下：

按照图 3，假设匹配结果中查询视频 q 有 m 个聚类，参考视频 r 有 n 个聚类。其中，$1\leqslant i<j<k<=m$, $1\leqslant u<v<w<x<y<z\leqslant n$。在图 3(a)中，每条有向边表示查询视频中的某个聚类与参考视频中的某个聚类之间的相似度 $\mathrm{sim}(c_q^m, c_r^n)>T$ (T 为阈值)。而在图 3(b)中，路径 $[<i,u>\rightarrow <j,w>\rightarrow <k,y>]$代表了以查询类 i 为起点的一条最长路径，对每一个查询类起点查找其最长路径，每条最长路径实际上代表了 2 段匹配的视频序列，其相似度计算公式定义如下：

$$\mathrm{sim}(q,r)=\frac{\sum_{0}^{m}\mathrm{sim}(c_q^i,c_r^j)}{m}\mathrm{lb}(1+m)\ (m\text{ 为路径长度})$$

4 实验及结果

实验环境：英特尔双核 CPU 2.53 GHz, 1 GB DDR2 SDRAM，算法实现使用 C++。

数据集：数据集使用 TRECVID2008[6]提供的官方数据，参考视频文件共 438 个，约 200 h，200 GB 视频数据；查询视频文件共 2 010 个，查询视频是从参考视频或者参考视频以外的视频中抽取一段，并实施一种或者多种转换后生成的，因此，查询视频长短不一，在几十秒和 1 小时之间，约 40 GB。

评价指标：评价使用 3 个指标：最小检测消耗率，拷贝定位精确率和平均拷贝检测时间。

最小检测消耗率：最小检测消耗率实际上利用了检索系统中最常用的 2 个指标检全率和检准率，本文使用它们的反向指标漏检率和误检率，分别用 P_{Miss} 和 R_{FA} 表示。最小检测消耗率的计算公式如下：

$NDCR = P_{\mathrm{Miss}} + \beta\times R_{\mathrm{FA}}$($\beta$ 为损耗系数)

拷贝定位精确度：用于衡量判断为拷贝的时间段与实际视频的时间段之间的配准程度，它是建立在正确的拷贝检测基础之上的，即必须先检测到正确的拷贝，然后才能在此基础上判断拷贝的时间定位精确度，拷贝定位精确度使用 F1 度量。

平均拷贝检测时间：由一个查询遍历完所有参考视频所消耗的时间来衡量。

实验对 TRECVID2008 定义的 10 种拷贝类型[6]进行了评价，实验结果如图 4~图 6 所示，结果表明基于局部描述子的检测方法(CD_LKF)明显优于基于全局底层特征的检测方法(CD_GLF)，特别是 CD_GLF 对“画中画”或含有“画中画”的拷贝类型(对应图中拷贝类型 2 和 9)基本上没有检测效果，而 CD_LKF 对实验中的 10 种类型都有很好的效果。

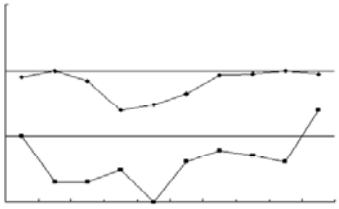


图 4　最小检测消耗时间

同时发现，基于 SIFT 特征的 CD_LKF 方法虽然对各种拷贝类型都有较好的效果，但由于它是基于局部点特征，其计算量相比全局的 OIS 特征至少多了 2 个数量级，检测时间也因此大大提高，因此要将其应用于大规模的视频数据库检测还有一定的难度，这也是下一步重点要解决和研究的工作。

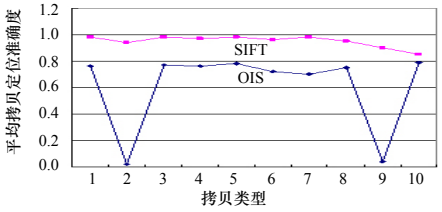


图 5　拷贝定位精确度

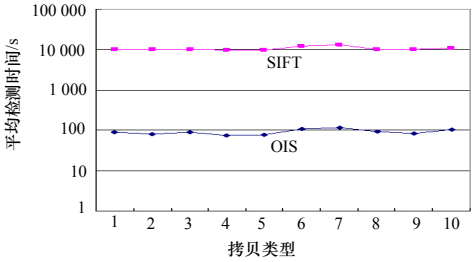


图 6　平均拷贝检测时间

5 结束语

本文重点阐述图像特征对各种视频拷贝类型检测性能的影响，提出了基于图的视频拷贝检测方法，先将匹配结果转换为匹配结果图，进而将视频拷贝检测转换成一个在匹配结果图中查找最长路径的问题。该方法有 2 个明显的优势，首先它弥补了全局底层特征描述力不足的缺陷(对于全局的底层特征，即使相似度为 1 的 2 帧图像也可能是完全不同的 2 幅图)，因为在相似度的计算公式中考虑了路径长度。其次，它充分考虑和利用了视频序列的时空特性，有很好的时间配准性。实验表明，本文方法能够应对多种视频拷贝类型，包括视频的有损压缩、大小和帧率的变化、全局的亮度、颜色、gamma 值的变化等问题。但该方法出有一些不足，虽然实验结果显示基于 SIFT 的检测方法有很好的效果，但由于其检测消耗的时间太长，实用效果不是很理想，因此在下一步的研究中，将改进基于局部描述子(如 SIFT 特征)的视频拷贝检测方法和性能，以提高整个视频拷贝检测系统的性能。

参考文献

[1] Hampapur A, Bolle R. Comparison of Distance Measures for Video Copy Detection[C]//Proc. of IEEE International Conference on Multimedia and Expo. Tokyo, Japan: IEEE Press, 2001.

[2] Yang Ke, Sukthankar R, Huston L. Efficient Near-duplicate Detection and Sub-image Retrieval[C]//Proc. of ACM International Conference on Multimedia. New York, USA: ACM Press, 2004.

[3] Lowe D G. Distinctive Image Features from Scale-invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91.

[4] Laptev I, Lindeberg T. Space-time Interest Points[C//Proc. of International Conference on Computer Vision. [S. l.]: IEEE Press, 2003: 432-439.

[5] Mikolajczyk K, Schmid C. A Performance Evaluation of Local Descriptors[J]. IEEE Trans. on Pat. Analysis and Machine Intelligence, 2005, 27(10): 1615-1630.

[6] TREC Video Retrieval Evaluation Home Page[Z]. (2008-10-23). http://www-nlpir.nist.gov/projects/trecvid/.

编辑　张正兴