

面向压缩域视频拷贝检测的主要边缘相对相位算法

张志杰, 邹建华

(西安交通大学系统工程研究所, 710049, 西安)

摘要: 为了提高视频拷贝检测的查全率和速度,提出了一种基于主要边缘相对相位的压缩域视频拷贝检测算法.该算法先通过压缩视频 I 帧中 DCT 系数的 AC 分量获取边缘信息,然后从边缘信息中抽取主要边缘相对相位,并将其作为视频特征,最后利用这种特征进行视频拷贝检测.各种拷贝攻击对主要边缘相对相位影响很小,因此本文算法是鲁棒的.此外,该算法直接从压缩域获取信息,避免了解压环节,从而减少了检测用时.实验结果表明,该算法相比于其他压缩域算法,查全率提高了 68%,相比于非压缩域算法,检测速度提高了约 5 倍.

关键词: 视频拷贝检测;压缩域;DCT 系数;边缘

中图分类号: TP391 **文献标志码:** A **文章编号:** 0253-987X(2010)10-0008-05

A Relative Phases Algorithm of Key Edges for Detecting Compressed Video Copies

ZHANG Zhijie, ZOU Jianhua

(Systems Engineering Institute, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A new algorithm to detect compressed video copies is proposed to improve the recall and speed of video copy detection based on the relative phases of key edges. The algorithm firstly computes edge information from the AC components of the DCT coefficients, which are directly obtained from all the I frames in compressed videos; The relative phases of key edges are computed from the edge information, and then used as the video signatures; Finally the video signatures are used to detect the video copies. Since various video copy attacks have little influence on the relative phases of key edges, the proposed algorithm is robust. Moreover, the algorithm avoids the decompression process so that the detection time is reduced. Experimental results show that the proposed algorithm outperforms not only the other compressed-domain algorithms 68% in recall, but also the uncompressed-domain algorithm about 5 times in speed.

Keywords: video copy detection; compressed-domain; DCT coefficient; edge

作为管理视频数据库和保护知识产权的技术,视频拷贝检测逐渐成为研究热点.在压缩视频中,压缩域 DCT(离散余弦变换)系数与各种空间特征(颜色、纹理、边缘等)关联密切.空间特征有不同的抽取速度和鲁棒性^[1],性能比较表明颜色、边缘特征较优.颜色特征与 DCT 系数的 DC 分量相关,已被用于拷贝检测^[2]或镜头分割^[3];边缘特征则与 DCT 系数的 AC 分量相关,已被用于镜头分割^[4-5].但是,目前还没有研究者将边缘特征用于视频拷贝检测,主要原因是边缘特征对视频拷贝攻击的鲁棒性较差.我们认为,与颜色特征相比,边缘特征与视频使用价值结合得更为紧密,视频在颜色发生较大变化时仍可传递信息,但是当边缘发生较大变化时视频会变得难以辨认,也就失去了传递信息的功能.因此,本文选择边缘来表征视频,通过设计一种鲁棒的边缘特征达到检测视频拷贝的目的.

1 检测算法

本文提出的检测算法首先利用 DCT 系数与边缘的关系,从视频压缩域获取边缘信息,然后基于边缘信息构建的视频特征,最后使用所构建的特征检测视频拷贝。

1.1 DCT

在众多视频压缩标准中,MPEG-X 和 H. 26X 的影响最大,其基础皆为 DCT. DCT 的定义为

$$F(u,v) = 1/4C(u)C(v)\left\{\sum_{i=0}^7\sum_{j=0}^7f(i,j)\cos[(2i+1)u\pi/16]\cos[(2j+1)v\pi/16]\right\} \quad (1)$$

式中:当 $u,v=0$ 时, $C(u)=C(v)=2^{-1/2}$,其他情况时 $C(u)=C(v)=1$; $f(i,j)$ 表示 DCT 变换前 8×8 像素块 (i,j) 坐标处的亮度; $F(u,v)$ 表示 DCT 变换后 8×8 系数块位于 (u,v) 坐标处的值. DCT 如图 1 所示. 在系数块(量化后,以下同)中, $F(0,0)$ 为直流(DC)分量,反映像素平均能量,其余 63 个系数为交流(AC)分量,反映像素间差异的频率和方向信息。

在压缩视频中,I 帧、P 帧和 B 帧的生成都基于 DCT. 其中 I 帧最为关键,是生成 P 帧和 B 帧的参考帧,所以 I 帧包含了视频基本信息. 本文算法只用到 I 帧。

1.2 AC 分量与边缘的关系

AC 分量与边缘密切相关. 描述边缘的 3 个量为边缘方向 θ 、边缘强度 h 和偏移量 d ,它们都可由 AC 分量计算而得^[6-7]. 边缘模型如图 2 所示. 下面介绍 θ 和 h 的计算过程。

(1)边缘方向 θ 的计算公式为

$$\theta = \arctan\left[\left(\sum_{i=1}^7 F(0,i)\right)/\left(\sum_{j=1}^7 F(j,0)\right)\right] \quad (2)$$

(2)边缘强度 h 边缘模型共有两种,如图 3 所示. 图 3a 的边缘强度可由式(3)计算

$$h = 2^{3/2}(\pi/8)^2 \{ |F(1,0)F(0,1)| / (| [F(1,0) - F(2,0)][F(0,1) - F(0,2)] |)^{1/2} \} \quad (3)$$

图 3b 的边缘强度可由式(4)计算

$$h = 2^{3/2}(\pi/8)^2 |F(0,1)| \times$$

表 1 AC 分量与边缘模型的关系

$F(1,0)\neq 0$ 且 $F(0,1)\neq 0$		$F(1,0)=0$	$F(0,1)=0$
$F(2,0)=0$ 或 $F(0,2)=0$	其他情况		
边缘模型二	边缘模型一	边缘模型二	边缘模型二

$$\left(\left|\frac{2^{1/2}F(1,0)F(1,1)}{2^{1/2}F(1,0)F(1,1)-F(2,0)F(0,1)}\right|\right)^{1/2} \quad (4)$$

边缘模型由图 2 所示像素块对应的 8×8 DCT 块中的 AC 分量之间的关系确定,如表 1 所示。

1.3 基于边缘的视频特征构建

由 1.2 节所述方法得到的 θ 和 h 可直接用作视频特征,但是当视频经过某些处理后, θ 和 h 往往发生较大变化,因此直接将 θ 和 h 作为特征并不合适. 以图 4 为例,图 4a 是原始视频的第 32 个 I 帧,图 4b 是原始视频经过旋转和模糊处理后相同位置的 I 帧,图 4c、图 4d 分别是图 4a、图 4b 的边缘图. 比较图 4c 和图 4d,容易看出在视频经过处理后, θ 和 h 都发生了变化。

为此,我们在 θ 和 h 的基础上设计出一种鲁棒的视频特征,具体过程如下。

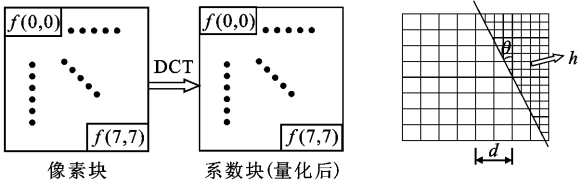


图 1 DCT

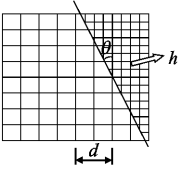
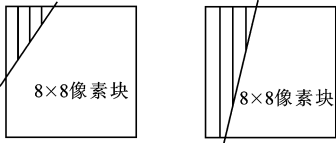


图 2 边缘模型



(a)边缘模型一 (b)边缘模型二

图 3 两种不同的边缘模型

(1)对视频的每一个 I 帧,计算其每一个 8×8 小像素块的 θ (四舍五入后取整数值)和 h . 例如,若 I 帧分辨率为 $M\times N$,则每个 I 帧含 $(M/8)\times(N/8)$ 个 8×8 小像素块,将第 1 个小像素块的 θ 和 h 记作 (θ_1,h_1) ,第 2 个小像素块的 θ 和 h 记作 (θ_2,h_2) ,依次类推。

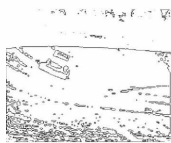
(2)将 $(\theta_1,h_1)\sim(\theta_{(M/8)\times(N/8)},h_{(M/8)\times(N/8)})$ 中有相同 θ 值的二元组合并. 例如,若 (θ_a,h_a) 、 (θ_b,h_b) 和 (θ_c,h_c) 有 $\theta_a=\theta_b=\theta_c$,则此 3 个二元组合并为 1 个新二元组 $(\theta_a,h_a+h_b+h_c)$ 。



(a)原始视频的第32个I帧



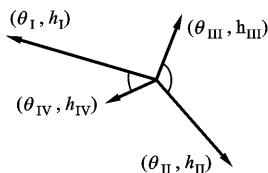
(b)经处理后视频的第32个I帧



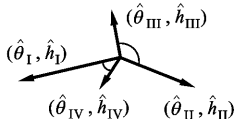
(c)图4a的边缘图



(d)图(4b)的边缘图



(e)图4a的主要边缘信息



(f)图4b的主要边缘信息

图4 原始视频和处理后视频的对比

(3)将新得到的一系列二元组按新 h 从大到小排序,并将新 h 最大的若干二元组取出,所得即为原 I 帧的主要边缘信息.如图4所示,图4e、图4f分别是图4a、图4b的主要边缘信息.图4e的主要边缘按新 h 从大到小依次是 (θ_1, h_1) 、 (θ_{II}, h_{II}) 、 (θ_{III}, h_{III}) 和 (θ_{IV}, h_{IV}) ,图4f的主要边缘依次是 $(\hat{\theta}_1, \hat{h}_1)$ 、 $(\hat{\theta}_{II}, \hat{h}_{II})$ 、 $(\hat{\theta}_{III}, \hat{h}_{III})$ 和 $(\hat{\theta}_{IV}, \hat{h}_{IV})$.比较图4e、图4f,主要边缘的绝对强度和绝对方向都发生了变化,但主要边缘的相对相位没有变化,例如, $\theta_1 - \theta_{IV} = \hat{\theta}_1 - \hat{\theta}_{IV}$.

(4)将主要边缘相对相位信息作为原 I 帧的特征.例如,将 $(\theta_1 - \theta_{II}, \theta_{II} - \theta_{III}, \theta_{III} - \theta_{IV})$ 作为图4a的特征,将 $(\hat{\theta}_1 - \hat{\theta}_{II}, \hat{\theta}_{II} - \hat{\theta}_{III}, \hat{\theta}_{III} - \hat{\theta}_{IV})$ 作为图4b的特征.

(5)若图4a所示帧所属的视频有 K 个 I 帧,则此视频的特征可记作 $\{(\theta_1 - \theta_{II}, \theta_{II} - \theta_{III}, \theta_{III} - \theta_{IV})^1, (\theta_1 - \theta_{II}, \theta_{II} - \theta_{III}, \theta_{III} - \theta_{IV})^2, \dots, (\theta_1 - \theta_{II}, \theta_{II} - \theta_{III}, \theta_{III} - \theta_{IV})^K\}$,其中, $(\theta_1 - \theta_{II}, \theta_{II} - \theta_{III}, \theta_{III} - \theta_{IV})^1$ 为第1个 I 帧的特征, $(\theta_1 - \theta_{II}, \theta_{II} - \theta_{III}, \theta_{III} - \theta_{IV})^2$ 为第2个 I 帧的特征,依次类推.图4b所示帧所属视频的特征表示方法同此.

1.4 视频特征匹配

各含有 K 个 I 帧的视频 V_1 、 V_2 进行拷贝检测

时,首先依照1.3节构建 V_1 、 V_2 的特征,然后使用式(5)进行相似度计算

$$S = 1 - \left(\sum_{k=1}^K |V_1^k - V_2^k| \right) / (\pi \times K) \quad (5)$$

式中: V_1^k 、 V_2^k 分别为视频 V_1 、 V_2 第 k 个 I 帧的特征.当 S 大于阈值 ϵ 时,拷贝存在.

2 实验结果与分析

本实验所用视频取自 TRECVID2008 年拷贝检测评测数据集^[8],共有 37 个视频片段,其中查询视频 1 个,目标视频 36 个.

视频的压缩格式为 MPEG-1,分辨率为 352×288 像素,帧率为 25 帧/s.查询视频时长 28 s,直接取自文献^[8],未经任何处理.目标视频包含 16 个拷贝视频和 20 个非拷贝视频,其中拷贝视频由查询视频施加视频攻击后得到,施加的攻击包括改变颜色、改变尺寸等,非拷贝视频则是由文献^[8]中无关视频截取得到.查询视频和部分目标视频如图5所示.

在视频特征构建环节,只取 2 个新 h 最大的二元组;在特征匹配环节,取阈值 $\epsilon = 0.75$.

2.1 查准率与查全率比较

查准率 p 和查全率 r 是信息检索领域两个重要的评价标准,定义为

$$p = A/B; \quad r = A/C \quad (6)$$

式中: A 是正确检出的视频拷贝个数; B 是检出的视频拷贝总数; C 是真正的视频拷贝个数.图6给出了文献^[2,9]算法与本文算法的查准率比较.文献^[9]算法是非压缩域算法,利用颜色直方图构建特征;文献^[2]算法与本文算法同属压缩域算法,但文献^[2]算法是基于 DC 分量来构建特征的.图6表明,3 种算法的查准率都比较高,但本文和文献^[2]的算法有着更高的查准率.图7给出了文献^[2,9]与本文算法的查全率比较,可见本文算法取得了最高的查全率;文献^[2]算法所用特征的鲁棒性较差,在某些拷贝攻击发生后,特征发生较大改变,因此它的查全率最低;文献^[9]算法的查全率高于文献^[2]算法,但低于本文算法.

2.2 检测速度比较

图8给出了文献^[2,9]算法与本文算法的检测用时,所用计算机配置为 Intel 酷睿 2 四核 CPU, 4 GB 内存,编程语言为 MATLAB.检测用时包括特征构建用时和特征匹配用时.图8表明,本文算法的检测速度比文献^[9]算法提高了约 5 倍,但只比文献^[2]算法略高.原因是本文与文献^[2]算法同属压缩



图 5 查询视频和部分目标视频

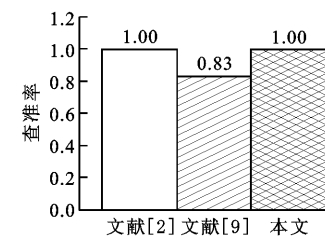


图 6 查准率

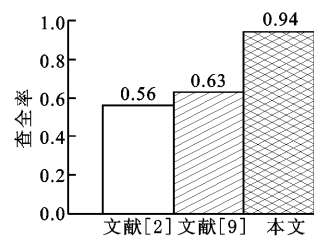


图 7 查全率比较

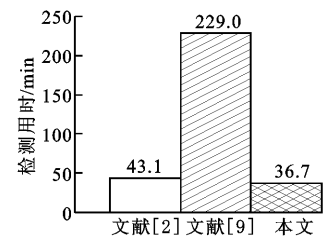


图 8 检测用时比较

域算法,没有解压环节,节省了时间,而文献[9]为非压缩域算法,解压环节耗费了大量时间,因此降低了检测速度.

3 结 论

本文提出了一种基于主要边缘相对相位的压缩域视频拷贝检测算法,首先由压缩域 DCT 系数的 AC 分量获取边缘信息,然后抽取主要边缘相对相位作为视频特征,最后基于这种特征进行视频拷贝检测.由于主要边缘相对相位在视频处理后几乎不发生改变,所以将其作为视频特征可有效抵御各种拷贝攻击.实验结果表明,与同类其人压缩域算法相比,本文算法有更高的查全率,与非压缩域算法相比,本文算法有更快的检测速度.

参考文献:

[1] WANG H, DIVAKARAN A, VETRO A, et al. Survey of compressed-domain features used in audio-visual indexing and analysis [J]. Journal of Visual Communication and Image Representation, 2003, 4(2): 150-183.

[2] 张勇东, 张冬明, 郭俊波, 等. 压缩域快速视频拷贝检测算法[J]. 通信学报, 2009, 30(3): 135-140.

ZHANG Yongdong, ZHANG Dongming, GUO Junbo, et al. Rapid video copy detection on compressed domain [J]. Journal on communications, 2009, 30(3): 135-140.

[3] WON C S, PARK D K, NA I Y, et al. Efficient color feature extraction in compressed video [C]// Proceedings

of SPIE Conference on Storage and Retrieval for Image and Video Data- bases VII. San Jose, CA, USA: SPIE, 1999: 677-686.

[4] ZABIH R, MILLER J, MAI K. A feature-based algorithm for detecting and classifying scene breaks [J]. ACM Multimedia System, 1995: 189-200.

[5] SONG B C, RA J B. Fast edge map extraction from MPEG compressed video data for video parsing [C]// Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases VII. San Jose, CA, USA: SPIE, 1999: 710-721.

[6] SHEN B, SETHI I K. Direct feature extraction from compressed images [C]// Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Data-bases IV. San Jose, CA, USA: SPIE, 1996: 404-414.

[7] LEE S W, KIM Y M, CHOI S W. Fast scene change detection using direct feature extraction from MPEG compressed videos [J]. IEEE Transactions on Multimedia, 2000, 2(4): 240-254.

[8] IMEDIA Research Group. Guidelines for the TRECVID 2008 CD task evaluation [EB/OL]. [2010-03-01]. <http://www-nlpir.nist.gov/projects/tv2008/tv2008.html>.

[9] TAN Y P, KULKARNI S R, RAMADGE P J. A framework for measuring video similarity and its application to video query by example [C]// Proceedings of 1999 International Conference on Image Processing. Piscataway, NJ, USA: IEEE, 1999: 106-110.

(编辑 武红江)

【优秀博士论文摘登】

基于语义的视频浏览系统中的关键技术的研究

作者: 钱学明 导师: 刘贵忠

单位: 西安交通大学电子与信息工程学院

如何快捷地从网络视频媒体库中定位到自己真正感兴趣的内容是视频分析和检索领域中一个富有挑战性的研究课题。为了解决该问题,通常需要对视频进行语义的分类,以充分挖掘其中语义对象、概念以及他们之间的内在因果关系,给用户提供类似于 Internet 上基于关键词的文本查询方式。另外,对视频媒体内容按照一个灵活的方式进行展示,以辅助用户进行内容的查询和浏览,就象翻书那样方便地从视频媒体库中找到自己感兴趣的内容。本文是针对基于语义的视频浏览系统中的关键技术研究而展开的,其中的工作包括如下几个方面。

(1)基于压缩域信息的特征分析和特征提取。由于视频数据量大,而且视频数据不同于图像和文字等媒体内容,其最大的特点是存在极大的时间和空间上的冗余,因此视频媒体在存储和传输前都要进行冗余的压缩编码。如何有效利用压缩域信息进行快速高效的特征分析和特征提取,是我们研究的主要内容之一。使用 I 帧 DCT 系数来近似表示图像的纹理,用 DC 图像及其直方图来近似表达原始图像及其直方图以快速进行分析,用压缩域中的运动矢量场来进行快速的运动特征描述等。这些为本文后续工作中采用压缩域信息的特征表示以及快速的特征分析和提取方法奠定了基础。

(2)镜头边界检测以及基于语义的镜头分类。镜头边界检测即场景切换类型检测,是视频分析中的一个基本环节。本文从镜头切换的数学模型出发推导了 Flashlight 和 Fade in/out 的累计直方图差的一般特性,并用压缩域的 DC 图像的直方图来进行快速的检测。从累计直方图的特性,不仅可以检测出 Fade in/out 并且能够进行 Fade in/out 所对应的语义类型识别。在对 Flashlight 和 Fade in/out 确认的基础上进行 Cut、Dissolve 和其他类型的场景切换检测,极大地提高了镜头边界检测的性能。进行镜头的语义分类是视频检索中的一个重要环节,从语义的镜头类型信息以及相应的音频数据类型信息能够有效地进行视频故事单元的主题思想理解。本文对视频按照其所在领域知识、视频编辑中的约定俗成的制作手法,以及摄像机的运动模式,将足球比赛视频中的镜头划分成一系列语义的类别,并且将音频按照时域和谱域的能量分布特性划分成纯说话、静音、纯音乐、含有背景噪声的说话以及含有背景音乐的说话片段等 5 种类型。这种场景切换类型识别、语义的镜头类型特征以及音频数据类型等都为我们进行后续高级语义事件和故事单元的检测和分类提供了重要的信息。

(3)字幕检测、定位、跟踪、分割以及字幕类型划分方法。利用 MPEG 压缩域中 I 帧 DCT 系数所表达的纹理特征进行快速、高效的字幕检测和定位,并用压缩域中的 DCT 系数特征来对字幕的出现和消失帧予以快速的跟踪,最终融合包括前背景以及视频字幕的时间冗余特性来进行高效的字幕分割。对 H. 264/AVC 和 MPEG 压缩域中的字幕检测性能进行了对比分析,将字幕按照其活动性以及存在时间的长度信息划分成滚动字幕、长期字幕、说话内容字幕和标题字幕等 4 种类型,以辅助进行基于语义的事件和故事单元检测和分类。

(4)全局运动估计、基于摄像机运动模式的镜头细分和基于全局/局部运动相结合的应用。我们使用压缩域中的运动矢量场来进行快速的全局运动估计,其中包括基于运动矢量组的全局运动估计和基于遗传算法的全局运动估计。利用全局运动信息,将足球比赛视频中的全景镜头进行语义的细分,这种细分后的语义镜头类型信息为进行足球视频中的语义事件挖掘提供了重要的参考依据。另外,提出了基于 GM/LM 视频字幕遮挡区域恢复以及视频通信系统中的错误恢复,达到了较好的恢复效果。

(5)语义事件和故事单元的检测和分类。融合语义的镜头类型信息、字幕类型信息、摄像机运动模式、视频领域相关知识来将体育比赛视频序列划分成进球、射门、犯规、定位球和普通等 5 种完备的事件集合,并且在此基础上对精彩事件按照摄像机运动模式分类到比赛中的两支球队中。融合视觉信息、字幕类型信息和音频类型特征来进行新闻视频故事单元的检测和分类。这种事件和故事单元的分类方式为进行基于高级语义特征的视频检索、摘要和浏览提供了可能。

(6)一个统一的灵活方便的视频摘要和浏览系统框架。在事件和故事单元的检测和分类基础上,按照书目编排的目录即 ToC 结构来有效进行新闻和体育视频内容组织,并提出了一种通用的基于 ToC 的视频内容浏览系统框架。在该框架中,按照事件和故事单元的分类情况,给出了一种可分级的视频摘要和浏览方案。不同级所生成的摘要能够很好地提供对视频内容浏览的形式,使用户可以像浏览书本那样方便快捷地进行视频内容浏览,并对感兴趣的内容进行快速的定位。