

基于视觉感知的时空联合视频拷贝检测方法

潘雪峰^{1),2)} 李锦涛¹⁾ 张勇东¹⁾ 唐 胜¹⁾ 余乐军^{1),3)} 夏 添^{1),2)}

¹⁾(中国科学院计算技术研究所 北京 100190)

²⁾(中国科学院研究生院 北京 100190)

³⁾(北京师范大学电子信息科学与技术系 北京 100875)

摘 要 视频拷贝检测在视频信息检索和版权保护等方面有着重要意义. 文中依据视觉感知特性, 认为现有的基于单帧的拷贝检测方法未充分利用视频的时空相关性, 进而在利用这一重要特性进行拷贝检测方面进行了探讨. 依据自然图像序列的时空相关性以及视觉系统对时空信息的感知特性, 通过对时空切片单元块 *Sub* 进行 DCT 分析得到与视频内容相关的时空联合特征, 并基于该特征提出视频拷贝检测算法. 实验表明, 文中所提出的视频拷贝检测方法对视频内容变化敏感而对相同内容视频格式上的变化鲁棒, 表现出优于现有方法的性能.

关键词 视频拷贝检测; 视觉感知; 时空相关; 时空切片; DCT

中图法分类号 TP391 DOI号: 10.3724/SP.J.1016.2009.00107

Spatiotemporal Video Copy Detection Based on Visual Perception Analyses

PAN Xue-Feng^{1),2)} LI Jin-Tao¹⁾ ZHANG Yong-Dong¹⁾ TANG Sheng¹⁾ YU Le-Jun^{1),3)} XIA Tian^{1),2)}

¹⁾(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

²⁾(Graduate University of Chinese Academy of Sciences, Beijing 100190)

³⁾(Department of Electronic Information Science and Technology, Beijing Normal University, Beijing 100875)

Abstract A novel spatiotemporal video copy detection method based on visual perception features is proposed in this paper. The method differs from the conventional video copy detection algorithms in that it does not using key frame based features any more. By extracting visual perception based spatiotemporal correlation of slice in frequency domain using DCT, the proposed method possesses both robustness and discriminability for video copy detection. It copes well with not only the format, resolution variation of video but also the aspect ratio change. Even more, the method is robust to display format conversion such as adding pillar-box. The experiment results show the proposed method is effective and robust.

Keywords video copy detection; visual perception; spatiotemporal correlation; spatiotemporal slice; DCT

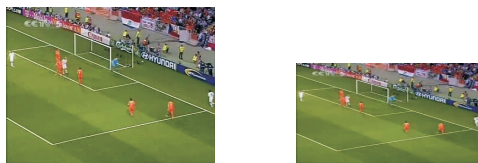
1 引 言

数字化技术极大地方便了视频信息的存储、传

输和复制,使得数字视频内容的数量飞速增长.而在实际应用中受到网络传输带宽、视频质量要求等条件的限制,相同的视频内容经常被编码为不同的格式,以适应不同的需求.我们称这种在内容上与原视

收稿日期:2006-12-01;最终修改稿收到日期:2008-10-19. 本课题得到国家“九七三”重点基础研究发展规划项目基金(2007CB311100)、国家“八六三”高技术研究发展计划项目基金(2007AA01Z416)、国家自然科学基金(60773056)和北京市科技新星计划(2007B071)资助.
潘雪峰,男,1977年生,博士研究生,主要研究方向为视频处理、机器学习. E-mail: xspan@ict.ac.cn. 李锦涛,男,1962年生,研究员,博士生导师,主要研究领域为多媒体技术、虚拟现实技术. 张勇东,男,1973年生,博士,副研究员,主要研究方向为图像处理、视频处理. 唐 胜,男,1972年生,博士,副研究员,主要研究方向为多媒体信息检索、视频处理. 余乐军,男,1977年生,博士,主要研究方向为图像处理、视频编码. 夏 添,男,1980年生,博士研究生,主要研究方向为多媒体信息检索、机器学习.

频相同,而在格式上发生改变的视频为拷贝视频. 通常的拷贝视频,相对于原视频常见的格式转变方式有编码方式转换、画面尺寸变化、画面比例变化、添加边框等等. 图 1 给出其中几种比较常见的例子.



(a) 原始视频帧MPEG 352×288 (b) 格式变化为AVI 320×180

(c) 格式变化为AVI 320×180并加上竖直边框

图 1 视频格式变换

检测拷贝视频在视觉信息处理中有着重要的实际用途:首先视频拷贝检测在视频信息检索和对搜索结果进行过滤、排序方面具有重要意义^[1-3]. 其次,在媒体追踪方面,视频拷贝检测可以用来自动地监控特定内容视频的播出状况^[4]. 此外,在数字视频版权保护方面,相对于传统的数字水印技术而言,由于具有无需在媒体中加入附加信息、特征提取可在媒体发布之后进行等良好的特性,拷贝检测也开始受到关注^[5]. 为此,研究者提出了多种视频拷贝检测方法,但由于视频格式和内容的多样性,如何选取有效的特征,准确地检测视频拷贝依然是一个未解决的问题. 下面介绍视频拷贝检测中已有的相关工作和面临的困难.

为便于介绍,我们先引入如下记号: V ——原视频; V' ——待判定视频. 则当 V 与 V' 内容相同时,称它们相互之间存在拷贝关系,记为 $COPY(V', V)$.

可见,拷贝检测的实质即在于判定两段视频是否具有相同的内容. 现有的视频拷贝检测方法多采用基于帧匹配的框架^[6]:首先选取反映 V 内容的帧序列 I_{key} 以及反映 V' 内容的帧序列 I'_{key} . 然后通过比较 I_{key} 和 I'_{key} 中对应帧的相似性来衡量 I_{key} 和 I'_{key} 的相似程度,并判别视频是否具有相同的内容. 在进行帧序列匹配时,一些图像相似度计算方法被直接推广到了视频拷贝检测中. 例如基于颜色直方图的方法^[7]、基于边缘的方法^[8]以及基于运动信息的方法^[6]等. 但这些方法分别存在如下的问题:采用颜色直方图无法描述图像中各种颜色的空间分布,因此这类方法在进行视频拷贝检测时描述能力比较弱. 同时由于视频编码时会对颜色信息进行量化处理,编码后的视频帧在色彩上会发生偏移,影响检测结果,这是基于颜色直方图的方法所无法避免的问题.

基于边缘的方法,其计算比较复杂,而且由于视频格式转换时,会出现亮度变化和块效应而导致边缘信息产生变化,导致方法失效. 而基于运动信息的方法,采用的宏块运动矢量特征往往无法正确描述实际的运动信息,因此也难以取得理想效果.

近来,研究者又提出基于分块亮度顺序 (Ordinal Measures, OM) 的视频拷贝检测方法^[1,9-10]. 文献^[6]指出基于 OM 的方法相对现有其它方法而言,性能是最优的. 这种方法对视频亮度、色调、尺寸变化都比较鲁棒. OM 概念最初在文献^[11]中提出,文献^[10]中用 OM 来识别体育运动中的特定动作片段,但没有涉及对视频格式变化的处理. 文献^[1]论述了如何用 OM 方法检测经格式转换的视频拷贝. 文献^[9]提出一种结合时序信息的 OM 方法,能检测添加边框的拷贝视频,显示出了优于普通 OM 方法的性能. 该方法为了对添加边框这种格式变化方式鲁棒,使用的 OM 分块数目不能超过 2×2 . 但由于 OM 分块数少,对视频内容的区分能力较弱,该文采用了引入视频序列同一分块位置随时间的变化特征来弥补这一不足. 同时该文也指出,为了继续加强描述能力,还是必须增加 OM 分块数.

综合上述,要检测出经过格式变化后的拷贝视频,检测方法必须满足两个要求,即一方面要对因格式转变而产生的视频数据变化鲁棒,另一方面又要对因视频内容变化产生的视频数据变化敏感. 如何解决这一问题是视频拷贝检测的核心. 因此,提取在鲁棒性和区分性方面都更有效的特征成为了整个检测方法的关键.

本文和已有工作的最大不同在于不再采用基于帧的拷贝检测框架,而是基于视觉感知提出一种视频时空联合特征,并基于该特征进行视频拷贝检测. 本文方法依据自然图像序列的感知特性,通过分析时空切片的频域特征,有效地利用了视频内容的时空相关性. 由于所采用的特征反映的是视频内容变化对视觉感知的刺激,而非是某个时刻视频帧的低级特征,因而对于经格式转换的视频内容有着良好的鲁棒性. 本文方法取横向和纵向两种切片,由于给视频添加字幕框或竖边框只会影响到横、纵切片中的一个,因此对于添加边框是鲁棒的. 在生成视频的灰度切片时,由于统一了视频切片的尺寸,并且是在视频画面的中央位置取切片,故而所选特征对视频的画面尺寸和比例变化鲁棒. 同时所提取的特征体现了视频内容对视觉的刺激,因而能够有效区分不同内容的视频. 实验表明所提出的方法对于视频内容变化敏感而对多种视频格式变化鲁棒,具有优

于现有方法的性能。

本文第 2 节依据视觉感知特性来分析所要采用的工具,即时空切片;据此,第 3 节提出基于感知的视频时空联合特征提取和拷贝检测算法;第 4 节给出实验结果,说明本文方法的有效性;第 5 节总结全文。

2 视觉感知特性与时空切片

如前文所述,要解决视频拷贝检测的核心问题,获取适当的特征是关键。因此在这部分基于视觉感知来分析如何选取视频特征。研究表明视觉敏感性 K 与自然图像序列的能量谱 S 之间有如下的关系 (N 为噪声能量)^[12]:

$$K = \frac{(1/S)^{1/2}}{(1+N/S)^{3/2}} \quad (1)$$

而视频作为由采集设备采集到的自然图像序列而存在,其内容在空间上和时间上是有相关性的。自然图像序列的能量谱 S 与空间频率 f 、时间频率 ω 之间的关系可以表示为^[13]

$$S(f, \omega) \propto \frac{1}{f^{m+1}} G(\omega/f) \quad (2)$$

该公式表明,自然图像序列的能量谱和图像序列的空间频率及时间频率密切相关,而且能量谱的时空相关性是不可分离的^[13],这种时空关联性正是视频区别于静态图像的关键。而现有基于单帧的视频特征选取方法,却未充分利用这一重要特性。基于以上分析,我们认为在自然图像序列中,选取与能量谱相关的时空联合特征能够比较有效地刻画视频特征,反映视觉对视频内容的敏感性。该特征在拷贝检测中的应用值得探讨。

本文中,我们选取视频的时空切片作为分析时空相关性的工具。下面介绍时空切片的概念并揭示其所包含的时空联合信息。

时空切片是指在视频的不同位置上于不同的时间进行采样而得到的图像^[14]。将视频在时间轴上展开为帧序列,然后在各帧相同位置进行采样,即可得到该视频信号在所指定位置上随时间变化的信息。对长度为 L 帧、画面宽为 M 、高为 N 的视频序列 V ,本文直接提取每帧处于画面中间位置的像素线来生成时空切片,提取过程如下:

对视频第 i 帧 $I[i]$,以该帧左上角为原点,右方作为 X 正向,下方作为 Y 正向建立坐标系 $I[i]_{xoy}$,则 $I[i]$ 水平方向像素线 h_i 与竖直方向像素线 v_i 分别由以下像素组合生成:

$$\begin{aligned} h_i &: I[i]_{xoy} \left(x, \frac{N}{2} \right), \quad x=1, 2, \dots, M, \\ v_i &: I[i]_{xoy} \left(\frac{M}{2}, y \right), \quad y=1, 2, \dots, N \end{aligned} \quad (3)$$

然后用提取的各帧像素线拼接出 V 的横向时空切片 S_V^H 和纵向时空切片 S_V^V 。

我们用 $S_V^H = \langle h_1, \dots, h_L \rangle$ 表示由视频 V 的纵向中间位置水平方向像素线 h_i 拼接生成的时空切片 (h_i 为第 i 帧的像素线)。

我们用 $S_V^V = \langle v_1, \dots, v_L \rangle$ 表示由视频 V 的横向中间位置竖直方向像素线 v_i 拼接生成的时空切片 (v_i 为第 i 帧的像素线)。

下面分析时空切片中像素间的相关性。以水平方向的切片 S_V^H 为例,在水平方向时空切片图像中,若以图像的左上角为原点,建立如下坐标系 S_{xoy} (图 2 为我们对一段足球视频按上述方法提取的时空切片及建立的坐标系),则切片 S_V^H 在 S_{xoy} 中 (x, y)



图 2 切片坐标系(04 年欧洲足球锦标赛,半决赛,捷克对荷兰下半场一部分)

位置上的像素值与视频中第 x 帧在 $I[x]_{xoy}$ 中 $\left(y, \frac{N}{2}\right)$ 处的像素值相等,即

$$S_{xoy}(x, y) = I[x]_{xoy} \left(y, \frac{N}{2} \right) \quad (4)$$

在 S_{xoy} 中, S_V^H 在 Y 方向上相邻的像素点是同一帧的像素线上相邻的点,反映的是像素点的空间相关性, S_V^H 在 X 方向上相邻的像素点则是相邻帧中位

于同一位置上的像素点,反映了像素的时间相关性,这表明时空切片中融合了视频的时空相关信息。下面我们基于时空切片所含的时空相关信息来提取特征并进行视频拷贝检测。

3 视频拷贝检测

基于以上对视觉感知和时空切片特点的介绍,

本节给出视频特征提取方法和拷贝检测算法. 首先提取 V 的横向时空切片 S_V^H 和纵向时空切片 S_V^V ; 取 V' 的横向时空切片 $S_{V'}^H$ 和纵向时空切片 $S_{V'}^V$, 基于所提取的切片, 本文拷贝检测方法的两个关键步骤分别为

1. 特征提取. 分别提取 S_V^H 、 S_V^V 、 $S_{V'}^H$ 、 $S_{V'}^V$ 的特征;
2. 拷贝检测. 计算 S_V^H 、 $S_{V'}^H$ 特征的相似度以及 S_V^V 、 $S_{V'}^V$ 特征的相似度, 当两个相似度的较大值超过给定阈值 τ 时, $COPY(V, V')$ 为真, 否则为假.

下面详细介绍特征提取方法和拷贝检测算法.

3.1 特征提取及 Sub 表示

首先基于时空切片提取视频的特征. 将视频时空切片的高度变为某定值(为便于下文变换域分析, 将高度值设定为 2^n), 并转换为灰度图像, 称为灰度切片.

我们用 $S_V^{H_{gray}}$ 表示视频 V 横向灰度切片; $S_V^{V_{gray}}$ 表示视频 V 纵向灰度切片.

设灰度切片宽为 W , 高为 H , 按从左到右的顺序将切片分割成多个 $H \times H$ 的块, 每个块称之为一个切片单元块 Sub (Slice Unit Block).

我们用 Sub 表示灰度切片经分割后生成的与切片等高的正方形图像块;

$SUB = \langle Sub[1], \dots, Sub[W/H] \rangle$ ($[\cdot]$ 为向下取整) 表示灰度切片从左至右分割后生成的 Sub 序列.

对取出的各 Sub 块进行离散余弦变换(DCT). 根据 DCT 变换的性质, 二维 DCT 变换基体现了图像沿坐标轴方向变化的频率^[15], 而 Sub 的坐标方向分别为时间和空间方向, 所以其 DCT 变换基体现了图像序列在时间和空间上的变化频率 ω 和 f . 依据前文分析, f 和 ω 与图像序列的能量谱 S 和视觉敏感性 K 直接相关, 因此通过对 SUB 进行 DCT 变换可获得整个视频序列的时空相关性和视觉感知特征.

实验中, 取不同视频, 对它们的 Sub 进行 DCT 变换, 统计各个 DCT 分量系数的变化. 图 3 中是多个 Sub 经 DCT 变换后各分量系数的方差. 可以看出, DCT 低频系数在不同视频的 Sub 之间有着显著的不同. 本文按 Zig-Zag 顺序, 取一部分低频交流系数来生成各个 Sub 的特征向量, 称为 $SubV$.

这样得到的 $SubV$ 对于视频内容的变化是敏感的, 同时由于生成视频的灰度切片时, 首先对视频的切片统一尺寸, 而且所取的切片位于视频画面的中央位置, 因此 $SubV$ 对视频的画面尺寸和比例变化

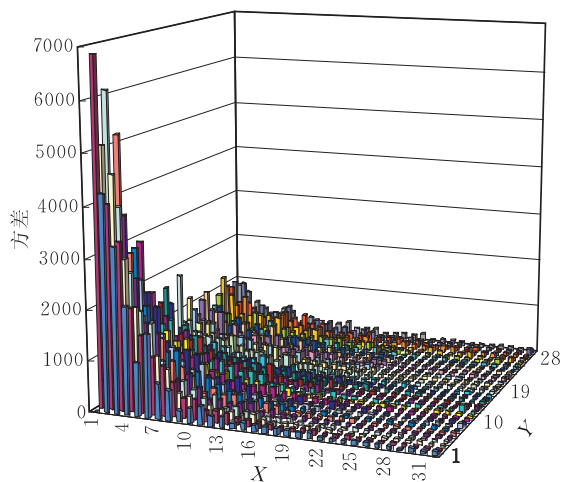


图 3 DCT 系数方差分析

是鲁棒的. 再者, 由于 $SubV$ 反映的是视频内容对视觉感知的刺激, 而非是某个视频帧的低级特征, 因而受格式转变中颜色变化等因素影响较小. 这样各个切片分块的 $SubV$ 所组成的向量组可以有效反映视频的内容而基本不受格式变化影响. 接下来, 基于特征向量组, 我们提出一种新的拷贝检测算法.

3.2 拷贝检测算法

沿用前文定义, 我们引入如下记号: SUB_V 表示视频 V 的 Sub 序列; Sub_V^i 表示 SUB_V 中的第 i 个 Sub 块; $SubV_V^i = \langle SubV_V^i[0], \dots, SubV_V^i[P-1] \rangle$ 表示对 Sub_V^i 进行 DCT 变换后, 按 Zig-Zag 顺序依次取出前 P 个位置的 DCT 交流系数所生成的特征向量; $SUBV_V = \langle SubV_V^0, SubV_V^1, \dots, SubV_V^L \rangle$ 表示由 $SubV_V^i$ 按块顺序组成的特征向量组.

对要判定的视频 V' 和原视频 V , 设它们切片分块数为 L . 本文通过计算 $SUBV_{V'}$ 和 $SUBV_V$ 的相似度来判定 V' 是否为 V 的拷贝视频.

定义 1. 特征向量组相似度 VFVS (Video Feature Vector Similarity). 其计算公式如下:

$$VFVS(SUBV_{V'}, SUBV_V) = 1 - \frac{\sum_{i=1}^L d(SubV_{V'}^i, SubV_V^i)}{\sum_{i=1}^L (abs(SubV_{V'}^i) + abs(SubV_V^i))} \quad (5)$$

其中 $d(\cdot)$ 为 L_1 距离; $abs(\cdot)$ 为向量中各元素的绝对值之和.

以 $SUBV_V^H$ 记视频 V 横向切片的特征向量组; $SUBV_V^V$ 记视频 V 纵向切片的特征向量组, 定义视频 V 和视频 V' 的相似度为

定义 2. 视频 V 和视频 V' 的相似度 $Simi(V', V)$. 其计算公式如下:

$$Simi(V',V)=\max(VFVS(SUBV_{V'}^H,SUBV_V^H),VFVS(SUBV_{V'}^V,SUBV_V^V))$$

(6)

当 $Simi(V',V)$ 超过给定阈值 τ 时,判定视频 V' 是视频 V 的拷贝视频. 判定公式 $COPY\langle V',V\rangle$ 如下:

$$COPY\langle V',V\rangle=\begin{cases} \text{真}, & Simi(V',V)>\tau \\ \text{假}, & \text{否则} \end{cases}$$

(7)

关于阈值 τ 的选取,我们将在实验部分进行讨论.

4 实验和结果分析

针对前文所述的视频拷贝检测的核心问题,我们通过实验验证本文提出的视频特征对格式变化的鲁棒性和对内容变化的敏感性,并给出采用本文方

法进行拷贝检测的结果.

4.1 特征鲁棒性和内容敏感性分析

4.1.1 鲁棒性分析

视频特征的鲁棒性主要体现在对不改变视频内容的视频格式转换鲁棒. 为此,取两段足球视频,两段电影视频进行实验分析. 以不同格式的同一视频内容之间的相似度作为衡量特征鲁棒性的标准. 在实验中所取视频的原始格式为 MPEG1,画面尺寸为 352×288 . 我们将视频进行了 5 种格式的转换,包括了编码方式变化、视频尺寸变化、视频画面比例的变化、视频竖边框的添加这几种情况,具体的格式如下: 352×288 AVI、 320×180 AVI、 512×288 MPEG-1、 320×180 AVI 有竖边框、 320×240 MPEG-1 有竖边框.

按前文所述方法提取这些视频水平和竖直方向的切片,计算它们特征的相似度. 表 1 为每种经格式转换后的视频与原视频之间的相似度均值.

表 1 经格式转换的视频与原视频的相似度均值

转换后的视频格式	纵向 VFVS	横向 VFVS	相似度
352×288 AVI	0.88245	0.88623	0.88623
320×180 AVI	0.87302	0.86946	0.87302
512×288 MPEG-1	0.90397	0.89257	0.90397
320×180 AVI 加竖边框	0.86541	0.41225	0.86541
320×240 MPEG-1 加竖边框	0.87394	0.63628	0.87394

从表 1 中可见采用本文的算法,经格式转换后的视频,即使变化相当显著(如第 4 种格式转换,对视频的编码方式、尺寸、画面比例都进行了变化而且添加了竖边框),但依然和原视频保持了较高的相似度. 这显示出了本文方法对视频格式变化的鲁棒性.

4.1.2 内容敏感性分析

作为一种视频特征,必须对视频内容变化敏感,本文方法在这方面也表现出了良好的性能. 图 4 为 6 段电影、体育、和新闻视频进行横向切片提取后得到的各个特征向量组中各特征向量的绝对值之和.

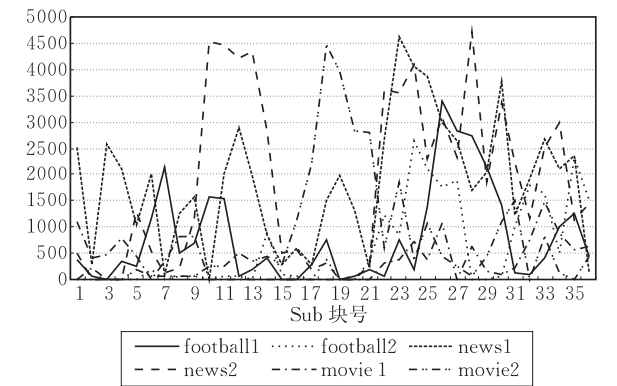


图 4 不同视频内容的特征差异

可以看到,不同内容的视频其特征向量组中各向量的元素绝对值和差异较大.

进而我们将一段约 18000 帧的 MPEG1 视频分为 24 段,每段 800 帧,求得各段的特征向量组. 然后将每一段的特征向量组分别和其它 23 段视频的特征向量组进行比较,依前文方法计算其相似度. 设同一片段与其自身的特征向量组相似度 S_{same} 以及与其它片段特征向量组的相似度 S_{diff} 服从高斯分布,表 2 为其分布特性. 对比表 1,可以看到用本文方法,内容相同格式不同的视频的相似度远高于不同内容视频之间的相似度. 本文方法有效地区分了和内容相同与不同的视频片段,表现出了对内容的敏感性.

表 2 内容相同片段与不同片段的相似度特性

相似度	均值	标准差
S_{same}	1	0
S_{diff}	0.264375	0.064546

此外,为检测本文方法对同类型视频内容的敏感性. 我们对一段约 18000 帧的足球视频提取特征. 取自第 26 个 Sub 块开始,长为 800 帧的片段,分别与视频的各个片段计算相似度,结果如图 5 所示.

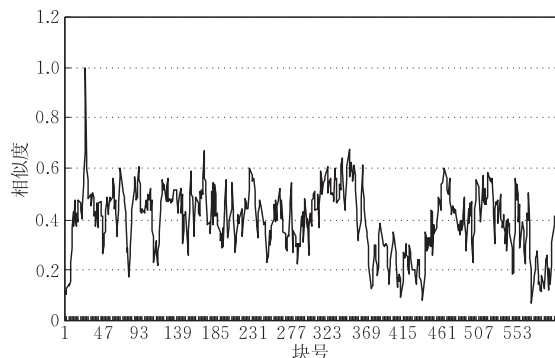


图 5 同类视频内容的特征相似性.

可以看出,本特征对于同类视频依然保持了较好的区分度,并能对长段视频中的短片段进行定位.

4.2 视频拷贝检测实验

在实验中,我们采用了约 50000 帧的 MPEG1 视频作为原始视频,包含新闻、体育、电影等视频内容.将视频分为 63 个等长片段,对这些片段进行如表 1 中所列出的视频格式转换,共得到 315 个转换格式后的拷贝视频.对每个原始视频,各有 5 个拷贝视频.用本文所提出的方法来检测 315 个转换格式后的视频中与各个原视频内容相同的拷贝视频.

实验中,所取的切片分块大小为 32×32 ,选取的 DCT 交流系数个数为 35,阈值 τ 从 0.45 变化到 0.95.我们对不同的阈值 τ 计算了拷贝检测的查准率和查全率.以 C_d 记算法检测得到的视频拷贝个数,以 C_d^r 记正确检出的原视频拷贝的视频个数,以 C^r 记在检测集中原视频实际有的视频拷贝个数.我们按以下公式计算拷贝检测的查准率 P 和查全率 R :

$$P(\tau) = \frac{C_d^r}{C_d}, \quad R(\tau) = \frac{C_d^r}{C^r} \quad (8)$$

图 6 为 τ 取不同值时,算法在数据集上的查准率和查全率.

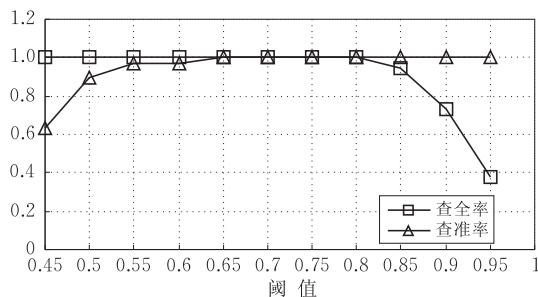


图 6 阈值变化时本文算法的查全率、查准率

从图 6 中可以看到,本文所提出的视频拷贝检测方法在 τ 值位于 0.65 和 0.8 之间时,查准率和查

全率都稳定在 100%.

我们将实验结果与文献[9]的方法进行了对比.对实验数据集中的片段按文献[9]的方法进行拷贝检测,并计算查准率和查全率.按文献[9]中所述,以 0.5 为加权系数所得到的检测效果最好,所以在实验中,取加权系数 $\alpha = 0.5$.并依据文献[9]结果,分别取 0.1, 0.15 和 0.2 作为阈值进行实验.图 7 为按文献[9]方法得到的查准率和查全率与本文部分实验结果的对比,横坐标为所选阈值(由于文献[9]方法阈值取 0.2 时其查准率为 0.26,小于 0.6,所以在图中未显示).

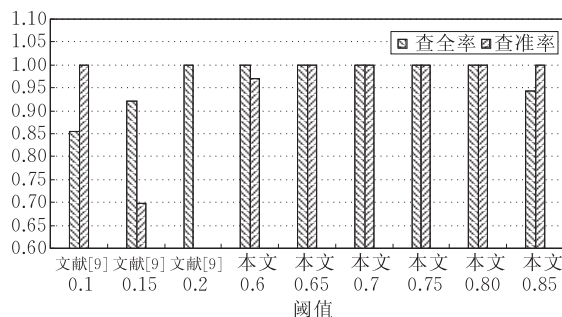


图 7 本文方法与文献[9]方法对比:取不同阈值时两种方法的查全率和查准率

从图 7 中可以看出,文献[9]的方法随着阈值的增大,查全率上升而查准率下降.当阈值为 0.1 时,查准率为 1,查全率为 0.86;当阈值为 0.2 时,查全率为 1,查准率降至 0.26,无法在一个阈值上使查全率和查准率同时达到高水平.其综合性能在阈值取 0.15 时达到最佳,此时的查全率和查准率分别为 0.92 和 0.7.而本文方法,当阈值在 0.65 和 0.8 之间时,拷贝检测的查全率和查准率都能维持在高水平上(在实际应用中即可在这一区间内选取阈值 τ 获得理想效果),这表明本文所选取的特征对于视频格式转化的鲁棒性和对视频内容变化的敏感性都较强,从而能够比较明显地区别原视频的拷贝视频和与原视频内容不同的其它视频.同时也说明,我们基于视觉感知和视频时空相关特征提出的视频拷贝检测方法是有效的.

4.3 检测算法复杂度对比

在计算复杂度方面,以一个长度为 n 帧的原视频 V 和拷贝视频 V' 为例.按文献[9]中的方法,当分块数为 2×2 时,在特征提取完后计算时序变化 D_T , V 内每一帧和相邻帧要作一次比较得到相邻帧差结果 d 需 $4(n-1)$ 次减法, V' 的每一帧和相邻帧也要做一次比较,需 $4(n-1)$ 次减法, V 和 V' 对应的相

邻帧差结果做差运算, 需要 $4(n-1)$ 次减法, 求所有对应帧相邻帧差的和约需 $4(n-1)$ 次加法, 再做一次除法进行归一化. 计算 V 和 V' 对应帧的 OM 的相似度 D_s , 首先计算对应帧对应位置的 OM 差值需 $4n$ 次减法, 再对所有帧的各对应 OM 差值求和需要 $4n$ 次加法, 此外也还需要一次除法进行归一化. 其复杂度大约为 $24n$ 次加法.

按本文方法, 设时空切片图像块的尺寸为 $m \times m$, 选取的 DCT 交流系数个数为 p , 生成的 V 和 V' 特征向量组中的向量个数均为 n/m . 计算各向量对应元素之间的距离所用减法次数为 p , 基于对应元素距离求向量距离所需加法次数为 $p-1$, 则得到向量组所有向量的距离共需 $(n/m) \times (2p-1)$ 次加法. 此外对各向量的距离求和需要 $n/m-1$ 次加法. 因此计算向量组距离共需约 $(n/m) \times 2p$ 次加法. 而计算所有元素的绝对值和所用的加法次数也约为 $(n/m) \times 2p$. 此外还需要一次除法和一次减法, 以获得最后的相似度. 在实验中我们取 $m=32$, $p=35$, 其复杂度约为 $4.375n$ 次加法. 所以在检测算法上, 本文方法的复杂度约为文献[9]方法的 $4.375n/24n=18.23\%$.

5 结 论

本文提出一种基于视觉感知的时空联合视频拷贝检测方法. 该方法与已有方法的不同在于不再采用将视频拷贝检测转化为帧匹配的模式来处理问题. 而是依据视觉感知特性, 通过分析视频时空切片, 提出一种新的时空联合视频特征, 进而提出视频拷贝检测算法. 实验表明本文所使用的特征提取方法能有效地区别原视频经过格式转变而内容未发生变化的拷贝视频和与原视频内容不同的其它视频, 体现出了优于现有检测方法的性能.

在进一步的工作中, 我们将继续研究本文所提出的方法在实际互联网环境下的应用, 一方面关注视频内容在时间上的准确对齐, 另一方面还将关注对不同帧率视频进行有效下采样以克服帧率变化对特征提取的影响.

参 考 文 献

- [1] Hua Xian-Sheng, Chen Xian, Zhang Hong-Jiung. Robust video signature based on ordinal measure//Proceedings of the International Conference on Image Processing. Beijing, China, 2004: 685-688
- [2] Cheung S-S, Zakhor A. Efficient video similarity measurement with video signature. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(1): 59-74
- [3] Cheung S-S, Zakhor A. Fast similarity search and clustering of video sequences on the world-wide-web. IEEE Transactions on Multimedia, 2005, 7(3): 524-537
- [4] Hampapur A, Bolle R M. Feature based indexing for media tracking//Proceedings of the IEEE International Conference Multimedia and Expo(ICME). New York, 2000: 67-70
- [5] De Roover C, De Vleeschouwer C, Lefebvre F, Macq B. Robust video hashing based on radial projections of key frames. IEEE Transactions on Signal Processing, Part 2, 2005, 53(10): 4020-4037
- [6] Hampapur Arun, Hyun Ki-Ho, Bolle Ruud. Comparison of sequence matching techniques for video copy detection//Proceedings of the Storage and Retrieval for Media Databases. 2002: 194-201
- [7] Naphade M R, Yeung M R, Yeo B L. A novel scheme for fast and efficient video sequence matching using compact signatures//Proceedings of the SPIE, Storage and Retrieval for Media Databases. 2000, 3972: 564-572
- [8] Hampapur Arun, Bolle Ruud M. Comparison of distance measures for video copy detection//Proceedings of the IEEE International Conference on Multimedia and Expo. 2001: 944-947
- [9] Kim Changick, Vasudev Bhaskaran. Spatiotemporal sequence matching for efficient video copy detection. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(1): 127-132
- [10] Mohan Rakesh. Video sequence matching//Proceedings of the International Conference Audio, Speech and Signal Processing (ICASSP). 1998, 6: 3697-3700
- [11] Bhat Dinkar N, Nayar Shree K. Ordinal measures for image correspondence. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(4): 415-423
- [12] Dong Dawei W. Spatiotemporal inseparability of natural images and visual sensitivities//Zanker Johannes M, Zeil Jochen eds. Motion Vision—Computational, Neural, and Ecological Constraints. Berlin Heidelberg, New York: Springer Verlag, 2001
- [13] Dong Dawei W, Atick Joseph J. Statistics of natural time-varying images. Computation in Neural Systems, 1995, 6(3): 345-358
- [14] Peng S L, Medioni G. Interpretation of image sequences by spatio-temporal analysis//Proceedings of the Workshop on Visual Motion. 1989: 344-351
- [15] Iain E G. Richardson: Video Codec Design, Developing Image and Video Compression Systems. UK: Wiley, 2002



PAN Xue-Feng, born in 1977, Ph. D. candidate. His major research interests include video processing and machine learning.

LI Jin-Tao, born in 1962, professor, Ph. D. supervisor. His major research interests include multimedia processing and V. R. technology.

Background

More and more video contents are encoded in different formats to meet different requirements. How to find the video content in different formats is important for video retrieve and search. Some format-independent video copy detection methods were proposed. The ordinal measure (OM) has shown superior performance than motion and color based features for format-independent video content representing.

According to visual perception theory, the visual sensitivity related motion information is very important to human visual perception. And it will be remained stably in encoding process. So a format-independent motion describing method based on visual sensitivity is proposed in this paper. The spa-

ZHANG Yong-Dong, born in 1973, Ph. D., associate professor. His major research interests include image processing and video processing.

TANG Sheng, born in 1972, Ph. D., associate professor. His major research interests include multimedia retrieval and video processing.

YU Le-Jun, born in 1977, Ph. D.. His major research interests include image processing and video codec.

XIA Tian, born in 1980, Ph. D. candidate. His research interests focus on multimedia retrieval and machine learning.

tiotemporal slice DCT is used to extract the effective features for video content. Experiments show the proposed describing method is sensitive to variation of video content and is robust to variation of video format and the method proposed performance well in video copy detection.

The work in this paper is partially supported by the National Basic Research Program (973 Program) of China (2007CB311100), the National High Technology and Research Development Program (863 Program) of China (2007AA01Z416), the National Nature Science Foundation of China (60773056), and the Beijing New Star Project on Science & Technology (2007B071).