一种新的半脆弱视频水印方案

吕安强
(华北电力大学电子与通信工程系,保定 071003)

摘 要 半脆弱视频水印是一种进行视频数据内容完整性认证的重要技术。为了得到具有更强认证能力的半脆弱视频水印信息,提出了一种新的方案,该方案采用 DCT 块组能量关系特征和块灰度均值特征相结合的双特征提取方法构成水印信息,然后将水印信息进行 Turbo 编码,再利用改进的 DEW 算法实现水印的嵌入。双特征提取算法可以克服单特征提取的不完备性,增强篡改判断和定位能力,Turbo 编码可以提高水印信息的鲁棒性,降低认证虚警率。实验结果表明,该算法在不破坏视觉质量的基础上,能够对常见的篡改操作进行完备的认证,虚警概率小。

关键词 视频水印 半脆弱 特征提取 Turbo 码
中图法分类号: TP309 **文献标识码**: A **文章编号**: 1006-8961(2009)10-1966-06

A Novel Semi-fragile Video Watermarking Scheme

LÜ An-qiang
(*The Department of Electronic and Communication Engineering*, *North China Electric Power University*, *Baoding* 071003)

Abstract Semi-fragile video watermarking is an important technology of videos' content authentication. A novel semi-fragile watermarking scheme is proposed in this paper in order to improve authentication abilities. The scheme obtained the watermarking information by adopting double-feature extraction which was a combination of DCT blocks' relation and grayscale blocks' relation. The watermark was embedded into the host video with an improved DEW algorithm after it had been Turbo coded. The double-feature extraction algorithm can conquer single-feature's imperfection to improve the abilities of tamper estimation and orientation. Turbo code can improve the robustness of watermark and decrease the probability of false positive. Experimental results demonstrate that it will not degrade the visual quality of original videos, and it can authenticate normal tamper operations. Besides the scheme have little probability of false positive.

Keywords video watermark, semi-fragile, feature extraction, Turbo code

1 引 言

近年来多媒体应用取得了惊人的进展。数字媒体易于编辑、合成、复制和传播的优点在给人们带来方便的同时,也使它在知识产权保护和真实性、完整性认证等方面成为人们关注的焦点,这就推动了以知识产权保护和完整性认证为目标的数字水印技术的研究。数字水印技术是利用人的视、听觉特性和媒体内容的冗余性,有控制地将一些标识信息嵌入多媒体中,这些标识信息以后可以用作版权证明、完整性认证、拷贝控制和内容注释等目的。

用于完整性认证的水印又可以分为脆弱水印[1-2]或半脆弱水印[3-4]。前者通常设计为对像素值的每个可能改变非常敏感。大多数情况下脆弱水印算法通过 LSB 修改嵌入水印,这些算法不能容忍"内容保持"类操作。不同于脆弱水印对所有数据的改变都敏感,半脆弱水印的脆弱性是有选择的,它

收稿日期:2008-05-19;**改回日期**:2008-09-23

第一作者简介:吕安强(1979 ~),男。讲师,2005 年于华北电力大学获通信与信息系统专业硕士学位。主要研究方向为数字图像处理、数字水印。E-mail:lvaqdz@163.com

只对反映媒体内容特征的数据改变敏感。即半脆弱水印被设计成能够对“恶意篡改”类操作敏感,而能够经得住“内容保持”类操作。

半脆弱水印的研究始于数字图像的完整性认证,但随着多媒体视频技术的发展,视频监控、电视会议、视频娱乐等多媒体业务的完整性认证工作迫在眉睫。目前,对视频半脆弱水印的研究已经初步展开。视频半脆弱水印的任务是将视频特征作为水印或水印的一部分,在不可见性前提下鲁棒地嵌入到视频数据中,并能完整地提取出水印,判断是否遭到恶意篡改,找到具体篡改的位置。

Fridrich 提出一种基于图像分块,采用类似矢量量化方法的半脆弱水印算法[5],但算法对 JPEG 压缩不够稳健。Kundur 提出基于阈值选择的多层 Harr 小波域半脆弱水印算法[6],但同样存在 JPEG 压缩不够稳健,且阈值选择困难、水印不够安全等缺点。Lin 提出一种基于图像子块 DCT 系数的水印嵌入算法[7],该算法能有效地区分 JPEG 压缩和其他恶意篡改,但特征提取方法单一,不能完备地提取图像的特征。另外,文献[5]~[7]在经过特征提取得到水印后,只对水印做了简单的置乱处理就进行嵌入,以至无法为水印提供足够的鲁棒性,如果在认证过程中不能完整地提取水印,那么篡改判断和定位就无法进行。

综合以上分析可以看出,目前大多数的半脆弱水印都采用单一的特征提取算法,即特征提取不完备,因此无法完备地判断各类攻击并进行精确定位;另外,在水印嵌入之前没有对水印进行鲁棒性处理,给水印的完整提取和篡改的正确判断带来了困难。

在分析大量算法的基础上,提出了一种双特征提取算法,可完备地提取视频特征;另外,为了保证水印的鲁棒性,在水印嵌入之前进行 Turbo 编码,使提取出的水印具有一定的容错能力。

2　水印的生成与处理

2.1　特征提取

视频数据的特征可以从空域、时域和频域 3 个方面考虑。灰度特征可以从空域角度体现视频特征,帧间相关性可以从时域角度体现视频特征,DCT 能量块关系可以从频域角度体现视频特征。综合考虑,本文的视频特征码由子块灰度均值和 DCT 能量块帧间关系与帧内关系构成,从而完整提取空域、时域和频域 3 个方面的视频特征。另外,为了抵抗 MPEG Ⅰ / Ⅱ 的影响,只从关键帧中提取特征。而且从数据域方面考虑,视频特征主要集中在亮度分量中,因此本文特征提取的操作对象是视频亮度信息。

首先,根据 MPEG 算法提取出关键帧 I_n,其中 $n\in[1,N]$,N 是关键帧的总数。根据式(1)计算 Torus 1 维自同构映射[8],得到关键帧之间的一一映射关系 $A\rightarrow B\rightarrow C\rightarrow\cdots\rightarrow A$,其中的 $A,B,C,\cdots$ 都是关键帧,此帧间映射顺序是计算帧间关系的依据。

$$F'=[(k1\times F)\,mod N]+1 \tag{1}$$

式中,$F,F'\in[1,N]$ 为关键帧序号;$k1\in[1,N]$ 是像素数,属于私有密钥。

然后,计算灰度特征信息。将关键帧分成 8×8 的小块 $\boldsymbol{B}_i$,$i\in[1,M]$,M 是一帧中的总块数。进而将每一块 $\boldsymbol{B}_i$ 分割成 4 小块 $\boldsymbol{B}_{ij}$,$j\in[1,4]$,分隔方法如图 1 所示。

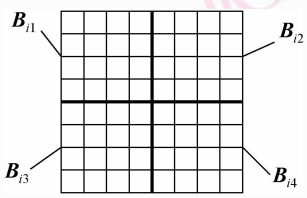

图 1　$\boldsymbol{B}_i$ 块的子块分隔

Fig. 1　Segmentation of B_i blocks

计算块 $\boldsymbol{B}_i$ 的灰度平均值 $\boldsymbol{B}_{i\_avg}$ 和子块 $\boldsymbol{B}_{i1}$ 的灰度平均值 $\boldsymbol{B}_{i1\_avg}$,根据式(2)生成块 $\boldsymbol{B}_i$ 的灰度特征比特 w_h_i。

$$w\_h_i=\begin{cases}1 & 若\ B_{i\_avg}>B_{i1\_avg}\\ 0 & 其他\end{cases} \tag{2}$$

接下来,计算帧内与帧间的 DCT 能量关系特征信息。根据前面分割成的关键帧块 $\boldsymbol{B}_i$,计算每一块的 DCT 系数,进行量化后得 $\boldsymbol{C}_i=\{c_{i1},c_{i2},\cdots,c_{i64}\}$,再根据块的相邻性将 $\boldsymbol{B}_i$ 块组合成矩形的块组,每组内块的数目为 L,L 是偶数或 1,其大小根据视频水印嵌入容量确定,于是 I_n 可以表示成 $\boldsymbol{G}=\{g_1,g_2,\cdots,g_M\}$,且 $N=L\times M$。图 2 为 $L=4$,$M=20$ 时的分组情况。

利用伪随机序列生成密钥 $k2$ 和 $k3$,用 $k2$ 对帧内组块进行配对,用 $k3$ 对帧间组块进行配对,如图 3 所示。

块组配对完成后,按照下列步骤计算帧内特征

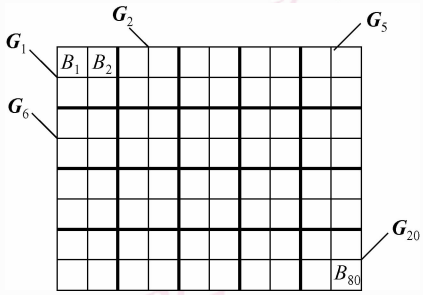

图 2 $L=4$ 时的关键帧块分组情况

Fig. 2 Segmentation grouping of key frames when $L=4$

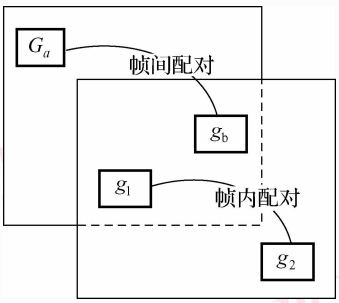

图 3 帧内与帧间块组配对情况

Fig. 3 Segmentation group partnership in intra-frame and inter-frame

比特 w_{z1} 和帧间特征比特 w_{z2}：

(1)根据式(3)计算每一块的能量 E_i；

$$E_i = \sum_{i=1}^{64} (a_i c_i) \tag{3}$$

式中，a_i 是 DCT 系数权值；

(2)根据式(4)计算块组能量 E_m；

$$E_m = \sum_{i=1}^{L} E_i \tag{4}$$

式中，$m \in [1, M]$；

(3)根据式(5)和式(6)计算帧内组对的能量差 $\Delta z1$ 和帧间组对的能量差 $\Delta z2$；

$$\Delta z1 = E_{g1} - E_{g2} \tag{5}$$

$$\Delta z2 = E_{ga} - E_{gb} \tag{6}$$

(4)根据式(7)和式(8)计算特征比特。

$$w_{z1} = \begin{cases} 1 & \text{若 } \Delta z1 \geqslant 0 \\ 0 & \text{其他} \end{cases} \tag{7}$$

$$w_{z2} = \begin{cases} 1 & \text{若 } \Delta z2 \geqslant 0 \\ 0 & \text{其他} \end{cases} \tag{8}$$

最后，经过特征提取得到的特征比特由 w_h，w_{z1} 和 w_{z2} 构成，把它们合并在一起组成特征码 $w_t = w_h + w_{z1} + w_{z2}$。

2.2 水印的构成和预处理

采用的水印数据由两大部分构成：其一是版权信息 w_c，它可以是代表公司标志的图像，也可以是序列号等信息，但都要转换成二进制比特；其二是前面求出的特征码 w_t。把这两部分数据合并到一起称为水印数据 $w = w_c + w_t$。

为了保证水印的安全性，需要将水印数据做置乱处理。根据密钥 $k4$ 生成整数伪随机序列，并以序列中的数值为位置信息将水印 w 置乱。

为了提高水印的鲁棒性，使其能够具有较强的抗攻击能力，根据通信理论中的香农定理，将水印攻击认为是信道噪声，利用纠错码的方法，使水印二值信息具有一定的容错能力。在纠错码的选择上，采用了目前世界上纠错能力最强的编码方法，即第 3 代移动通信中的核心纠错编码技术 Turbo 码[9]，对置乱后的水印进行编码。Turbo 码码率是 1/3，传输函数如式(9)所示。

$$\begin{cases} G(D) = \left[1, \dfrac{g_1(D)}{g_0(D)}\right] \\ g_0(D) = 1 + D^2 + D^3 \\ g_1(D) = 1 + D + D^3 \end{cases} \tag{9}$$

式中，D 是 SRC 编码器的延迟算子。

将待嵌入的二值水印比特 w 输入此编码器，得到 3 倍于编码输入数据量的编码输出比特 w_{turbo}，多出来的比特用于水印纠错。需要说明的是，w_{turbo} 中所有的比特都要嵌入宿主视频中，以便在水印提取后送给解码器进行纠错。

3 水印算法设计

3.1 水印的嵌入

(1)提取原始视频数据中的亮度分量 Y，根据 MPEG 算法计算关键帧，从关键帧中提取特征码 w_t，方法如前面所述；

(2)将版权信息 w_c 与 w_t 合并得到水印 w，置乱后进行 Turbo 编码，得到待嵌入的水印 w_{turbo}；

(3)提取原始视频数据中的色度分量 U，根据 MPEG 算法计算关键帧，使用改进的 DEW 算法[13]嵌入水印。

3.2 水印的提取

水印的提取是嵌入的逆过程，按以下步骤进行：

(1)提取原始视频数据中的色度分量 U，根据 MPEG 算法计算关键帧，使用改进的 DEW 算法提取

出水印 $w_{turbo'}$;

(2)对 $w_{turbo'}$进行 Turbo 解码,得到 w';

(3)根据事先保存的密钥将 w'分解成版权信息 $w_{c'}$和特征码信息 $w_{t'}$,再将 $w_{t'}$分解成灰度特征信息 $w_{h'}$,DCT 能量块组帧内特征信息 $w_{z1'}$和帧间特征信息 $w_{z2'}$。

3.3　特征重建和认证

从提取出的水印数据中恢复视频特征数据,记做 T',此数据和原始水印中的特征码 w_t 应该是基本相同的;再根据前面用到的特征提取方法从可能遭受攻击的含水印视频中重新提取特征信息,记做 T'',此数据说明了遭受攻击后的视频的特征;最后,比较 T'和 T''可以判断视频对象是否遭受恶意篡改。如果发生了篡改,还可进一步探知被篡改的位置。有两种可能的情形:

(1)水印中提取的特征码和重新生成的特征码逐位相等时,则内容可以被认为是"真实"的;

(2)特征码比较中出现 H%(容忍阈值,由经验确定)以上比特不一致情况,可以认为该视频已经被篡改。由于块组是编号的,从而可以根据组号确定篡改位置。但是一个特征位代表 2 个组之间的关系,为了确定哪个组遭受攻击,可以采用的对策是观察这 2 个组邻近的那些组,因为数据块组的尺寸一般很小,而通常一种针对内容的篡改会影响不止一个组。所以通过考察相邻组是否有问题即可做出判断。

4　实验结果

选用 352 × 288 的视频流"mobile. cif(300 帧)","vectra. yuv(142 帧)"和"tempete. cif(260 帧)"进行实验,如图 4 所示。

(a) mobile.cif　　(b) vectra.yuv　　(c) tempete.cif

图 4　选用的 3 组视频

Fig. 4　Three frames adopted

选取这 3 组视频的原因是它们分别代表了不同的视频特点,"mobile. cif"是具有场景渐变、镜头伸缩特点的视频;"vectra. yuv"是场景变化明显的视频;"tempete. cif"是镜头伸缩幅度大、图像细节变化明显的视频。对这 3 种视频做测试,实验结果将具有一定的普遍性和完备性。

4.1　不可见性

水印嵌入需要满足的首要条件就是不可见性。为了定量分析本文算法的不可见性,采用式(10)所示的峰值信噪比均值 PSNR_{mean}来评价含水印视频的质量好坏。

$$\mathrm{PSNR}_{mean}=\frac{1}{N}\sum_{k=1}^{N}10\log\left[\frac{255^2}{\sum_{i=1}^{352}\sum_{j=1}^{288}[X'_{ij}-X_{ij}]^2}\right] \tag{10}$$

式中,N 是关键帧的数量,也是嵌入水印的帧数,X_{ij}和 X'_{ij}分别表示某幅原始视频帧和含水印视频帧中的像素值。

PSNR_{mean}可以反映水印嵌入前后视频帧中对应像素点之间的均方误差情况,此值越大,说明嵌入前后的差别越小,即不可见性越好。表 1 列出了 3 组视频嵌入水印后的 PSNR_{mean}值。

表 1　3 组视频嵌入水印后的峰值信噪比均值

Tab. 1　PSNR mean of three watermarked videos

mobile. cif	vectra. yuv	tempete. cif
48. 51	48. 63	48. 47

目前,大多数文献资料中的峰值信噪比都在 39 ~ 45 dB 之间,本文算法带来的峰值信噪比与目前的最优值相比提高了 3. 5 dB,这是由于采用的是改进的 DEW 算法,且水印嵌入了视觉不敏感的色度分量中。

4.2 篡改检测与定位

在特征提取时采用 DCT 块组能量关系和灰度块均值关系相结合的方法,能够起到优势互补的作用。在检测过程中,分 3 种情况判断篡改的发生,其一,二者都判断为篡改的位置就认定该位置遭到篡改;其二,根据前者判断为篡改,而后者判断为非篡改的,认定为该位置遭到篡改,因为 DCT 能量比灰度特征更能从本质上判断特征的改变;其三,根据前者判断为非篡改,而后者判断为篡改的,需要根据经验阈值 D 认定是否遭到篡改。此阈值的设定主要是针对 MPEG 压缩,因为适量 MPEG 压缩后,根据前者一般不会判断为遭篡改,而后者会判断为大面积遭到篡改。根据采用的 3 组视频,阈值设定为 0.5,即根据后者判断出 50% 以下的位置遭到篡改就认定是篡改,高于此值就认为是 MPEG 压缩,认定为非篡改。

为了说明算法的优越性,比较了 3 种不同方案的测试数据。3 种方案分别为:(1)采用 DCT 单特征提取加 Turbo 编码的方案,用来对比单特征与双特征的完备性;(2)采用双特征提取方案,但不对水印信息进行 Turbo 编码,用以说明 Turbo 编码对水印鲁棒性的作用;(3)本文算法方案。测试数据见表 2。

表 2 不同方案性能比较

Tab. 2 Performance comparison of different schemes

单位:%

	DCT 单特征提取方案	双特征但无 Turbo 码方案	本文算法方案
MPEG(70%)	1.2	5	0
删除 5%	90.1	112.3	100
替换 5%	92	116.4	99.8
移动 5%	95.8	106.9	100

表 2 中数据是 3 组视频测试结果的均值。第 1 行是 70% 的 MPEG 压缩率下,方案 1 和方案 2 分别认定不同比例的像素块遭到篡改,出现虚警的概率,可见本文方案的虚警率是最低的。其余 3 行分别列出了删除、替换或移动 5% 的像素块后,3 种方案所能判断出的像素块占实际遭篡改像素块总数的比例,此数值越接近 100% 就越能说明判断的准确性。从后 3 行数据可以看出,方案 1 只采用单特征提取算法,因此无法完备地认定所有的篡改,方案 2 没有进行 Turbo 纠错编码,导致部分水印数据遭篡改后无法完整地提取,因此出现虚警,只有本文算法能够进行较为正确地判断。

对 3 组视频分别进行了部分删除、部分移动和部分替换的操作,以验证篡改定位能力。以视频 tempete. cif 为例,验证结果如图 5 ~ 图 7 所示,图中用白色小块指示篡改定位结果。

实验结果说明算法具有精确的定位能力。另外,利用帧间 DCT 块组的能量关系还可认定帧序变化方面的篡改,如帧丢弃、帧序颠倒等。

(a) 遭部分删除的视频 (b) 篡改定位结果

图 5 删除了一片树叶和一节栅栏

Fig. 5 One piece of leave and one barrier are deleted

(a) 遭部分移动的视频 (b) 篡改定位结果

图 6 移动了一片树叶和一节栅栏

Fig. 6 One piece of leave and one barrier are moved

(a) 遭部分替换的视频 (b) 篡改定位结果

图 7 用一只小鸟替换了右侧栅栏上的草

Fig. 7 Grass was replaced with one bird

4.3 虚警率

半脆弱水印必须能够尽量区分合法修改和恶意篡改,即应具有较低的虚警率。针对视频数据经常受到的合法修改 MPEG 压缩做了测试,对 3 组嵌入水印的视频进行压缩率分别为原始比特率 90%,70% 和 40% 的压缩,实验结果见表 3 和表 4。

表 3　本文算法在不同压缩率下的虚警率

Tab. 3　Probability of false positive in different compress ratios

	90%	70%	40%
mobile. cif	0	0	1. 2%
vectra. yuv	0	0	1. 12%
tempete. cif	0	0	2. 25%

表 4　单特征且无 Turbo 编码的算法在不同压缩率下的虚警率

Tab. 4　Probability of false positive in different compress ratios with single feature and non-turbo code

	90%	70%	40%
mobile. cif	2. 1%	5. 4%	9. 7%
vectra. yuv	2. 9%	6. 1%	9. 2%
tempete. cif	2. 4%	7. 2%	11. 3%

比较表 3 和表 4 中的数据可以看出,在常用压缩率下,本文算法的虚警率为零,即使在高压缩率时,本文算法也比单特征提取且无 Turbo 编码的常规算法的虚警率低。这里主要有两个方面的原因:(1)采用了双特征提取算法,其中的一个特征是 DCT 能量块组之间的关系,而 MPEG 压缩前后,视频帧中的 DCT 块组能量大小关系是不变的;(2)采用了 Turbo 编码,即使在压缩过程中造成水印的改变,也能在解码时得到纠正,从而降低了虚警率。

5　结　论

针对目前半脆弱水印特征提取算法的不完备性和篡改认证虚警率高的问题,提出了一种新的半脆弱视频水印算法。该算法利用双特征提取和第 3 代移动通信中的 Turbo 编码技术,使特征提取更加完备,水印更加鲁棒。从实验结果可以看出,双特征提取具有更强的篡改判断和定位能力,而 Turbo 编码使水印认证的虚警率得到降低。

致　谢　本文受到华北电力大学青年教师基金项目(200611032)的支持,在此表示感谢。

参考文献(References)

1　Sattar F, Barkat B. A new time-frequency based fragile watermarking scheme for image authentication [A]. In: Proceedings of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing[C], Victoria, BC, Canada, 2003: 992-995.

2　Zhang Xiao-hua, Meng Hong-yun, Liu Fang. A new kind of efficient fragile watermarking technique [J]. Acta Electronica Sinica, 2004, **32**(1): 114-117.

3　Liu Q, Jiang X M. A unified digital watermark algorithm based on singular value decomposition and spread spectrum technology [J]. Chinese Journal of Electronics, 2005, **14**(4): 621-624.

4　Maeno K, Sun Q, Chang S F, *et al*. New semi-fragile watermarking technique using random bias and non-uniform quantization [A]. In Proceedings of SPIE Security and Watermarking of Multimedia Contents IV[C], San Jose, CA, USA, 2002: 659-670.

5　Fridrich J. Image watermarking for tamper detection [A]. In: Proceedings of the International Conference on Image Processing[C], Chicago, IL, USA, 1998: 404-408.

6　Kundur D, Hatzinakos D. Digtal watermarking for telltale tamper proofing and authentication [J]. Proceedings of the IEEE, 1999, **87**(7): 1167-1180.

7　Lin C Y, Chang S F. Semi-fragile watermarking for authenticating JPEG visual contents [A]. In: Proceedings of SPIE Security and Watermarking of Multimedia Contents[C], San Jose, CA, USA, 2000: 140-151.

8　Voyatzis G, PITAS. Chaotic mixing of digital images and applications to watermarking [A]. In: Proceedings of the European Conference on Multimedia Applications Services and Techniques[C], Louvain-La-Neuve, Belgium, 1996: 687-689.

9　Berrou C, Glavieux A, Thitimasjshima P. Near Shannon limit error-correcting coding and decoding: Turbo-codes[A]. In: Proceedings of ICC'93, 1993: 1064-1070.