

基于代价策略的 H.264/AVC 半脆弱水印算法

张维伟¹, 赵宸², 黄德天¹, 张沛³, 杨义先³

(1. 华侨大学 工学院, 福建 泉州 362021; 2. 国家计算机网络与信息安全管理中心, 北京 100031;

3. 北京邮电大学 信息安全中心, 北京 100876)

摘要: 针对目前基于 H.264/AVC 标准的视频半脆弱水印算法存在的问题, 提出一种基于代价策略的 H.264/AVC 半脆弱水印算法。通过消除基于内容的特征码之间的相关性, 提高内容认证水印的顽健性。为实现半脆弱水印既能准确定位篡改区域, 又对视频质量和码率的影响最小, 引入了一个代价函数, 通过该函数在残差 4×4 块的 DCT 系数中选择一个最佳的系数嵌入水印。实验表明, 所提算法嵌入水印对视频质量和码率的影响较小, 半脆弱水印对恶意篡改的图像区域能有效、准确地检测和定位。

关键词: 半脆弱水印; 代价函数; H.264/AVC; 篡改定位; 顽健性

中图分类号: TP309

文献标识码: A

Semi-fragile video watermarking algorithm for H.264/AVC based on cost strategy

ZHANG Wei-wei¹, ZHAO Chen², HUANG De-tian¹, ZHANG Pei³, YANG Yi-xian³

(1. College of Engineering, Huaqiao University, Quanzhou 362021, China;

2. National Computer Network and Information Security Management Center, Beijing 100031, China;

3. Information Security Center, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: In order to solve some existing problems on detecting malicious tampering by video watermarking based on H.264/AVC video compression standard, a semi-fragile video watermarking algorithm for H.264/AVC based on cost strategy was proposed. To further improve the robustness of the feature codes in semi-fragile video watermarking algorithm, a new generation algorithm for feature codes by eliminating the correlation of them was proposed. In order to meet the requirement that the semi-fragile watermarks should be uniformly embedded into the video frames to locate the tampered region accurately, a minimal cost function was introduced to select a DCT coefficient in the 4×4 block residual to embed watermark which had minimal impact on the video quality and bit rates. Experimental results show that the proposed algorithm can significantly reduce impact on video quality and bit rate because of embedding the watermarking by adopting the minimum cost function and it is sensitive to malicious manipulations. The retrieved feature code and watermarking can detect and locate malicious manipulations.

Key words: semi-fragile watermarking; cost strategy; H.264/AVC; tamper localization; robustness

1 引言

随着网络和多媒体技术的发展, 视频点播、社交媒体等新型数字化视频产品越来越受到欢迎, 成

为人们生活中不可或缺的一部分。然而, 随着多媒体处理和编辑技术的发展和广泛使用, 人们可以轻易地编辑和修改新型媒体作品。如果经过伪造和篡改的多媒体涉及国家安全、科研成果、医疗鉴定和

收稿日期: 2014-10-07; 修回日期: 2014-12-22

基金项目: 华侨大学科研基金资助项目 (13BS415); 泉州市科技计划基金资助项目(2014Z112); 福建省自然科学基金资助项目(2015J05125); 福建省科技厅专项基金资助项目(2013H2002)

Foundation Items: The Research Foundation of Huaqiao University(13BS415); The Science and Technology Development Foundation of Quanzhou City(2014Z112); The Natural Science Foundation of Fujian Province(2015J05125); The Special Foundation of Fujian Provincial Department of Science and Technology(2013H2002)

法庭证据等, 可能会造成严重的后果。因此, 视频内容的真实性与完整性认证已成为近年来多媒体内容安全领域的研究热点^[1]。半脆弱水印可以为视频完整性认证提供一个潜在的解决方案^[2,3]。

针对数字图像内容篡改检测的算法已经得到了学术界的大量研究, 很多脆弱水印算法^[4-7]具有较好的篡改区域检测、定位甚至复原能力。近年来, 随着视频产品的广泛应用, 针对视频内容的篡改检测技术也逐渐受到学者的关注。Chen 等^[8]提出了基于混沌半脆弱水印的视频内容认证方案。该方案把视频的时域信息调制成混沌系统的输入参数, 混沌系统的输出作为水印嵌入到基于块的 DCT 域中。通过该水印可以检测时域和空域的篡改。Shi 等^[9]提出了一种基于运动目标的双水印视频内容认证方案, 首先把视频每一帧的索引号作为可逆水印嵌入到相应帧的运动目标用于检测时域篡改; 再把每帧的内容轮廓信息和运动目标的细节与认证码进行绑定作为另一个水印嵌入到视频帧中, 用于空域篡改的检测及篡改内容的复原。Masoumi 等^[10]利用 CDMA 技术把水印嵌入到视频运动场景的小波子带的中频系数, 水印能检测恶意的内容篡改。

文献[8~10]提出的方案不仅定位精度高, 而且都能检测视频时域和空域的恶意篡改, 但都是基于未压缩域的视频认证方案。然而, 视频的信息量通常较大, 大部分视频数据在存储和传输时都要进行压缩编码, 而压缩编码会对水印信息造成一定的损失。另外, 在实时性要求较高的场合, 基于未压缩域的视频认证方案也无法满足要求。因此基于压缩编码标准的视频水印方案越来越受到关注。

H.264/AVC^[11]是由 ISO/IEC 的 MPEG 标准委员会和 ITU-T 的 VCEG 共同制定的视频编码标准。与 H.263 或 MPEG-4 相比, 在同等图像质量情况下, 其码率能降低一半^[12]。H.264/AVC 已被广泛应用于视频点播服务、流媒体网站和其他一些如 YouTube 的视频分享网站, 是当前市场最受欢迎的压缩标准。显然, 文献[8~10]的方案无法直接应用于 H.264/AVC 领域。

近年来, 学术界也提出了不少基于 H.264/AVC 的视频认证水印算法。Hong 等^[13]对视频信号的时空域特性进行分析, 将视频在 H.264/AVC 压缩编码过程中宏块的帧内或帧间的预测模式当成特征码, 并把该特征码当成基于内容的水印嵌在压缩码流的应用层。该算法嵌入水印对视频质量和码率造成

的影响降到最小, 而且水印能有效检测常见的“内容篡改”类攻击。Kim 等^[14]提出的算法把认证水印嵌在帧间编码宏块的运动矢量 LSB 上或帧内编码的模式系数上。该算法嵌入水印后, 在保持视频质量和码率变化不大的同时, 水印的容量也很大。另外, 也能有效检测内容篡改。Li 等^[15]提出把认证水印嵌在运动矢量的水平和垂直分量中, 为了减少嵌入水印对视频质量的影响, 算法选择在幅值较大的运动矢量分量中嵌入水印。Xu 等^[16]在嵌入水印过程中首先标识参考帧的指数哥伦布码, 然后建立水印信息和这些码字之间的映射规则。算法根据映射规则调制水印信息对应的指数哥伦布码字嵌入水印。提取水印信息时, 不需要原始视频参与, 只要从码流解析指数哥伦布码字。该算法对视频质量和码率的影响极小。由于重编码或者转码都会移除水印, 算法以此判断视频是否被篡改。

已提出的这些用于视频内容完整性认证的水印算法^[13-16]对于视频内容的任何修改都很敏感, 属于脆弱性水印。但是由于对内容修改的高度敏感性, 使脆弱水印算法无法区别对内容的恶意篡改和一些普通的视频信号处理(如压缩、滤波等)操作, 而这些处理并不会改变视频的内容。但在许多视频内容认证场合, 人们往往希望能区分视频内容被改变的类型, 即内容被改变是“内容保持”类改变, 还是人为的“恶意篡改”。显然采用脆弱水印技术无法满足上述要求, 这就需要借助半脆弱数字水印。

目前, 基于 H.264/AVC 的半脆弱视频水印算法的公开文献还很少。Chen 等^[17]在 I 帧使用块的子带索引因子调制量化后的 AC 系数嵌入水印信息。该算法嵌入水印对视频质量和码率的影响较大。林等^[18]首先根据一个宏块上下 2 部分子块残差 DCT 系数量化后的低频能量之间的关系构造内容特征码, 再通过修改子宏块中拖尾系数的编码方式嵌入半脆弱认证水印。该算法在水印嵌入前后视频的质量及码率基本保持不变, 但算法定位篡改区域的精度仅为宏块级别(16×16), 而且由于同一宏块的相关性, 算法的特征码顽健性不高。王等^[19]根据 I 帧宏块之间残差量化 DCT 能量的不变特性生成内容特征码, 并根据 DCT 系数的特点, 通过符号编码的方式把特征码嵌到较小的 DCT 系数中。为了保持水印嵌入前后视频流码率的恒定, 算法引入了误差准则函数, 通过该函数把水印嵌在对码率影响最小的 4×4 块的 DCT 系数上。算法能定位视频图像中遭到恶意篡改

的某些区域,但由于算法是选择在 4×4 块的 DCT 系数满足幅值小于某个阈值的中低频分量嵌入水印,虽然在一定程度上减小嵌入水印对视频质量和码率的影响,但如果块中不存在满足嵌入条件的 DCT 系数时,则该块无法嵌入水印,即使该块遭到恶意篡改,也无法检测和定位。Xu 等^[20]首先利用宏块中 2 个块组的差分能量值在重编码前后保持相对不变的特性生成特征码,利用符号编码的方式把水印嵌在 4×4 块对角线的 DCT 系数上。为了能够定位时域上的帧篡改,算法对每个 P 帧的帧序号进行二进制编码,并进行扩频处理,再把扩频后的信息嵌在 P 帧的 DCT 系数上。算法利用一个滑动窗口统计在 I 帧提取的水印和特征码之间的差值中“1”的分布,定位对空域上内容的篡改。该算法能够定位空域和时域上的篡改,但在空域上定位的精度也仅为宏块级别。不久前, Mehdi 等^[1]提出将宏块和帧的序号作为水印嵌入到残差 4×4 块的 DCT 系数的最后一个非零系数上。算法对视频和码率的影响较小,且半脆弱水印能定位时域和空域的篡改区域。但算法并没有在每个 4×4 块上嵌入半脆弱水印,因此,算法可能无法检测和定位空域上的篡改区域。

已提出的基于 H.264/AVC 的视频半脆弱水印算法存在嵌入水印对视频质量和码率影响较大(如文献[17])、对空域恶意篡改区域定位精度不高(如文献[1, 18~20])等问题。针对这些问题,本文提出一种新的基于 H.264/AVC 的视频半脆弱水印算法。首先,通过消除特征码之间的相关性提高其顽健性。其次,为了提高定位篡改区域的准确性及水印的不可见性,本文引入了一个最小代价函数,即嵌入水印对视频质量和码率影响代价最小的函数,通过该函数在 DCT 域中选择最佳的水印嵌入位置。和已有的算法相比,本文提出的算法可以较明显地降低嵌入水印对视频质量和码率的影响,并且半脆弱水印对于可接受的“内容保持”类修改具有顽健性,而对“内容篡改”类修改具有敏感性,并且能对恶意篡改的区域准确定位。

2 特征码的生成

在基于特征码的水印算法中,特征码的顽健性直接影响水印的顽健性。因此,有效且顽健的特征码提取算法至关重要。Lin 等^[21]证明了图像在较高质量因子的 JPEG 压缩前后不同 DCT 系数块相应位

置上系数的大小关系保持不变。在 MPEG 压缩领域,同样存在这种不变特性^[22]。他们根据这种不变特性生成顽健的特征码。然而,与 JPEG 或 MPEG 压缩编码标准不同,H.264/AVC 采用帧内预测技术,重编码前后 I 帧帧内的预测模式和残差 DCT 系数大小的关系都会发生一定程度的改变^[23],即压缩前后不同 DCT 系数块相应位置上系数的大小关系不再保持不变。因此,在 JPEG 或 MPEG 领域根据压缩前后 DCT 系数大小关系的不变特性生成的特征码不能直接应用于 H.264/AVC 领域。

针对 H.264/AVC 的新特性,Xu 等^[20]提出在每个宏块(16×16)中随机选择 8 个 4×4 块,每 4 个 4×4 块分为一组,分别计算这 2 组 4×4 块残差的 DCT 系数与预测块 DCT 系数的绝对值之和(也称为能量),再根据能量的大小关系产生 1 bit 特征码。然而,由于视频图像的像素块之间存在相关性,特别是背景区域面积比较大的图像,相关性较强。因此,在同一个宏块中,2 个块组的能量大小可能很接近甚至相等。视频在经过编码和重编码后,由于预测模式和残差 DCT 系数的改变,2 个块组的能量大小关系可能发生改变。

针对这个问题,本文提出一种新的特征码生成算法。由于 H.264/AVC 帧间编码采用运动补偿技术,编码效率较高,P 帧和 B 帧可供水印嵌入的空间很小^[24],因此本文选择在 I 帧嵌入水印。但为提高特征码的顽健性,本文提出的算法对第 1 个 I 帧不嵌入水印,而是直接提取特征码,再把该特征码嵌在第 2 个 I 帧,依此类推。虽然视频的首帧并没有嵌入水印,只是提取特征码,但这并不影响算法的实用性。因为如果只是篡改视频的第 1 帧,并不会对视频的观看者造成任何影响,而视频的其他 I 帧都进行水印的嵌入,可以进行篡改检测和定位。第 1 帧不嵌入水印是为了通过置乱的方式消除每 1 个 I 帧块组能量之间的相关性,提高特征码的顽健性。每 1 个 I 帧提取特征码的算法如下。

step1 依次从 I 帧的左到右、上到下计算每个 4×4 块的 DCT 系数能量。

假设 I 帧图像的大小为 $M \times N$,则 I 帧 4×4 块的个数 m 为 $\frac{M}{4} \times \frac{N}{4}$ 。每个 4×4 块的能量为

$$E_i = \sum_{u=0}^3 \sum_{v=0}^3 |P_i(u,v)| + \sum_{u=0}^3 \sum_{v=0}^3 |R_i(u,v)| \quad (1)$$

其中, $i=1,2,\dots,m$, $R_i(u,v)$ 为第 i 个残差块的 DCT

系数值， $P_i(u, v)$ 表示第 i 个预测块的 DCT 系数值， (u, v) 为系数在块中的位置， E_i 为块的能量。

step2 重复 step1，直到获得整个 I 帧 4×4 块的能量值，再按从左到右、从上到下的顺序把这些能量值排成一维向量 $\mathbf{E} = \{E_1, E_2, \dots, E_m\}$ 。

step3 生成一维 Logistic 伪随机序列

$$x_{k+1} = \mu x_k (1 - x_k) \quad (2)$$

其中，令 $1 \leq k \leq m$ ， $\mu = 3.65$ 。

step4 对 Logistic 序列的值按从小到大的顺序进行排序，记录排序后各个元素在原来 Logistic 序列中的位置值，再把该值作为向量 $\mathbf{E} = \{E_1, E_2, \dots, E_m\}$ 各个元素新的位置值进行置乱，得到新的一维向量 \mathbf{E}' 。

step5 假设每个 I 帧要生成 n 个特征码，先把一维向量 \mathbf{E}' 分成 $\frac{2m}{n}$ 组，再分别计算每个组的能量之和，根据相邻的 2 个组能量之和的差值生成特征码。

表 1 为在同一实验平台上本文算法和文献[20]算法特征码顽健性之间的比较。实验基于 H.264 参考软件模型 JM12.4。实验中使用的测试序列为 CIF(352×288)格式的 Foreman、Mobile、Coastguard、Hall、Flower。实验条件为每种序列都是在帧率为 15 帧/秒，帧数为 100 帧，并按 IPPPPPP... 的顺序进行编码，每隔 10 帧编码一个 I 帧。表 1 中的数值表示在量化参数 QP 分别为 26、28 和 30 的重编码条件下，提取正确的特征码比特数与在第一次编码($QP=28$)时提取的特征码总比特数的比值。从表 1 可以看出，经过 H.264/AVC 重压缩后，本文的算法优于文献[20]算法。这是由于对每个 I 帧的 4×4 块 DCT 系数能量进行置乱，有效消除了块组能量之间的相关性，提高了特征码的顽健性。

3 水印的嵌入

半脆弱水印不仅应满足抵抗“内容保持”类修改的顽健性和不可见性，而且还应满足检测“内容

篡改”类修改的敏感性^[25,26]。因此，在设计半脆弱水印算法时，既要考虑在图像上嵌入大量而均匀的半脆弱水印，又要考虑嵌入水印对图像质量和码率的影响。结合 H.264/AVC 的编码特性，本文提出一种新的基于代价策略的半水印算法。

3.1 水印嵌入的最小代价策略

半脆弱水印的特性要求即使是在视频图像的平坦区域也要嵌入水印，这对算法提出了更高的要求。针对如何在 H.264/AVC 编码过程中的 DCT 系数上嵌入水印，Zhang 等^[27]提出了基于 H.264/AVC 的视频顽健性水印算法，该算法首先将作为水印信息的灰度图像进行压缩编码，然后对该编码结果进行扩频处理，最后将扩频后的水印信息嵌在残差 4×4 块 DCT 域的某个固定位置的对角线系数上。该算法没有充分考虑人眼的视觉特性，水印的不可见性较差；另外，在固定位置系数上嵌入水印，通常这个系数是零值，在零值上嵌入水印，水印顽健性较差。文献[15]采用误差准则函数在 4×4 块中选择对视频图像像素值影响最小的 DCT 系数嵌入水印。该算法计算代价小，只要对嵌入水印的 DCT 系数进行反量化，再选择对视频图像像素值影响最小的 DCT 系数嵌入水印。但算法中的加权参数需要根据实验确定，并且误差准则函数只是考虑嵌入水印对单个系数产生的量化误差的影响，而不是对整个块的视频质量的影响。

本文采用一个最小代价选择函数，根据这个函数把水印信息嵌在对整个 4×4 块图像质量和码率代价影响最小的 DCT 系数上。具体的算法如下。

step1 在量化后的 4×4 DCT 系数矩阵 \mathbf{C} 上嵌入水印

$$\mathbf{C}_i^w = \mathbf{C} + \Delta_i^w \quad (3)$$

其中， Δ_i^w 为水印矩阵，该矩阵除了在位置 i 上有一个非零值，其余的元素都为 0，非零值的大小表示

表 1 本文算法与文献[20]特征码顽健性比较

测试序列	$QP=26$		$QP=28$		$QP=30$	
	文献[20]	本文算法	文献[20]	本文算法	文献[20]	本文算法
Foreman	0.895 2	0.963 0	0.912 2	0.982 5	0.867 5	0.981 3
Mobile	0.956 3	0.986 5	0.949 8	0.987 6	0.943 1	0.988 3
Coastguard	0.887 9	0.968 3	0.907 2	0.966 3	0.873 8	0.968 5
Hall	0.875 6	0.967 7	0.889 3	0.987 2	0.867 2	0.982 7
Flower	0.814 5	0.979 8	0.853 2	0.983 2	0.813 2	0.982 1

嵌入水印的强度。

step2 对嵌入水印后的矩阵进行反量化和 DCT 反变换, 得到残差矩阵

$$R_i^w = \text{IDCT}(\text{dequant}(C_i^w)) \quad (4)$$

其中, $\text{dequant}(C_i^w)$ 为反量化函数, $\text{IDCT}()$ 表示 DCT 反变换运算。

step3 重构嵌入水印后的 4×4 块图像块矩阵

$$Y_{\text{RE}i}^w = R_i^w + Y_{\text{RE}} \quad (5)$$

其中, Y_{RE} 是预测模式下的当前 4×4 块的预测值矩阵。

step4 计算图像失真函数值

$$D(i) = \text{SAD}(Y_{\text{org}}, Y_{\text{RE}i}^w) \quad (6)$$

其中, Y_{org} 是当前 4×4 块的图像像素值, 函数 $\text{SAD}()$ 为绝对残差和, 定义如下

$$\text{SAD} = \sum_{m=0}^3 \sum_{n=0}^3 |Y_{\text{org}}(m,n) - Y_{\text{RE}i}^w(m,n)| \quad (7)$$

其中, (m,n) 为图像像素在 4×4 块中的位置值。

step5 计算嵌入水印并进行变换和量化后的 DCT 系数所需的比特数 $\text{Bit}(i)$ 。

step6 计算代价函数

$$J(i) = D(i) + \lambda \text{Bit}(i) \quad (8)$$

其中, λ 为加权参数, 其大小由量化参数确定。

step7 选择使代价函数值 J 最小的 i 作为最佳的水印嵌入位置。

根据最小代价选择函数, 可以得到 4×4 块中最佳的水印嵌入位置。

3.2 水印嵌入算法

水印的嵌入算法如下。

step1 在残差 4×4 块中采用最小代价选择函数选择水印的嵌入位置, 并记录该系数的位置信息作为解码端提取水印的密钥。嵌入水印的算法为

$$cw_i' = \begin{cases} \max\{cw_i, level\}, & W_i = 1 \\ 0, & W_i = 0 \end{cases} \quad (9)$$

其中, cw_i 为 DCT 系数, cw_i' 是嵌入水印后的 DCT 系数, W_i 为水印信息, $level$ ($level$ 为正整数) 是水印嵌入的强度。

step2 循环 step1, 直到嵌入所有的水印信息。

3.3 水印的提取

水印的提取算法如下。

step1 得到嵌入水印的位置信息。

step2 采用下列方式提取水印信息

$$W_i' = \begin{cases} 1, & |cw_i'| > 0 \\ 0, & \text{其他} \end{cases} \quad (10)$$

其中, cw_i' 是嵌入水印后的 DCT 系数, W_i' 为提取的水印信息。

step3 循环 step1 和 step2, 直到提取所有的水印。

4 视频内容认证

在 H.264/AVC 的解码端进行视频内容是否被恶意篡改的认证。首先分别提取各个 I 帧的特征码和水印信息(第 1 帧只提取特征码), 并把当前 I 帧提取的特征码与下一个 I 帧提取的水印信息进行比较, 以确定篡改类型和进行篡改区域定位。

4.1 篡改判别准则

首先把提取的一维特征码和水印向量转化为二维矩阵, 定义水印的差值图像

$$W^* = |w_{ij} - w'_{ij}| \quad (11)$$

其中, w_{ij} 表示提取的特征码二维矩阵, w'_{ij} 表示提取的水印二维矩阵。

差值图像 W^* 上 “1” 的分布可以反应视频 I 帧被改变的类型。如果对视频 I 帧进行滤波或添加噪声等 “内容保持” 类修改, 则 W^* 上 “1” 的分布呈均匀而随机; 如果是 “内容篡改” 类修改, 则 W^* 上 “1” 的分布相对比较集中。

Hu 等^[28]根据 W^* 上 “1” 的分布定义了差值图像中的稠密点和稀疏点。即在差值图像中若某个像素值为 “1”, 并且其 8 邻域中至少有一个 “1”, 把该点称为稠密点; 反之, 定义为稀疏点。并根据稠密点所占的比例判断图像被修改的类型

$$r = \frac{N_d}{N_s + N_d} \quad (12)$$

其中, N_s 为稀疏点的总个数; N_d 为稠密点的总个数。

判断图像被修改的类型

$$\begin{cases} \text{图像未受篡改,} & N_d + N_s = 0 \\ \text{内容保持类篡改,} & r < T \\ \text{恶意篡改,} & r \geq T \end{cases} \quad (13)$$

其中, 阈值 T 为判断图像被修改类型的临界值。 T 越大, 对恶意篡改的判断条件就越严格, 算法安全性也就越高, 但可能出现误判; T 越小, 对

恶意篡改的判断条件就越宽松，但可能出现漏判。因此， T 的大小由实验决定。本文实验中 T 取 0.8。

4.2 篡改区域的定位

根据式(13)对差值图像进行判定，如果图像被修改的类型为恶意篡改，则对篡改区域进行定位。定位之前对篡改矩阵进行去除虚警和漏警噪声处理^[29]。

去除虚警噪声。如果某一个报警点 $w^*(i, j)$ 的 8 连通域内仅有少于 1 个的报警点，即

$$\sum_{i,j=1}^1 w^*(i, j) \leq 2 \quad (14)$$

则判定该位置为虚警，在篡改矩阵中将其删除。

去除漏警噪声。如果某一个非报警点 $w^*(i, j)$ 的 8 连通域内有多于 4 个的报警点，即

$$\sum_{i,j=1}^1 w^*(i, j) \geq 4 \quad (15)$$

判定该位置为漏警，在篡改矩阵中将其进行填补。

式(14)和式(15)中， $w^*(i, j)$ 是矩阵 W^* 中的元素。

5 实验结果与分析

本文实验基于 H.264 参考软件模型 JM12.4，实验中使用的测试序列为 CIF(352×288)格式的 Foreman、Mobile、Coastguard、Hall、Flower。测试的条件：帧率为 15 帧/秒、帧数为 100 帧、并按 IPPPPPP... 的顺序进行编码，每隔 10 帧编码一个 I 帧，QP 值设为 28。式(9)中， $level$ 的取值为 2。

5.1 水印不可见性及对比特率的影响分析

图 1 和图 2 所示分别为 Foreman 和 Mobile 序列的第 30 帧(I 帧)在嵌入水印前后图像质量的对比。由图可见，从主观上无法判断 Foreman 和 Mobile 序列嵌入水印前后图像质量的差别。在本文实验中，嵌入水印后 5 种测试序列的亮度分量 PSNR 值

平均下降 0.27 dB。



(a) Foreman 第 30 帧原始图像 (b) Foreman 第 30 帧水印图像

图 1 Foreman 序列嵌入水印前后的比较



(a) Mobile 第 30 帧原始图像 (b) Mobile 第 30 帧水印图像

图 2 Mobile 测试序列嵌入水印前后的比较

表 2 为在同一实验平台上本文算法和文献[27]算法水印性能的比较。实验条件为每种序列都是在帧率为 15 帧/秒，帧数为 100 帧，并按 IPPPPPP... 的顺序进行编码，每隔 10 帧编码一个 I 帧，QP 值设为 28。2 种算法嵌入的水印容量相同，即每个 I 帧嵌入 396 bit，10 个 I 帧共嵌入 3 960 bit。由表 2 可知，本文算法嵌入水印后对 PSNR 值和码率的影响都比文献[27]小。其主要原因有 2 点：首先，本文算法采用了最小代价函数，在对每个 4×4 残差块整体像素值和码率影响最小的 DCT 系数上嵌入水印，而文献[27]是在 4×4 块对角线的一个固定位置上嵌入水印；其次，采用最小代价函数嵌入水印不仅会减少对当前块 PSNR 值和码率的影响，也会减少因嵌入水印产生的误差漂移对临块的影响，从而使临块的帧内预测更加准确，降低码率。

表 2 本文算法和文献[27]在嵌入水印后对 PSNR 值和码率的影响比较

测试序列	PSNR/dB			码率/(kbit·s ⁻¹)		
	原始	文献[27]	本文	原始	文献[27]	本文
Foreman	37.49	36.96	36.98	283.9	323.58	354.27
Mobile	35.43	34.86	35.22	1 260.20	1 307.02	1 296.12
Coastguard	35.45	35.24	35.25	865.95	887.90	921.89
Hall	38.30	37.64	37.65	206.89	238.17	275.39
Flower	36.44	35.89	36.18	1 096.27	1 135.86	1 139.13

5.2 水印顽健性分析

为了测试本文算法水印的顽健性，在表 2 的测试实验条件下，测试水印抵抗各种“内容保持”类修改的顽健性。图 3 为嵌入水印且受到各种攻击的视频图像，图 3(a)为重编码攻击，图 3(b)~图 3(d)分别为高斯噪声攻击、椒盐噪声攻击和高斯低通滤波，从图中可以看出，图 3(b)~图 3(d)对视频质量都有比较明显的影响，特别是加入高斯噪声的图 3(b)。

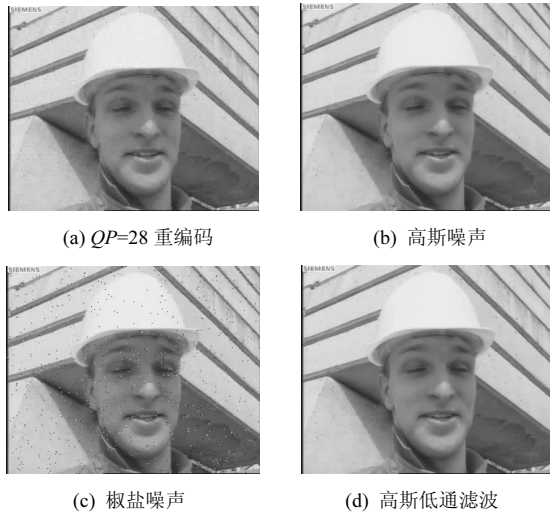


图 3 Foreman 序列的第 1 帧在各种攻击下的图像

表 3 为本文算法和文献[27]水印的顽健性比较。表中的数值为在解码端提取的正确水印比特数与嵌入水印总比特数的比值

$$sim = \frac{NUM_d}{NUM_e} \quad (16)$$

其中， NUM_d 表示在解码端解码时正确提取水印的比特数， NUM_e 表示嵌入水印的总比特数， $NUM_d \leqslant NUM_e$ 。

对于顽健性水印，要求水印载体受到各种攻击后水印仍能被准确地提取，因此 sim 值越大表示水印的顽健性越强。若 sim 为 1，则表示嵌入的水印

信息全部被正确提取。表 3 表明，本文算法的半脆弱水印能有效抵抗常规信号处理攻击，顽健性也优于文献[27]。对于 5 种测试序列，半脆弱水印对于重编码和高斯低通滤波的顽健性较强，原因是重编码和高斯低通滤波攻击对视频质量的改变较小；而高斯噪声对视频图像的改变最大，水印抵抗高斯噪声的能力最弱。

从表 3 还可以看出，本文算法的 Hall 序列水印的顽健性最差，主要原因是与其他的测试序列相比，Hall 序列 I 帧平滑区域较多，纹理区域较少，而且序列之间时域变化平缓。因此，Hall 序列 I 帧残差块经过整数 DCT 变换和量化后，非零 DCT 系数较少。而在零系数上嵌入水印，顽健性较差。反之，Mobile 序列 I 帧纹理区域较多，运动较剧烈，水印的顽健性也较好。

5.3 篡改区域的定位

本文算法在视频 I 帧的每个 4×4 块中嵌入 1 bit 水印，定位精度为 4×4 。当没有对视频的 I 帧图像进行添加噪声、滤波等常规信号处理，而仅改变内容(视频图像中的对象被删除)，如图 4(b)中的篡改(图中黑色圆圈内为遭篡改部分)发生时，利用本文算法对篡改区域进行定位，定位结果如图 4(c)所示。从图 4(c)可以看出，能准确定位内容遭篡改的区域。

半脆弱水印算法对篡改区域的定位性能可通过虚警检测率 P_{FP} 、误警检测率 P_{FN} 和平均检测率 P_D 测定^[30]，其定义分别为

$$P_{FP} = \frac{n_{FP}}{N} \quad (17)$$

其中， n_{FP} 表示图像中未被篡改的块却被误检为被篡改的块， N 为总的块数。

$$P_{FN} = \frac{n_{FN}}{N} \quad (18)$$

其中， n_{FN} 为图像中被篡改的块被误检为未被篡改的块。

表 3 本文算法和文献[27]水印的顽健性比较

测试序列	QP=26		高斯噪声(方差 0.000 5)		椒盐噪声(密度 0.01)		高斯低通滤波(5×5)	
	文献[27]	本文	文献[27]	本文	文献[27]	本文	文献[27]	本文
Foreman	0.847 2	0.910 9	0.841 4	0.784 2	0.766 7	0.832 5	0.849 5	0.873 1
Mobile	0.888 9	0.926 6	0.800 0	0.821 7	0.838 4	0.871 4	0.868 7	0.935 0
Coastguard	0.913 6	0.938 8	0.829 0	0.813 0	0.800 8	0.852 8	0.891 4	0.911 8
Hall	0.792 7	0.850 3	0.763 1	0.723 7	0.705 8	0.840 1	0.743 4	0.851 1
Flower	0.780 6	0.847 8	0.740 2	0.756 8	0.761 4	0.840 2	0.803 5	0.850 9

$$P_D = 1 - \frac{P_{FP} + P_{FN}}{2} \quad (19)$$

在实际的应用场合，视频图像内容通常会遭受“内容保持”类修改和“内容篡改”类修改的叠加攻击，为了测试水印抵抗该类攻击的性能，本文测试 Mobile 序列嵌入半脆弱水印后分别在同时遭受复制粘贴这一“内容篡改”类修改和重编码、加椒盐噪声(如图 5(a))和高斯噪声(如图 6(a))这 3 种“内容保持”类修改时水印的性能，测试结果如表 4、表 5 和表 6 所示。从表 4 可以看出，当对视频 I 帧图像的内容篡改 20%，同时进行重编码，对恶意篡改内容定位的误检率在 5%以下。从表 5 和表 6 可以看出，对内容进行同样的恶意篡改，同时添加噪声后，误检率仍在 7%以下。

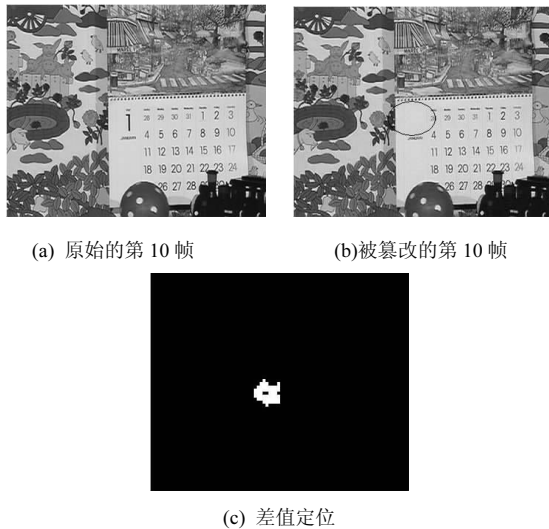


图 4 视频帧篡改区域定位

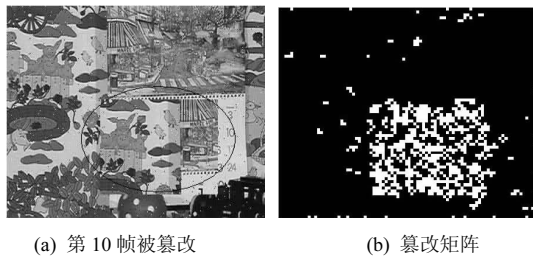


图 5 Mobile 序列第 10 帧加椒盐噪声及复制粘贴篡改定位

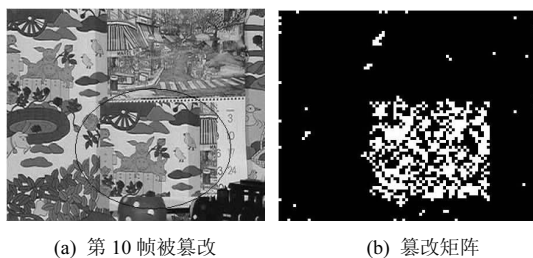


图 6 Mobile 序列第 10 帧加高斯噪声及复制粘贴篡改定位

表 4 篡改 20%，QP=26 重编码水印性能

检测参数	帧序号						
	10	20	30	40	50	60	70
P_{FN}	2.74	2.52	2.43	3.11	2.75	2.73	2.39
P_{FP}	5.17	4.63	5.38	5.35	5.43	5.56	5.98
P_D	95.89	96.14	96.79	95.43	95.68	95.24	95.70

表 5 篡改 20%，加 0.1%椒盐噪声水印性能

检测参数	帧序号						
	10	20	30	40	50	60	70
P_{FN}	2.79	2.21	2.62	2.26	2.71	2.89	2.13
P_{FP}	6.13	6.12	5.62	6.74	6.63	6.87	6.07
P_D	95.24	95.64	95.77	95.89	95.15	95.64	95.69

表 6 篡改 20%，加 0.01%高斯噪声水印性能

检测参数	帧序号						
	10	20	30	40	50	60	70
P_{FN}	7.76	8.25	8.43	8.62	8.05	8.83	8.24
P_{FP}	3.56	3.79	3.64	3.28	4.53	3.74	4.32
P_D	94.43	94.18	93.92	93.79	93.86	93.85	93.37

图 5(a)为 Mobile 序列第 10 帧同时添加椒盐噪声及复制粘贴的图，图 5(b)为该帧篡改定位图。图 6(a)为 Mobile 序列第 10 帧同时添加高斯噪声及复制粘贴的图，图 6(b)为该帧遭篡改后的定位图，从图中可以看出，当内容被恶意篡改时，水印差值图像的“1”分布比较集中，而由于噪声引起的“1”比较分散和随机，通过“1”的分布可以有效定位篡改区域。

本文算法能有效判断内容被修改的类型及能有效定位遭恶意篡改的内容，主要原因是提取的特征码和嵌入的半脆弱水印算法具有良好的顽健性。如果提取的特征码或嵌入的水印算法无法抵抗重编码等“内容保持”类攻击，则水印差值图像中“1”的分布会比较集中，这会增大误检率，影响定位的准确率。

6 结束语

本文提出了一种基于代价策略的 H.264/AVC 的半脆弱水印算法。首先，提出了一种新的特征码生成算法，通过置乱的方式消除特征码之间的相关性以提高其顽健性。其次，根据半脆弱水印的特性要求及结合 H.264/AVC 的编解码特性，提出了一种根据最小代价策略选择最优水印嵌入位置的算法。实验结果表明：特征码具有较好的顽健性；采用代价策略函数可以较明显地降低嵌入水印对视频质量和码率的影响；半脆弱水印算法能有效区别内容

修改类型, 并能对恶意篡改区域进行精确定位。

下一步的研究工作将致力于使半脆弱水印算法能有效抵抗时域的恶意篡改攻击。

参考文献:

- [1] MEHDI F, SHERVIN S, MEHDI S, *et al.* Tampering detection in compressed digital video using watermarking[J]. *IEEE Transactions on Instrumentation and Measurement*, 2014, 63(5):1057-1072.
- [2] QUELUZ M P. Authentication of digital images and video: generic models and a new contribution[J]. *Signal Processing: Image Communication*, 2001, 16(5):461-475.
- [3] DOERR G, DUGELAY J L. A guide tour of video watermarking[J]. *Signal Processing: Image Communication*, 2003, 18(4):263-282.
- [4] ZHANG X P, WANG S Z, QIAN Z X, *et al.* Reference sharing mechanism for watermark self-embedding[J]. *IEEE Transactions on Image Processing*, 2011, 20(2):485-495.
- [5] HE H J, CHEN F, TAI H M, *et al.* Performance analysis of a block-neighborhood-based self-recovery fragile watermarking scheme[J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(1):185-196.
- [6] PIPER A, SAFAVI-NAINI R. Scalable fragile watermarking for image authentication[J]. *IET Information Security*, 2013, 7(4):300-311.
- [7] CHEN F, HE H J, TAI H M, *et al.* Chaos-based self-embedding fragile watermarking with flexible watermark payload[J]. *Multimedia Tools and Applications*, 2014, 72(1):41-56.
- [8] CHEN S, LEUNG H. Chaotic watermarking for video authentication in surveillance applications[J]. *IEEE Transactions on Circuits and System for Video Technology*, 2008, 18(5):704-709.
- [9] SHI Y J, QI M, YI Y G, *et al.* Object based dual watermarking for video authentication[J]. *Optik*, 2013, 124:3827-3834.
- [10] MASOUMI M, AMISI S. Content protection in video data based on robust digital watermarking resistant to intentional and unintentional attacks[J]. *The International Arab Journal of Information Technology*, 2012, 11(2):204-212.
- [11] ISO/IEC 14496-10, ITU-T Rec.H.264. Advanced video coding[S]. 2003.
- [12] HOROWITZ M, JOCH A, KOSENTINI F, *et al.* H.264/AVC baseline profile decoder complexity analysis[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2003, 13(7):704-716.
- [13] HORNG S J, FARFOURA M E, FAN P Z, *et al.* A low cost fragile watermarking scheme in H.264/AVC compressed domain[J]. *Multimedia Tools and Applications*, 2014, 72(3):2469-2495.
- [14] KIM T, PARK K, HONG Y. Video watermarking technique for H.264/AVC[J]. *Optical Engineering*, 2012, 51(4): 047402.
- [15] LI Q, WANG R D, XU D W. Fragile video watermarking algorithm based on H.264/AVC[J]. *Materials Research Innovations*, 2013, 17: 94-100.
- [16] XU D W, WANG R D. Watermarking in H.264/AVC compressed domain using Exp-Golomb code words mapping[J]. *Optical Engineering*, 2011, 50(9): 097402.
- [17] CHEN T Y, CHEN T H, LIN Y T, *et al.* H.264 video authentication based on semi-fragile watermarking[A]. *International Conference on Intelligent Information Hiding and Multimedia Signal Processing[C]*. 2008.659-662.
- [18] 林志高, 孙锬锋, 蒋兴浩. 基于 VLC 域的 H.264/AVC 视频流内容级认证水印算法[J]. *上海交通大学学报*, 2011, 45(10):1531-1535.
LIN Z G, SUN T F, JIANG X H. A content level watermarking scheme for H.264/AVC video authentication in VLC domain[J]. *Journal of Shanghai Jiaotong University*, 2011, 45(10):1531-1535.
- [19] 王小静, 杨高波, 朱宁波. 基于半脆弱水印的 H.264/AVC 视频流的内容级认证[J]. *通信学报*, 2009, 30(11):71-78.
WANG X J, YANG G B, ZHU N B. Content authentication of H.264/AVC video stream based on semi-fragile watermarking[J]. *Journal of Communications*, 2009, 30(11):71-78.
- [20] XU D W, WANG R D, WANG J C. A novel watermarking scheme for H.264/AVC video authentication[J]. *Signal Processing: Image Communication*, 2011, 26:267-279.
- [21] LIN C Y, CHANG S F. Semi-fragile watermarking for authentication JPEG visual content[A]. *Proceedings of SPIE Conference on Security and Watermarking of Multimedia Contents[C]*. 2000.140-151.
- [22] LIN C Y, CHANG S F. Issues and solutions for authenticating MPEG video[A]. *Proceedings of the SPIE Security and Watermarking of Multimedia Contents[C]*. San Jose, CA, 1999.54-65.
- [23] KIM D W, CHOI Y G, KIM H S, *et al.* The problems in digital watermarking into intra-frames of H.264/AVC[J]. *Image and Vision Computing*, 2010, 28: 1220-1228.
- [24] SHAHID Z, CHAUMONT M, PUECH W. Considering the reconstruction loop for data hiding of intra-and inter-frames of H.264/AVC[J]. *Signal Image and Video Processing*, 2013, 7(1): 75-93.
- [25] LIN E T, DELP E J. A review of fragile watermarkings[A]. *Proceedings of the ACM Multimedia and Security Workshop[C]*. Orlando, 1999.25-29.
- [26] FRIDRICH J. Methods for tamper detection in digital images[A]. *Proceedings of the ACM Workshop on Multimedia and Security[C]*. Orlando, 1999.19-23.
- [27] ZHANG J, HO A T S, QIU G, *et al.* Robust video watermarking of H.264/AVC[J]. *IEEE Transactions on Circuits and Systems-II: Express Briefs*, 2007, 54(2):205-209.
- [28] HU J Q, HUANG J W, HUANG D R, *et al.* Image fragile watermarking based on fusion of multi-resolution tamper detection[J]. *Electronics Letters*, 2002, 38(24):1512-1513.
- [29] 王向阳, 陈利科. 一种新的自适应半脆弱水印算法[J]. *自动化学报*, 2007, 33(4):361-366.
WANG X Y, CHEN L K. A novel adaptive semi-fragile watermarking scheme based on image content[J]. *Acta Automatica Sinica*, 2007, 33(4):361-366.
- [30] 段贵多, 赵希, 李建平等. 一种新颖的用于图像内容认证、定位和恢复的半脆弱数字水印算法研究[J]. *电子学报*, 2010, 38(4): 842-847.
DUAN G D, ZHAO X, LI J P, *et al.* A novel semi-fragile digital watermarking algorithm for image content authentication, localization and recovery[J]. *Acta Electronica Sinica*, 2010, 38(4): 842-847.

作者简介:



张维纬 (1982-), 男, 福建泉州人, 博士, 华侨大学讲师、硕士生导师, 主要研究方向为数字版权管理与数字水印、大数据与云计算安全等。

赵宸 (1985-), 男, 山东青岛人, 博士, 国家计算机网络与信息安全管理中心助理研究员, 主要研究方向为网络安全等。

黄德天 (1985-), 男, 福建龙岩人, 博士, 华侨大学讲师、硕士生导师, 主要研究方向为机器视觉、图像视频处理等。

张沛 (1982-), 男, 湖南株洲人, 北京邮电大学博士生, 主要研究方向为数字水印、知识可视化、通信导航等。

杨义先 (1961-), 男, 四川绵阳人, 北京邮电大学教授、博士生导师, 主要研究方向为网络与信息安全、密码学、编码理论等。