一种子块相关的半脆弱数字水印

王 勋 凌 云
(浙江工商大学计算机与信息工程学院,杭州 310035)

摘 要 数字水印是多媒体数字产品版权保护和内容完整性认证的重要技术之一。为得到一种鲁棒性更好的数字水印,在分析 MHB 算法缺陷的基础上,采用认证码生成块与嵌入块及嵌入块之间互相关的策略,提出了一种更加安全的子块相关的半脆弱数字水印算法,并增强了算法抵抗恶意攻击的能力。实验结果表明,该算法在不破坏宿主图像视觉质量基础上,对 JPEG 有损压缩具有较强的鲁棒性,并可有效地检测出对图像内容的局部恶意篡改。

关键词 半脆弱性水印 DCT 变换 JPEG 压缩
中图法分类号: TP309.7 **文献标识码**: A **文章编号**: 1006-8961(2007)05-0826-05

A Semi-fragile Watermarking with Blocks Relativity

WANG Xun, LING Yun
(*College of Computer & Information Engineering, Zhejiang Gongshang University, Hangzhou* 310035)

Abstract Digital watermarking is a important technology of multimedia digital product's copyright protection and content authentication. Based on through analysis of the fragile watermark scheme presented by Marvel *et al.*, a novel safer semi-fragile image digital watermarking algorithm with blocks relativity is proposed in this paper. Simulation experiments show that it will not degrade the visual quality of original images, and maintain certain robustness for JPEG compression. Besides the scheme can be used to detect the counterfeiting attack to specific local content.

Keywords semi-fragile watermarking, DCT transform, JPEG compression

1 引 言

随着 Internet 迅速普及,基于 Web 服务的各种应用日益增长,使多媒体信息的交流达到前所未有的广度和深度。然而,多媒体信息在传输过程中会遭受各种无意或有意的篡改攻击,使得人们对数字媒体(数字图像、数字音频、数字视频)的完整性和内容的真实性产生质疑。因此,如何在 Internet 网络环境中,对数字媒体内容的真实性、完整性实施有效保护已成为一个严峻的现实问题。

从信息隐藏技术发展而来的脆弱性数字水印技术为人们提供了有效的解决途径。它的基本措施就是在数字声音、图像、文档或者视频流中嵌入可见或不可见的水印信息,通过对水印检测来鉴别数字产品的真实性和完整性[1,2],从而成为保护数字产品版权的可靠手段。

关于多媒体内容的完整性认证,尤其是图像的完整性验证,是近几年来随着网络技术的发展而产生并发展起来的,当前的解决方案主要集中在脆弱性数字水印技术。脆弱水印作为一项新型数字媒体认证技术,这几年得到长足发展,但随着研究的深入,脆弱性水印的研究热点已经转向在对特定修改具有鲁棒性,但对其他恶意修改具有较强敏感性的半脆弱性数字水印[3~6],并且逐步从静止图象扩展到数字音频和视频等领域。由于图像信息在数字媒体信息中具有代表性,因此本文将以图像为例来研究半脆弱水印。

基金项目:浙江省自然科学基金项目(Y105579);浙江省教育厅重点资助项目(20050642)
收稿日期:2004-10-27;改回日期:2005-12-10

第一作者简介:王 勋(1967~),男。浙江大学 CAD&CG 国家重点实验室 2002 春博士研究生,浙江工商大学副教授。主要从事多媒体数字水印、虚拟现实、智能信息处理和 GIS 研究。E-mail: wx@mail.zjgsu.edu.cn

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[3],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[4],但同样存在 JPEG 压缩不够稳健,且阈值选择困难等缺点。Lin 提出一种基于图像子块 DCT 系数的水印嵌入算法[5],该算法能有效地区分 JPEG 压缩和其他恶意篡改,但在系统安全性方面存在缺陷。Holliman 对基于块方式的系列算法进行了研究,指出了基于块的水印方案存在安全问题,并给出了可能的解决方案[6]。在对文献[3~8]中提出的算法详细分析基础上,采用认证码生成块与嵌入块及嵌入块之间互相关的策略,提出一种在 DCT 域嵌入的半脆弱性水印算法。该算法可以有效地进行数字图像的真实性、完整性认证,并能抵抗 JPEG 压缩。

2　抗 JPEG 压缩的半脆弱水印算法

半脆弱水印对特定操作的鲁棒性和对其他恶意篡改的敏感性,一方面能保证用户对多媒体信息的合法操作,如为方便网上传输需对图像实行 JPEG 有损压缩,用户对经过 JPEG 有损压缩后的图像内容同样视为真实有效,且提取出的水印准确无误,另一方面对其他的恶意攻击却具有敏感性。

2.1　基于 DCT 变换的 JPEG 压缩算法

JPEG 算法[9]是联合图像专家组提出的一种静态图像压缩标准,在网络上应用甚为广泛。其算法流程大致如图 1 所示。

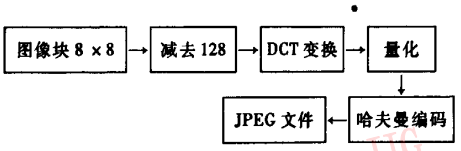


图 1　基于 DCT 变换的 JPEG 压缩算法

Fig. 1　DCT based JPEG algorithm

假设有 $m \times n$ 大小的图像 I(m 和 n 都是 8 的倍数),把 I 分为互不重叠的 8 × 8 子块,再对每块分别进行 DCT 变换,把数据从空间域变换到频率域,然后对 DCT 系数进行量化,最后通过编码器对数据最小化。

从信号处理角度来看,JPEG 算法的目的在于:把大量的数据简化为较小的、真正有意义的数据,删除只带极少视觉效果的信息,并且利用数据的空间特性进一步压缩数据。其解压缩过程为压缩的逆过程。

2.2　MHB 算法分析

在脆弱性水印的嵌入算法中,首先要做的是对原始图像进行特征提取,然后通过一个可靠的方法把水印信息隐藏到图像中去,得到添加水印后的图像。在许多实际的应用场合中,往往需要水印能够抵抗一定程度的有益的数字信号处理操作,如 JPEG 压缩等。在 JPEG 压缩算法中,DCT 变换可以认为基本上是无损的,真正使数据受损的是量化步骤。许多抗 JPEG 压缩的半脆弱水印算法就是抓住这个特点,使嵌入的水印即具有一般脆弱水印的特性而又有抗 JPEG 压缩的效果。

Marvel 等提出了一种抗 JPEG 压缩的 MHB 算法[7],MHB 算法的基本原理是将原图像分成 N 个 8 × 8 的 DCT 块,选取其中 H 个块作为水印的嵌入载体,利用剩下的块图像信息并通过 hash 函数产生一个特定长度的认证码(message authentication code, MAC),然后用 Stego-JPEG 算法[7]依次嵌入到上述选取的水印载体中。其中 Stego-JPEG 嵌入算法具体过程如下:

(1) 对每个选取的 DCT 图像块量化后的系数进行求和,即 $\Delta = \sum_{i,j=1}^{8} C_{i,j}$,其中 $C_{i,j}$ 为图像量化后的系数;(2) 如果 Δ 的奇偶性与嵌入信息相符(如奇代表 1,偶代表 0),则保持图像块系数不变;(3) 如果 Δ 的奇偶性与嵌入信息不符,则调整某个系数 $C_{i,j}$,使其为 $\hat{C}_{i,j} = C_{i,j} \pm 1$。提取时仅需判断量化后的 DCT 块系数之和的奇偶性即可得到水印信息。当检测图像是否被篡改时,只要通过比较计算出的 MAC 和嵌入块中提取的信息串,相同则认为认证成功,否则为受攻击篡改。

由于量化是 JPEG 压缩方法中造成数据损失的原因,而 MHB 算法是在量化后的 DCT 系数中嵌入水印信息的,因此对抵抗 JPEG 压缩具有较好的效果。但仔细分析后会发现,MHB 算法的安全性存在如下严重缺陷:(1) MHB 算法中 MAC 是直接用像素值作为 hash 函数的输入生成的。然而经过 DCT 变换并量化后,图像信息必定有所改变,也就是说,作为 hash 函数的输入就会发生改变,这样就不可能得到初始生成的 MAC;(2) MHB 算法中 MAC 的产生部分和嵌入部分是完全分离的,这使得嵌入部分的改变不影响认证结果,也就是说,一旦攻击者篡改了嵌入块,并保持了原来嵌入算法的特性,就能保证产生的 MAC 和提取的 MAC 完全吻合,这说明图像篡改成功;(3) 嵌入过程按比特独立进行,各比特的

嵌入互不相干,从而使得攻击者可以改变一个嵌入块而不影响其他的嵌入块。

在分析了 MHB 算法安全缺陷的基础上,Li 等人对其进行了如下两点修改[8]:(1)把 DCT 量化后的数据作为 hash 的输入;(2)认证码 MAC 的嵌入部分和生成部分不严格的分开,把嵌入部分的某些信息同时作为 hash 函数的输入以生成 MAC。经以上修正后,MHB 算法的安全性得到较大提高。

除了 Li 等人的改进外,笔者发现 MHB 算法嵌入块之间的比特信息是相互独立嵌入的,彼此之间并没有联系。假如在算法透明的情况下,攻击者可以通过修改一个嵌入块而不影响其他的嵌入块,从而并不影响整个水印信息的提取结果。所以在攻击者已知水印信息 W 或拥有多幅隐含相同的水印的可信图像时,很容易在保持提取水印不变的情况下,恶意篡改图像。

以上嵌入子块之间相互独立的子块不相关水印嵌入算法之所以易受攻击,原因在于:每一个子块水印信息的隐藏和提取与其他子块并不相关,一个子块的改变并不影响其他子块水印信息的提取结果。

由于 JPEG 压缩也会造成图像的损失,DCT 系数必定有所改变,使得压缩后的水印图像不能完全被检测通过。同时,子块不相关水印嵌入法中,可以让攻击者改变局部块的系数实现对图像进行篡改,因为各子块之间水印的嵌入是相互独立的,这种方法破坏水印信息(提取部分)而篡改成功的概率很大;就算总体有差错,但也不会很明显,如果控制在 JPEG 压缩后的范围内,由于很难判断是受到攻击还是本身 JPEG 压缩的原因,这就提高了该算法漏检的可能性。

为了提高算法的抵抗攻击能力,区分出正常的改变和恶意攻击,本文把 MAC 的嵌入看作是一个整体嵌入,增加子块之间的相关性,一旦某一子块信息被篡改就无法正确提取其他块的信息。

3 子块相关的半脆弱水印算法

3.1 子块相关算法基本思想

根据上节的研究思路,提出一种采用子块相关的嵌入方法,其算法基本思想如下:

(1) 把图像像素分块后进行 DCT 变换和量化处理,得到 DCT 系数,记做 C,把他作为 hash 函数的输入源;

(2) 为了保证 hash 函数输入部分的一致性,本文在选取 H 个待嵌入块后,对这些 DCT 块进行预嵌入处理。预嵌入处理是指:假设这些待嵌入块每一块都要进行修改,则利用 Stego-JPEG 算法在每块数据中找出一个系数作为修改对象;接着把除了这些需要修改的系数外的所有其他系数组成 hash 函数的输入,从而得到可靠的 MAC,如果其长度不够,可以再次使用 hash 函数,即将上次的输出作为下次的 hash 函数的输入;

(3) 把 MAC 认证信息转换成二进制的比特流,如 $w=\{w_1,w_2,\cdots,w_n\}$,其中 n 为水印信息的长度,即本文上面选取的 DCT 待嵌入块个数;

(4) 要使嵌入方法具有块相关性,嵌入策略如下:第 1 位水印信息按照原方法嵌入到第 1 块 C_0 中,然后计算当前块 C_i 和前一块 C_{i-1} 的状态 t,其关系满足:if $\theta(C_i)>\theta(C_{i-1})$ 则状态 $t=1$,否则 $t=0$,其中 $\theta(C_i)$ 可以是子块 C_i 的均值或方差,也可以是由 C_i 决定的其他特征值;最后计算当前嵌入的位信息值 $\hat{w}_i=w_i\oplus t_i$,其中,$i\in[1,h]$,嵌入信息过程按照 Stego-JPEG 算法。

3.2 水印嵌入算法

综上所述,为保证水印信息的唯一性,采取 DCT 系数作为 hash 函数的输入;然后采用块相关性的嵌入方法嵌入水印信息。具体嵌入算法如下:

(1)对图像进行分块、DCT 变换和量化;

(2)根据 key 随机选出 H 个块,并确定其顺序;

(3)对随机选取的 H 个块的 DCT 系数做嵌入的预处理,即可得到 H 个可能被修改的系数,并把原来的所有 DCT 系数组成新的数据(data),但是不包括从预处理得到的 H 个系数值;

(4)计算 $MAC(MAC=hash(key,data))$ 时,本文采用 MD5 算法[7]。并把 MAC 转换成二进制比特流 $w=\{w_1,w_2,\cdots,w_H\}$,长度为 H;

(5)初始化子块之间相关性的状态 t_0 为 0(为了保证嵌入的第 1 位保持原有的不变),其他 t_i 根据子块之间的特征值的大小比较得到:

$$\text{if}\quad \theta(C_i)>\theta(C_{i-1})\quad t_i=1$$

$$\text{else}\quad t_i=0$$

(6)将步骤(5)中得到的 H 个比特信息和状态进行异或操作,得到相应的 $\hat{w}_i=w_i\oplus t_i$,其中,$i\in\{1,\cdots,H\}$;

(7)按照 Stego-JPEG 算法嵌入水印信息 $\hat{w}$。

3.3 水印的提取和检测

水印提取时,分为两个阶段:(1)求水印的认证

码 *MAC*;(2)提取水印图像的水印信息。前者类似于嵌入方法,把 *data* 组织好以后在密钥 key 控制下通过 hash 函数生成;后者先根据 Stego-JPEG 算法求出直接的序列 $w=\{w_1,w_2,\cdots,w_H\}$,然后算出状态序列 t,再进行异或即得到水印信息,即本文所说的 MAC。

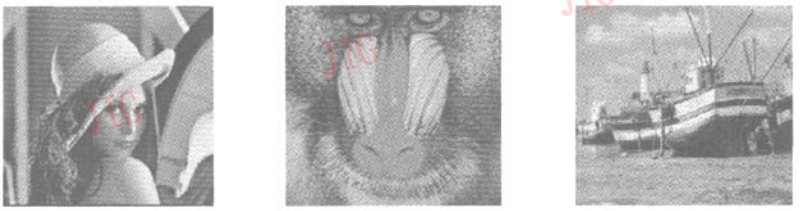
检测水印只要把计算出的 MAC 与随机选取的 H 个块中提取的信息进行比较,判断其是否被篡改。本文采用篡改估计函数(tamper assessment function, TAF)来度量提取水印与原始水印的相似程度:

$$TAF(w,\hat{w})=\frac{1}{N}\sum_{1}^{n}w\oplus\hat{w}$$

这里⊕代表异或,n 是水印序列长度,w 是生成的 MAC 认证信息,$\hat{w}$ 是提取出的水印信息。如果提取的水印序列与原序列的 TAF 值高于某个阈值(实验结果表明,TAF 值取 0.35 比较合理),则认为图像受到篡改,水印遭到破坏;否则认为图像完好。

4　实验结果与讨论

为了证实子块相关算法的有效性,实验采用 256×256 的 8bits 灰度图 Lena,Baboon 和 Ship 作为测试对象,原图如图 2 所示。按上面嵌入算法得如图 3 所示嵌入水印后的图像,经计算,图 3 中图像的 PSNR 在 46~48dB 之间,远远大于一般意义上的最低限度 28dB,表明经嵌入水印后图像与原图非常相似,同时,在视觉上嵌入水印后的图像质量也无明显下降,表明该水印算法具有很好的透明性。



(a) Lena　　(b) Baboon　　(c) Ship

图 2　原始图像

Fig. 2　Original image

(a) Lena (*PSNR*=48.204dB)　　(b) Baboon (*PSNR*=47.948dB)　　(c) Ship (*PSNR*=46.205dB)

图 3　嵌入水印后的图像

Fig. 3　Image embedded watermarking

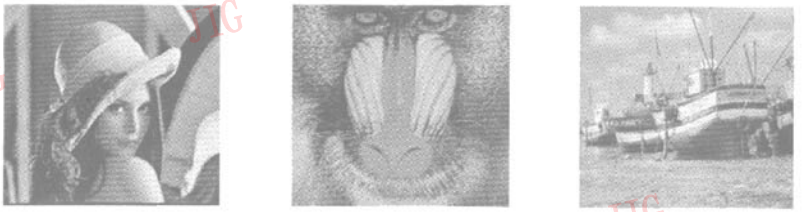
图 4 为嵌入水印的图像经 JPEG 压缩后的图像,其中压缩因子 $Q=50$,计算得到的图像的峰值信噪比和篡改估计函数值如表 1 所示。表中数据表明,经压缩后图像的信噪比明显下降,但篡改估计函数值没有超过 0.2,水印信息仍能够正常提取,这说明该算法能够抵抗 JPEG 压缩。

表 1　经 JPEG 压缩的结果

Tab. 1　The result after JPEG compression

	PSNR(dB)			*TAF*		
	Lena	Baboon	Ship	Lena	Baboon	Ship
JPEG 压缩	34.108	35.397	40.832	0.050781	0.109375	0.160156

为测试子块相关算法对其他恶意攻击的脆弱性,通过引入 3×3 中值滤波、图像平滑和一般篡改操作来对算法进行测试,表 2 是对各水印图像攻击后的检测结果。测试数据显示经这类攻击后,峰值信噪比(peak signal noise ratio, PSNR)都大幅度地下降,其中缩放处理后图像峰值信噪比下降最为明



(a) Lena (b) Baboon (c) Ship

图 4 JPEG 压缩因子 $Q=50$ 时嵌入水印后的图像

Fig. 4 Image embedded watermarking with JPEG compression rate $Q=50$

表 2 各类攻击的检测结果

Tab. 2 Attack detection result

	PSNR(dB)			*TAF*		
	Lena	Baboon	Ship	Lena	Baboon	Ship
中值滤波 3×3	29.566	24.190	30.972	0.547	0.527	0.555
图像平滑 3×3	28.933	24.602	30.958	0.387	0.352	0.346
随机噪音	26.612	27.414	27.242	0.453	0.461	0.551
缩放处理	20.742	20.733	24.250	0.508	0.426	0.445

显;篡改估计函数值明显提高,只有 Ship 图像的平滑攻击下接近 0.35,其他全部超过了 0.35 的阈值,这表明块相关算法对恶意攻击具有良好的脆弱性。

关于算法的安全性,由于攻击者不知道密钥,因此选择任意一块进行篡改,其成功的概率分析如下。当选择的块属于所有 N 个块中的 H 个块时,攻击不考虑水印嵌入算法中 Δ 的奇偶性,MHB 算法篡改成功的概率为 $\frac{H}{2N}$;本文算法由于考虑了子块的相关性,按平均计算,一个子块的改变不导致前后其余子块变化的概率是 1/4,因此我们算法篡改的成功概率为 $\frac{1}{4}\times\frac{H}{2N}$。当选择的块不属于 H 个块时(这种概率接近于 1,因为 N 远大于 H,$\frac{N-H}{N}\approx 1$),平均意义下 MAC 中有一半会被篡改,在不知道密钥的情况下,篡改图像通过认证的概率为 2^{-H}。对于本文的方法,由于 H 位中某一位的改变有 1/4 的可能不引起前后位的改变,因此篡改成功的概率为 2^{-4H}。可见,不管在哪种情况下,本文的算法都提高了算法的安全性。

5 结 论

通过对文献[7]中的 MHB 算法详细分析,发现 MHB 算法存在安全隐患,采用认证码生成块与嵌入块及嵌入块之间互相关的策略,提出一种嵌入子块相关性半脆弱水印算法。该算法不仅在安全性方面有所提高,而且实验显示对 JPEG 有损压缩具有较强的顽健性,对中值滤波、图像平滑及其他恶意攻击操作等具有较高的敏感性,可用于数字图像的真实性和内容完整性认证。

参考文献(References)

1 Wong P W. A public key watermark for image verification and authentication[A]. In: Proceedings of IEEE International Conference on Image Processing[C], Chicago, IL, USA, 1998: 425~429.

2 Yeung M, Mintzer F. An invisible watermarking technigue for image verification[A]. In: Proceedings of International Conference on Image Processing[C], Barbara, California, USA, 1997: 680~683.

3 Fridrich J. Image Watermarking for Tamper Detection[A]. In: Proceedings of the International Conference on Image Processing[C], Chicago, IL, USA, 1998: 404~408.

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

5 Lin Ching-yung, Chang Shihfu. Semi-fragile watermarking for authenticating JPEG visual content[A]. In: Proceedings of SPIE Security and Watermarking of Multimedia Contents, EI'00[C], San Jose, CA, USA, 2000: 140~151.

6 Holliman M, Memon N. Counterfeiting attacks on oblivious block-wise independent invisible watermark schemes[J]. IEEE Transactions on Image Processing, 2000, 19(3): 432~441.

7 Marvel L M, Hartwig G, Boncelet C. Compression-compatible fragile and semi-fragile tamper detection[A]. In: Proceedings of SPI[C], Washington, USA, 2000: 131~139.

8 Li De-quan, Su Pu-rui, Feng Deng-guo. A DCT-based multimedia authentication scheme[J]. Journal of Computer Research and DevelopmentM 2002, 39(6): 678~683.[李德全, 苏璞睿, 冯登国. 一个基 DCT 压缩编码的多媒体认证系统[J]. 计算机研究与发展, 2002, 39(6): 678~683.]

9 Gregory K. Wallace. The JPEG Still Picture Compression Standard [J]. Communications of the ACM, 1991, 39(4): 30~44.