

文章编号: 1004-4353(2020)01-0043-06

基于半脆弱水印的图像内容认证技术

杨珊, 李德*

(延边大学 工学院, 吉林 延吉 133002)

摘要: 为了认证图像内容的真实性和可靠性,提出了一种用于图像内容认证和篡改检测定位的半脆弱水印算法.首先利用小波变换的系数分量近似生成表征图像内容的特征水印;然后根据图像在小波变换后的能量分布特性,利用伪随机数发生系统确定特征水印的嵌入位置,以此实现水印的自适应嵌入.最后在小波系数中提取出嵌入的水印信息,并与图像生成的水印信息进行比对,以此实现图像的认证功能.实验结果显示,该算法能够区分常规信号的处理操作(如进行高质量的 JPEG 压缩、添加少量噪声)和恶意篡改操作(如图像剪切),并可对图像的篡改位置进行定位.

关键词: 半脆弱水印; 内容认证; 离散小波变换(DWT); 自适应嵌入; 篡改定位

中图分类号: TP309.7

文献标志码: A

Image content authentication technology based on semi-fragile watermarking

YANG Shan, LI De*

(College of Engineering, Yanbian University, Yanji 133002, China)

Abstract: In order to authenticate the authenticity and reliability of image content, a semi-fragile watermarking algorithm for image content authentication and tamper detection and localization is proposed. Firstly, the wavelet transform is used to approximate the watermarking of image content. Then, according to the energy distribution of the image after wavelet transformation, the pseudo-random number generation system is used to determine the embedding position of the feature watermarking, so as to realize the adaptive embedding of the watermarking. Finally, the embedded watermarking information is extracted from the wavelet coefficients and compared with the watermarking information generated by the image to realize the image authentication function. Experimental results show that the algorithm can distinguish between conventional signal processing operations (such as high quality JPEG compression, adding a small amount of noise) and malicious tampering operations (such as image cutting), and can locate the tampering position of the image.

Keywords: semi-fragile; content authentication; DWT transform; adaptive embedding; tamper location

0 引言

随着信息技术的快速发展,多媒体数据因便于存储、传输和分享,给人们获取和利用信息带来了极大便利.但同时,因多媒体数据存在容易被复制、修改以及难以追踪修改源等问题使得其安全问题受到人们的关注,如军事文件、法庭举证、新闻图像、历史文献、医疗图像等数字内容被非法篡改后就会给社会造成诸多影响;因此,在海量的多媒体数据环境中如何保护好图像内容的真实性、完整性和可靠性具

收稿日期: 2019-06-10

基金项目: 吉林省教育厅“十三五”科学技术研究项目(吉教科合字[2016]第 249 号)

* 通信作者: 李德(1972—),男,博士,教授,研究方向为数字内容安全、数字版权保护.

有重要意义. 研究表明, 水印技术是保护多媒体数据的一种有效手段, 水印按特性可以分为完全脆弱水印^[1]和半脆弱水印^[2]. 完全脆弱水印因对图像验证的精确度要求较高, 因此其应用范围较为有限, 如医学图像、法庭举证等方面; 半脆弱水印允许图像内容在一定范围内发生变化, 因此其在实际应用中更为广泛. 近年来, 许多学者对半脆弱水印技术进行了研究. 例如: 胡玉平等^[3]提出了一种基于小波系数形态滤波的半脆弱水印算法, 该方法可提高算法的安全性并减少运算量, 但其抗 JPEG 压缩效果较差. 赵春晖等^[4]提出了一种基于分块压缩感知的半脆弱零水印算法, 该算法计算简单, 保密性强, 实现了图像成像与水印生成的同步完成. 高铁杠等^[5]提出了一种利用图像块与 Logistics 混沌映射匹配的方法, 该算法在图像认证功能和检测恶意篡改方面表现良好, 可提高模糊、锐化等信号攻击的鲁棒性. 王向阳等^[6-7]提出了一种基于分组量化的自适应半脆弱水印嵌入方案, 该算法在误检率和漏检率等方面优于现有的半脆弱水印算法. 高雪丽等^[8]采用二维离散余弦变换提取低频系数的方法提高了图像在抵抗高斯噪声、裁剪攻击、压缩攻击方面的鲁棒性, 该方法同时还能够抵抗一定程度的帧丢失攻击. 上述研究中, 研究者都是将水印信息以固定的量化步长嵌入到图像中, 因此会影响图像的视觉效果. 为此, 本文提出一种基于变换域的半脆弱水印算法, 采用自适应量化步长的嵌入方法来提高水印信息的不可见性, 并通过实验验证本文方法的有效性.

1 相关理论与技术

1.1 半脆弱水印技术

半脆弱水印技术是利用图像的自身特征提取出具有特定含义的认证信息, 然后使用一定的方法和规则将这些特定含义的认证信息嵌入到被保护或需要提供版权认证的媒体数据中, 以此实现防伪溯源^[9]. 半脆弱数字水印技术应具有以下功能和特征:

1) 检测篡改及恢复功能. 半脆弱数字水印技术要对载体图像能够进行可靠地检测篡改, 且在此基础上还要能够检测出图像被篡改的位置和相应的篡改程度, 以及恢复被篡改的内容.

2) 水印的盲提取功能. 在不提供原始载体的情况下, 半脆弱水印技术要能够实现对待检测信息内容的认证, 即实现水印的盲提取功能.

3) 鲁棒性与脆弱性共存. 半脆弱水印技术应对图像的合理失真具有很强的鲁棒性, 同时对不合理的失真(恶意篡改)应具有脆弱性.

4) 透明性. 水印信息嵌入到原始载体图像可能会影响多媒体文件的视觉质量, 因此半脆弱水印技术在确定秘密水印的嵌入位置和方式时, 不仅需要满足算法的抗攻击性和定位准确性, 还要保证载体图像在视觉上具有不可感知性.

1.2 线性同余发生器

为了保证载体图像具有较好的视觉效果和水印信息的透明度, 即降低由嵌入水印信息而导致的图像失真对图像视觉质量的影响, 本文在选择水印嵌入位置时, 采用线性同余伪随机数发生器将水印信息随机嵌入到小波系数的高低频分量和低高频分量中, 以此达到分散水印信息的目的. 线性同余伪随机数发生器产生的是均匀分布的伪随机数, 其数学表达式为

$$x_{i+1} = (ax_i + c) \pmod{m}, \quad (1)$$

其中模 m 为素数, 乘数 a 为 m 的一个原根, 增量 $c \in [0, m)$, $m \in [0, \infty)$. 理论上伪随机数的周期 T 越趋近于无穷, 伪随机数的效果越好. 当 $c = 0$ 时, 可以得到线性同余伪随机数发生器的一般通用式(线性同余发生器):

$$x_{i+1} = ax_i \pmod{m}. \quad (2)$$

2 基于图像内容的小波域自适应量化水印算法

基于图像内容的小波域自适应量化水印算法主要包括水印信息的生成、嵌入和提取 3 部分,其流程如图 1 所示.

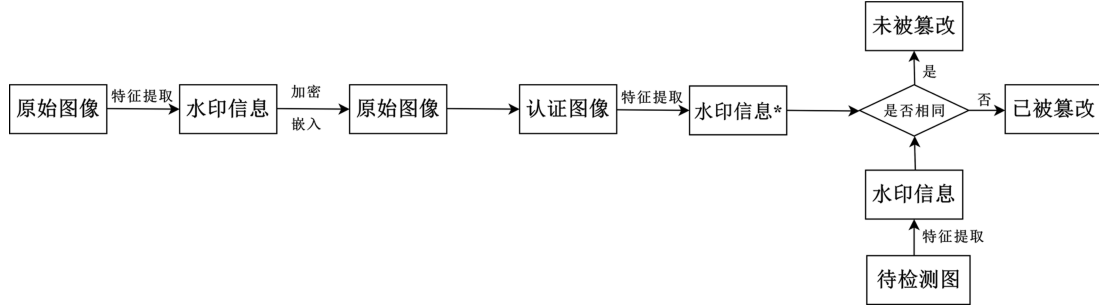


图 1 水印认证的流程图

2.1 水印信息的生成算法

小波变换的主要优点是能够提供丰富的图像特征信息,尤其是在水印信息提取方面^[10]. 研究^[11]表明,第 3 层小波分解的系数分量 $LL3(i, j)$ 含有图像的大部分能量,能有效表征图像特征,且在 JPEG 压缩前后具有一定的鲁棒性和很强的抗干扰能力,因此本算法在第 3 层小波系数中提取图像的特征水印. 生成水印信息的步骤如下:

Step1 对载体图像 I (灰度图像) 进行 3 层 DWT 变换,图像像素大小为 $M \times N$,小波系数 Z 大小为 $M/8 \times N/8$.

Step2 提取小波系数分量 $LL3(i, j)$ 的二进制的最高位,并将其作为载体的特征信息,以此生成大小为 $M/8 \times N/8$ 的待嵌入水印 W .

2.2 水印信息的嵌入算法

水印信息嵌入算法的具体步骤如下:

Step1 分块处理小波嵌入域. 首先对载体图像 I 进行 2 层 DWT 变换,得到图像的垂直细节分量 $HL2$ 和水平细节分量 $LH2$; 然后对 $HL2$ 和 $LH2$ 进行分块处理,每个分块的大小为 2×2 . 2 个小波系数的分量共分为 $M/8 \times N/8$ 块. 根据每个分块在小波系数中的位置,将 $HL2$ 的每个分块记为 $B_0(i, j)$, $LH2$ 的每个分块记为 $B_1(i, j)$, 其中 $1 \leq i \leq 8M, 1 \leq j \leq 8N$.

Step2 选择小波嵌入域. 首先利用伪随机数发生器产生服从均匀分布的数列,然后确定伪随机数发生器的初始值并生成 0,1 矩阵 $choice(i, j)$ ($1 \leq i \leq 8M, 1 \leq j \leq 8N$). 由该矩阵即可确定嵌入域的位置: 当 $choice(i, j) = 0$ 时,相应的水印信息嵌入到分块 $B_0(i, j)$ 中; 当 $choice(i, j) = 1$ 时,相应的水印信息嵌入到分块 $B_1(i, j)$ 中. 该方法不仅可以使水印信息分布得更均匀,而且还可提高水印的不可见性.

Step3 确定水印信息嵌入的量化程度. 将每个分块 $B_0(i, j)$ 或 $B_1(i, j)$ 都嵌入 1 bit 的水印信息. 为了保证水印信息嵌入的不可见性,利用小波系数自适应确定嵌入水印信息的大小. 定义参数 $weight(i, j)$ 为每一分块所能承载的最大量化程度,其表达式为

$$weight(i, j) = mean(i, j) + 3 \times std(i, j) + 10 \times entropy(i, j). \quad (3)$$

其中: $mean(i, j)$ 为 $B(i, j)$ 的均值, $std(i, j)$ 为 $B(i, j)$ 的标准方差, $entropy(i, j)$ 为 $B(i, j)$ 的熵.

Step4 确定水印信息嵌入的量化步长. 按 Step3 操作求出 $HL2$ 和 $LH2$ 所有分块所能承受的最大量化程度的矩阵 $weight_0(i, j)$ 和 $weight_1(i, j)$, 以此确定分块 $B(i, j)$ 的量化步长. 水印信息嵌入的量化步长公式为

$$\Delta(i, j) = c + (d - c) \times \frac{weight(i, j) - w_{\min}}{w_{\max} - w_{\min}}, \quad (4)$$

其中 w_{\min} 为 $weight(i, j)$ 分块的最小值, w_{\max} 为 $weight(i, j)$ 分块的最大值, c 和 d 为选定的量化步长的范围.

Step5 嵌入水印信息. 待嵌入的水印信息为 $W(i, j)$, 水印信息嵌入的位置为 $B(i, j)$, 自适应嵌入的量化步长为 $\Delta(i, j)$, 矩阵内的小波系数为 x_i , 分块 $B(i, j)$ 绝对均值的计算公式为

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n |x_i|. \quad (5)$$

定义 $(\bar{x} + \Delta(i, j)/2)/\Delta(i, j) = Q + r$, 其中 Q 为商, r 为余数. 根据 Q 和 W 的奇偶性知, 通过修改均值 \bar{x} 可实现水印信息的嵌入. 设修改后的均值为 \bar{x}' , 其计算公式为:

$$\begin{cases} \bar{x}' = \Delta(i, j) \times Q + \Delta(i, j)/2, & Q \text{ 和 } W \text{ 奇偶性相同;} \\ \bar{x}' = \Delta(i, j) \times Q - \Delta(i, j)/2, & Q \text{ 和 } W \text{ 奇偶性不同.} \end{cases} \quad (6)$$

定义均值的改变量 $g = \bar{x}' - \bar{x}$, 则小波系数更新方法为:

$$\bar{x}' = x_i \times (1 + \frac{g}{\bar{x}}). \quad (7)$$

式中 x_i 代表嵌入信息后的系数. 根据新产生的均值进行更新即可实现水印信息的嵌入, 再经过小波逆变换即可生成含水印的图像.

2.3 水印信息的提取算法

提取水印信息时, 首先对含水印图像 I_w 进行 2 层 DWT 变换, 以此得到 I_w 的小波系数分量 $HL2$ 和 $LH2$; 然后利用 2.2 中 Step1 的算法对垂直细节子带和水平细节子带进行分块(每块大小为 2×2), 并求出每块分块的最大水印信息嵌入程度 $weight'(i, j)$ 以及自适应量化步长 $\Delta'(i, j)$; 最后利用 2.2 中嵌入算法的初始值条件生成随机选择矩阵 $choice(i, j)$ 来控制水印信息的嵌入强度. 水印信息 W_m^* 可由下式计算得到:

$$\bar{x}_w/\Delta'(i, j) = \tilde{Q} + r', \quad (8)$$

其中 \tilde{Q} 为商, r' 为余数. 当 \tilde{Q} 为偶数时, $W_m^*(i, j) = 0$; 当 \tilde{Q} 为奇数时, $W_m^*(i, j) = 1$.

2.4 图像的认证和篡改定位

首先提取含水印图像 I_w 的特征信息 W_0^* , 然后通过比较 W_m^* 和 W_0^* 的不同点来判断图像中的篡改部分. 篡改定位矩阵的表达式为

$$T = |W_0^* - W_m^*|. \quad (9)$$

如果 $T = 0$, 说明载体图像没有发生篡改; 如果 $T \neq 0$, 则 T 中不为 1 的点被认为发生了篡改.

3 实验结果及分析

为了验证本文提出的半脆弱水印算法的有效性, 以像素为 256×256 的 Lena 图像为例, 选用正交小波函数 harr 作为离散余弦变换的小波基进行实验. 计算机的操作系统为 Windows 7 Professional $\times 64$, 软件为 Matlab R2017a $\times 64$.

1) 透明性检测. 首先对 Lena 图像进行嵌入水印的操作, 结果如图 2 所示. 由图 2 可以看出, 用肉眼难以发现两幅图像有无不同之处, 这表明本文算法具有良好的透明性. 为进一步验证本文算法对透明性的有效性, 分别计算本文算法、文献[12]算法和文献[13]算法的峰值信噪比(peak signal noise ratio, PSNR). 峰值信噪比的计算公式为

$$PSNR = 10 \lg \frac{255^2}{\frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [f(m, n) - g(m, n)]^2}. \quad (10)$$

式中 $f(m, n)$ 是原图的像素值, $g(m, n)$ 是修改后图的像素值. PSNR 的数值越大, 表明嵌入水印信息的

图像与载体图像之间的区别越小,即表明算法的透明性越好. 计算结果显示,本文算法、文献[12] 算法和文献[13]算法的 $PSNR$ 分别为 50.55 dB、40.28 dB 和 41.04 dB,该结果说明本文算法的透明性优于文献[12] 和文献[13] 算法.



(a) 原始载体图像 (b) 嵌入水印后的载体图像

图 2 原始载体图像和含水印图像的对比

2) 压缩攻击实验. 为了检验本文算法的抗压缩攻击性能,分别对采用本文算法和文献[14]算法得到的水印图像进行不同程度的 JPEG 压缩攻击实验,结果见表 1. 从表 1 中的数据可以看出,本文算法在压缩比小于 50% 时,能够较好地提取出嵌入的水印信息;虽然压缩率增大时篡改率随之增大,但总体效果较好,且优于文献[14]的效果. 该结果表明,本文算法对 JPEG 压缩攻击具有较好的鲁棒性.

表 1 不同 JPEG 压缩率下的图像篡改率 %

JPEG 压缩率	本文算法的篡改率 F	文献[14]算法的篡改率 F
10	45.5	60.2
20	34.4	48.5
30	23.8	30.1
40	21.7	16.8
50	4.2	9.5
60	2.6	3.8
70	2.3	2.5
80	2.3	2.2
90	2.3	2.1

3) 噪声攻击实验. 对图像分别进行高斯噪声攻击和椒盐噪声攻击,攻击类型的参数及其误码率 BER 如表 3 所示. 由表 3 可以看出,图像经过不同的高斯噪声和椒盐噪声攻击后,提取水印信息的误码率始终保持在 25% 以下,即具有较好的鲁棒性. 由此表明,本文算法允许对图像进行噪声攻击等常规的图像处理.

表 3 不同噪声攻击下图像的误码率 BER

噪声攻击类型	BER / %
椒盐噪声 1 (方差为 0.005)	13.5
椒盐噪声 2 (方差为 0.01)	19.8
高斯噪声 1 (方差为 0.005,均值为 0)	15.4
高斯噪声 2 (方差为 0.01,均值为 0)	23.6

4) 剪切攻击实验. 对图像进行不同剪切率(8.66%、11.36%、19.33%、38.54%)的攻击后其结果如图 3 所示. 由图 3 可以看出,本文算法能够较好地对图像篡改的位置进行定位,但只能定位出篡改位置的大致轮廓. 在 4 张定位图像中,未发生篡改的位置有白色噪声点出现,但这些噪点相对分散,并不会对算法的准确性造成较大影响.

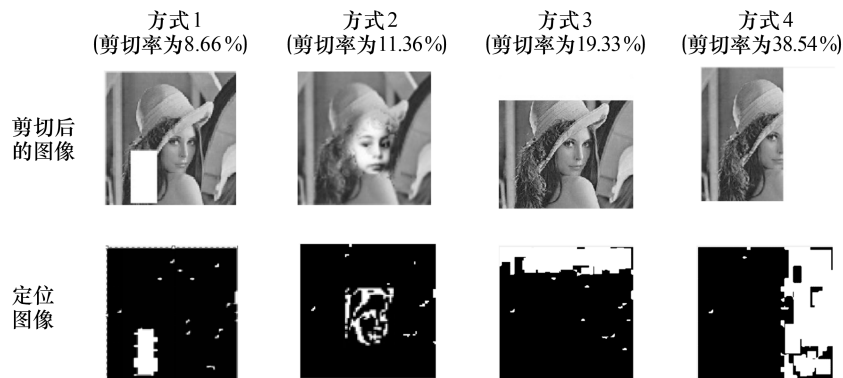


图 3 不同剪切率下针对 Lena 图像的篡改检测效果

4 结论

研究表明,本文提出的基于小波域半脆弱水印的图像内容认证算法提高了水印信息的不可见性,并实现了水印信息的盲提取.因本文方法对水印信息的提取和验证只需提供待认证的图像,因此该方法具有安全性和实用性,可应用于鉴别图像在复制、传输和存储等过程中出现的修改.图像受到剪切攻击时,本文算法能够准确定位被篡改的位置,但定位位置的边缘精度还有待进一步提高;同时,在未发生篡改的位置出现了少量相对分散的白色噪声点,这些噪声点一定程度地影响了算法的准确性;因此,在今后的研究中我们将进一步改进算法提取图像内容特征的方式,使提取的水印信息包含更多的图像边缘轮廓的有效特征,以此提高本文算法对篡改位置的定位精度以及减少噪声点对未被篡改区域的影响.

参考文献:

- [1] COLLBERG C S, THOMBORSON C. Watermarking, tamper-proofing, and obfuscation-tools for software protection[J]. IEEE Transactions on Software Engineering SE, 2002,28(8):735-746.
- [2] LIN C Y, CHANG S F. A robust image authentication method distinguishing JPEG compression from malicious manipulation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2001,11(2):153-168.
- [3] 胡玉平,张军.用于图像认证及恢复的半脆弱水印算法[J]. 计算机工程,2010,36(11):140-142.
- [4] 赵春晖,刘巍.基于分块压缩感知的图像半脆弱零水印算法[J]. 自动化学报,2012,38(4):609-617.
- [5] 高铁杠,谭琳,曹雁军.基于 Hash 函数和汉明码的半脆弱水印算法[J]. 光电子激光,2012,23(9):1808-1814.
- [6] 王向阳,杨红颖,陈利科,等.基于视觉掩蔽的半脆弱水印技术研究[J]. 中国图像图形学报,2005,10(12):1548-1553.
- [7] 王向阳,杨红颖,邬俊.一种基于自适应量化的半脆弱图像水印算法[J]. 小型微型计算机系统,2006,27(5):896-900.
- [8] 高雪丽,李德.基于边缘特征的动画零水印算法[J]. 延边大学学报(自然科学版),2017,43(4):353-358.
- [9] 侯振华,陈生潭.用于图像认证的半脆弱性数字水印研究[J]. 电视技术,2003(11):88-91.
- [10] 黄开兴,罗永,成礼智.基于方向小波和直方图平移的无损脆弱水印[J]. 计算机工程,2009,35(15):153-155.
- [11] 王俊文,刘光杰,戴跃伟,等.一种估计 JPEG 双重压缩原始量化步长的新方法[J]. 电子与信息学报,2009,31(4):836-839.
- [12] YANG H F, SUN X M. Semi-fragile watermarking for image authentication and tamper detection using HVS Model[C]//International Conference on Multimedia and Ubiquitous Engineering (MUE'07). Seoul: IEEE, 2007: 1112-1117.
- [13] QI X J, XIN X. A quantization-based semi-fragile watermarking scheme for image content authentication[J]. J Vis Commun Image R, 2011,22(2):187-200.
- [14] BAE S H, KIM M. A novel image quality assessment with globally and locally consilient visual quality perception [J]. IEEE Transactions on Image Processing, 2016,25(5):2392-2406.