块级篡改定位的JPEG图像脆弱水印

金喜子[1,2],姜文哲[3]
(1.吉林大学计算机科学与技术学院,吉林长春 130012;2.东北师范大学计算机学院,吉林长春 130117;
3.东北电力设计院,吉林长春 130021)

摘 要: JPEG是一种常见的图像格式,在JPEG图像中进行准确的篡改定位具有重要意义.本文提出一种新的JPEG图像脆弱水印方案,将每个小块主要内容的Hash比特重新分组,并将每组的模2和作为水印信息.也就是说每一个小块都对应多个水印比特,每个水印比特也对应多个小块.载体图像的每个小块中仅嵌入于1比特水印,保证了良好的隐蔽性.认证时依据图像内容与水印比特的整体匹配情况估计篡改率,再根据每个小块对应的水印信息的被破坏程度判别该小块是否曾被篡改.理论分析和实验结果标明该方法可以在篡改区域小于1%的情况下准确地找到所有篡改小块.

关键词: 脆弱水印;JPEG图像;图像认证
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2010) 07-1585-05

Fragile Watermarking Capable of Locating Tampered Blocks in JPEG Images

JIN Xi-zi[1,2], JIANG Wen-zhe[3]
(1. *College of Computer Science; JI LIN University, Changchun, Jilin 130021, China;*
2. *College of Computer Science; Northeast Normal University, Changchun, Jilin 130117, China;*
3. *Northeast Electric Power Design Institute, Changchun, Jilin 130021, China*)

Abstract: It is significantly meaningful to exactly locating the tampered areas in digital images. This paper proposes a novel fragile watermarking scheme for JPEG covers, in which the hash bits of all blocks with a size of 8×8 are pseudo-randomly reorganized as a series of subsets, and their sums with modulus 2 are regarded as watermark bits. This way, each block maps a number of watermark bits, and each watermark bit corresponds to a number of blocks. Then, a selected non-zero DCT coefficient in each block is used to accommodate one watermark bit. At authentication side, after estimating the tampering rate, the inconsistency between the calculated and extracted watermark bits of each block is used to judge it as "tampered" or "reserved". Theoretical analysis and experimental result show that the proposed method is capable of locating all tampered blocks when the tampering rate is less than 1%.

Key words: fragile watermarking; JPEG image; image authentication

1 引言

脆弱水印是数字水印的重要分支,可用于数字媒体的真实性和完整性认证.数字媒体的保护者事先将水印信息以隐蔽的方式隐藏于媒体内容中,认证时根据水印信息被破坏的情况察觉篡改行为[1,2].如果水印对正常的信号处理(如压缩、滤波等)不敏感,却能察觉恶意处理,则被称为半脆弱水印[3,4].脆弱水印或半脆弱水印不仅仅可以判断数字媒体是否经过篡改,而且可以用于确定被篡改内容的位置,这对于洞察篡改者的意图、进一步有效地打击篡改行为具有重要意义.

目前已有许多以未压缩格式图像为载体的脆弱水印技术研究成果,这些成果大多具备篡改定位功能,大致可分为两类:基于块的脆弱水印技术、基于象素的脆弱水印技术.基于块的技术往往将图像块主要内容的Hash结果作为脆弱水印嵌入到图像块中,如果图像遭到篡改,图像块主要内容的Hash结果与从图像块中提取出的脆弱水印将不再匹配,因此可察觉哪些图像块经过了篡改[5~7].为了提高篡改定位精度,研究者们提出了一系列基于象素的脆弱水印技术.在早期的这类方法中,水印信息往往嵌入于象素灰度值的低位中,认证时根据水印检测情况进行篡改定位[8~10].但是篡改后象

收稿日期:2009-03-15;修回日期:2010-01-12
基金项目:国家自然科学基金(No.60573067,No.60473042)

素的低位会以一定概率与所嵌水印信息吻合，所以对篡改象素的定位是不完全的.研究者近来发展出了基于统计机制的脆弱水印方案[11]，每个象素的对应的多比特水印信息隐藏于图像的多个位置，在篡改区域不大的情况下，可以根据多个位置的水印信息的正误情况准确区分原始象素和篡改象素.文献[12]融合了两类方法的优势，提出了基于分级机制的脆弱水印方案，认证时先找到篡改图像块再准确定位所有篡改象素.

JPEG是常见的图像压缩格式，准确地定位篡改区域对保卫JPEG图像安全具有重要意义.JPEG图像由许多8×8的小块以量化后的DCT系数形式组成，最理想的篡改定位效果就是找到所有被篡改的8×8小块.但JPEG图像中可用于负载水印信息的空间较少，所以准确定位篡改区域的难度较大.文献[13]提出了一种以JPEG图像为载体的脆弱水印方案，在每个8×8的小块中嵌入4比特水印信息用于篡改定位.该方案有两个弱点：一是嵌入信息较多，当载体图像质量因子较低时，水印嵌入会引起较明显的失真；二是尽管该方案根据邻域情况减少虚警，但仍然不能对篡改小块进行完全准确的定位.

本文提出一种新的JPEG图像脆弱水印方案，仅在每个小块中嵌入1比特水印信息，保证了良好的水印隐蔽性.认证时依据图像内容与水印比特的整体匹配情况估计篡改率，再根据每个小块对应的水印信息的被破坏程度判别该小块是否曾被篡改.理论分析和实验结果表明该方法可以在篡改区域小于1%的情况下准确地找到所有篡改小块.

2 水印嵌入过程

JPEG编码过程中，首先将图像分成8×8的小块，在每个小块内进行DCT变换，然后将不同位置的DCT系数按照不同的步长量化为整数.所以可以将JPEG图像小块看作8×8的二维整数数组.通常来说，图像内容集中于中低频，大多高频系数为零.

为了减少水印信息量，本文方法并不是将每个小块的Hash直接嵌入载体图像中，而是先求出每个小块主要内容的Hash，然后将Hash比特重新分组并作模2和，并将这些模2和的结果作为真正的水印信息.水印比特数与JPEG图像的小块数一致，因此在每个小块中仅需嵌入1比特水印，保证了良好的隐蔽性.

具体的水印嵌入过程如下：

(1)设JPEG载体图像包括N个小块，根据密钥在每个小块所有非0值的DCT系数中伪随机选取一个，记为$C_n(i_n,j_n)$，这里(i_n,j_n)表示第n个小块中所选非0系数的频率位置，$1\leqslant n\leqslant N$，$1\leqslant i_n,j_n\leqslant8$.按如下方式将所选系数进行偶数化，

$$C_n'(i_n,j_n)=\begin{cases}C_n(i_n,j_n)+\mathrm{mod}[C_n(i_n,j_n),2], & C_n(i_n,j_n)>0\\ C_n(i_n,j_n)-\mathrm{mod}[C_n(i_n,j_n),2], & C_n(i_n,j_n)<0\end{cases}\tag{1}$$

也就是说，如果$C_n(i_n,j_n)$是偶数，则保持不变；如果$C_n(i_n,j_n)$是奇数，则符号不变，绝对值加1变为偶数.

(2)对于每一个小块，用$C_n'(i_n,j_n)$替代$C_n(i_n,j_n)$.然后将64个系数值输入一个单向Hash函数，得到128位比特，共得到$128\cdot N$个Hash比特.

(3)根据密钥将$128\cdot N$个Hash比特进行伪随机置乱，并分成N组，每组包含128个Hash比特，并求128个Hash比特的模2和，将该结果称为水印比特，总共得到N个水印比特，记为$W_n(1\leqslant n\leqslant N)$.

(4)将N个水印比特与N个图像小块一一对应，并按如下方法在每个小块的所选非0系数中嵌入一个水印比特，

$$C_n''(i_n,j_n)=\begin{cases}C_n'(i_n,j_n), & W_n=0\\ C_n'(i_n,j_n)-\mathrm{sign}[C_n'(i_n,j_n)], & W_n=1\end{cases}\tag{2}$$

也就是说，如果水印比特为0，则保持$C_n'(i_n,j_n)$不变；如果水印比特为1，则保持$C_n'(i_n,j_n)$的符号不变，绝对值减1.

事实上，水印嵌入过程在每个小块中最多改变一个非0系数值，而且修改后的该系数值依然非0，且修改幅度不会超过1，因此具有较好的隐蔽性.具体而言，如果该系数原始值是偶数，水印嵌入后该系数保持不变或符号不变、绝对值减1；如果该系数原始值是奇数，水印嵌入后该系数保持不变或符号不变、绝对值加1.嵌入过程中引入密钥进行伪随机选取、置乱的作用是提高安全性，使无密钥的攻击者无法仿冒或提取水印信息.

3 篡改定位过程

假设别有用心者篡改了含水印图像的部分内容，我们称内容发生变化的小块为篡改小块，称篡改小块个数与N的比值为篡改率，记为α.篡改定位的目的就是准确地找到这些篡改小块的位置.由于水印比特是不同小块的Hash比特的模2和，每一个小块都对应多个水印比特，每个水印比特也对应多个小块.认证时先依据图像内容与水印比特的整体匹配情况估计篡改率，再根据每个小块对应的多个水印比特的被破坏程度判别该小块是否曾被篡改.

具体认证过程(篡改定位过程)如下：

(1)首先从待认证图像的所有小块中提取水印信息.根据同样的密钥在每个小块所有非0值的DCT系数中选取一个系数，若该系数是偶数，则提取水印比特

0;反之提取水印比特 1.将提取出的水印比特简记为 EWB(Extracted Watermark Bit).

(2)然后计算待认证图像的 Hash 比特.对于待认证图像的每一个小块,将所选非 0 系数按式(1)进行偶数化处理,然后将 64 个系数输入同样的单向 Hash 函数,得到 128 个 Hash 比特.

(3)根据密钥将由待认证图像计算得到的 Hash 比特进行同样的分组,并计算每组 128 个 Hash 比特的模 2 和,我们称之为 CWB(Calculated Watermark Bit).根据 CWB 与 EWB 的不一致率估计篡改率(具体估计方法见后).

(4)每一个小块都会有 128 个 Hash 比特,并分属 128 个 Hash 比特组,每个 Hash 比特组还包括 127 个由其他小块得出的 Hash 比特,其模 2 和即 CWB,也就是说每个小块对应于 128 个 CWB.这些 CWB 又一一对应于 EWB,记一个小块的 CWB 与 EWB 不一致的个数为 K_n.我们根据每个小块的 K_n 值判断该小块是否曾被篡改小块(具体判断方法见后).

如果含水印小块内容没有发生变化,根据密钥选取的系数一定是水印嵌入过程中用于承载水印比特的系数,在此情况下,EWB 是正确的.而对于篡改小块来说,非 0 系数的个数、位置、取值都有可能发生变化,此时 EWB 与嵌入的水印比特没有关系,因此其正确概率和错误概率都为 1/2.就所有 EWB 来说,其错误概率为 $\alpha/2$.如果含水印小块内容没有发生变化,计算得到的 128 个 Hash 比特与嵌入过程中的 128 个原始 Hash 比特完全相同.而对于篡改小块来说,由于单向 Hash 函数会由不同的输入得到截然不同的输出,因此 Hash 比特与原始 Hash 比特的相同概率为 1/2.对于由待认证图像计算得到的所有 Hash 比特来说,与对应的原始 Hash 比特的不一致概率为 $\alpha/2$.

考虑对计算得到的 Hash 比特进行分组,每组内 Hash 比特与原始 Hash 比特不一致的个数服从二项分布,

$$P_H(m)=\binom{128}{m}\cdot\left(\frac{\alpha}{2}\right)^m\cdot\left(1-\frac{\alpha}{2}\right)^{128-m} \tag{3}$$

那么 CWB 发生变化(不同于原始水印比特)的概率

$$e_C=\sum_{m\ \text{is odd}}[P_H(m)] \tag{4}$$

与嵌入过程步骤 4 类似,建立 CWB 与小块的一一对应关系,并与从对应小块提取出的 EWB 相比较,将不一致的比例记为 β.如前所述,EWB 的错误概率为 $\alpha/2$,因此

$$\beta=e_C\cdot(1-\alpha/2)+(1-e_C)\cdot\alpha/2 \tag{5}$$

简言之,CWB 与 EWB 的不一致率 β 是关于篡改率 α 的函数,图 1 给出了两者的关系.认证者可根据 β 值估计篡改率 α.

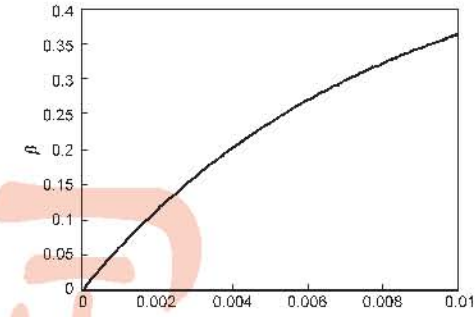

图1　CWB与EWB的不一致率β与篡改率α的关系

对于篡改小块来说,Hash 比特发生变化的概率是 1/2,所以 CWB 发生变化的概率也是 1/2,因此无论 EWB 以怎样的概率发生变化,CWB 与 EWB 不一致的概率都是 1/2.可见,篡改小块的 K_n 服从二项分布,

$$P_T(K_n=k)=\binom{128}{k}\cdot\left(\frac{1}{2}\right)^{128},\quad k=0,1,\cdots,128 \tag{6}$$

其概率集中于期望值 64 附近.而对于内容没有发生变化的原始小块来说,其 128 个 Hash 比特不会发生变化,但是每个 Hash 比特组中其余 127 个 Hash 比特会以 $\alpha/2$ 概率发生变化,EWB 也会以 $\alpha/2$ 概率发生变化,因此 CWB 与 EWB 不一致的概率与式(4)中的 e_C 相等.那么,原始小块的 K_n 服从另一个不同的二项分布,

$$P_O(K_n=k)=\binom{128}{k}\cdot(e_C)^k\cdot(1-e_C)^{128-k},k=0,1,\cdots,128 \tag{7}$$

该分布会因篡改率 α 的不同而变化.当 α 较小时,K_n 会远小于 64.因此,认证者可以根据每个小块对应的 K_n 值来判断其内容是否经过篡改,准则如下:如果

$$\hat{\alpha}\cdot P_T(K_n)>(1-\hat{\alpha})\cdot P_O(K_n) \tag{8}$$

则将该小块判为篡改小块;否则判为未篡改小块.这里 $\hat{\alpha}$ 是认证者根据 β 值获得的篡改率 α 的估计值.假设认证者对篡改率的估计完全准确,那么误判总概率(包括将篡改小块判为未篡改小块、将未篡改小块判为篡改小块两种情况)为

$$P_E=\sum_{K_n=0}^{128}\min[\alpha\cdot P_T(K_n),(1-\alpha)\cdot P_O(K_n)] \tag{9}$$

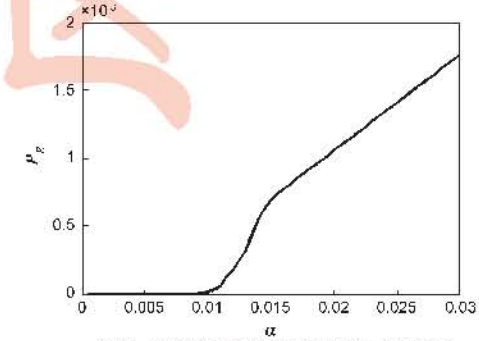

图2　误判总概率P_E与篡改率α的关系

图2给出了误判总概率 P_E 与篡改率 α 的关系.可见在篡改率小于1%时,误判总概率非常接近于0,即几乎可以准确定位所有篡改小块.

4 实验结果

在多幅具有不同质量因子的JPEG图像中进行水印嵌入,表1列出了水印嵌入引起的峰值信噪比的平均值.通常认为峰值信噪比高于38dB的影响是难以被视觉察觉的,可见本文方法具有较好的隐蔽性.图3给出了质量因子为70的一幅原始图像(大小为1080×1440),以及嵌入水印之后的版本(峰值信噪比为49.2dB),视觉上无法察觉两者的差异.

表1 水印嵌入在不同质量因子载体图像中产生的平均峰值信噪比

JPEG载体图像质量因子	100	90	80	70	60	50	40
水印嵌入产生的PSNR(dB)	66.2	60.1	54.1	49.3	46.2	43.9	42.4

(a)质量因子为70的原始图像 (b)嵌入水印之后的版本

图3

对含水印图像图3(b)进行篡改,改变了图像右下角车牌中的两个号码(篡改率为0.69%),篡改图像如图4(a)所示.对篡改后的图像进行认证,篡改定位结果如图4(b)所示,被认定为篡改小块的位置用黑色代替.认证结果完全准确,既无原始小块被认定为篡改小块、也无篡改小块被认定为原始小块.从图4(b)可以看出,由于实现了准确的篡改定位,篡改区域周围的真实内容(车牌的第一和第三个数字)依然可以用于刑侦或法庭证据.

(a) 篡改后的图像 (b) 篡改定位结果

图4

对含水印图像进行不同程度的篡改并进行认证.当篡改率大于1%时,会出现误判现象(包括将原始小块误认为篡改小块或将篡改小块误认为原始小块两种情况).当篡改面积相对于整幅图像较小时,误判现象大多是将原始小块误认为篡改小块,而且分散于整幅图像中.表2列出了不同篡改率时的误判总概率,表2第二行给出的误判总概率理论值,即式(9)与图2中的结果,第三行给出的是实际结果,吻合于理论结果.

表2 不同篡改率水平下的误判总概率的理论值和实际值

篡改率 α	0.69%	0.88%	1.17%	1.56%	1.97%	2.34%	2.90%
误判总概率的理论值	3.1×10^{-9}	2.0×10^{-6}	1.4×10^{-4}	7.4×10^{-4}	1.0×10^{-3}	1.3×10^{-3}	1.7×10^{-3}
误判总概率的实际值	0	0	1.2×10^{-4}	8.0×10^{-4}	1.1×10^{-3}	1.3×10^{-3}	1.8×10^{-3}

5 结论

本文以JPEG图像为载体设计了一种新的脆弱水印方案.该方案将每个小块主要内容的Hash比特的模2和作为水印信息,大大减少了水印信息量.载体JPEG图像的每个小块中仅需嵌入1比特水印,保证了良好的水印隐蔽性.认证时根据待认证图像内容计算水印信息,并与从待认证图像中提取的水印信息进行比较,根据整体匹配情况估计篡改率,再根据每个小块对应的水印信息的被破坏程度判别该小块是否曾被篡改.当篡改区域小于1%时,本文方法可以准确定位所有篡改小块.如果对含水印图像进行不同质量因子或不同量化矩阵的重新编码、或在含水印图像中迭加较强的噪声,这相当于对含水印图像进行大范围的篡改操作,因此不能定位篡改位置.理论分析准确的篡改定位能力极限并设计可准确定位更大篡改区域的脆弱水印方法将是下一步研究内容.

参考文献:

[1] F A P Petitcolas, R J Anderson, M G Kuhn. Information hiding-a survey[J]. Proc. IEEE, 1999, 87: 1062 - 1078.

[2] 王国栋,刘粉林,刘瑗,姚刚.一种能区分水印或内容篡改的脆弱水印算法[J].电子学报,2008,36(7):1349 - 1354.
Wang Guodong, Liu Fenlin, Liu Yuan, Yao Gang. An image authentication scheme with discrimination of tampers on watermark or image[J]. Acta Electronica Sinica, 2008, 36(7): 1349 - 1354. (in Chinese)

[3] O Altun, G Sharma, M U Celik, M F Bocko. A set theoretic framework for watermarking and its application to semifragile tamper detection[J]. IEEE Trans. on Information Forensics and Security, 2006, 1(4): 479 - 492.

[4] K Maeno, Q Sun, S Chang, M Suto. New semi-fragile image authentication watermarking techniques using random bias and

nonuniform quantization [J]. IEEE Trans. on Multimedia, 2006,8(1):32 – 45.

[5] P W Wong, N Memon. Secret and public key image watermarking schemes for image authentication and ownership verification[J]. IEEE Trans. on Image Processing, 2001, 10(10): 1593 – 1601.

[6] S Suthaharan. Fragile image watermarking using a gradient image for improved localization and security[J]. Pattern Recognition Letters, 2004, 25: 1893 – 1903.

[7] H Yang, A C Kot. Binary image authentication with tampering localization by embedding cryptographic signature and block identifier[J]. IEEE Signal Processing Letters, 2006, 13(12): 741 – 744.

[8] S Liu, H Yao, W Gao, Y Liu. An image fragile watermark scheme based on chaotic image pattern and pixel-pairs[J]. Applied Mathematics and Computation, 2007, 185(2): 869 – 882.

[9] H He, J Zhang, H Tai. A Wavelet-Based Fragile Watermarking Scheme for Secure Image Authentication[A]. in Proceeding of 5th International Workshop on Digital Watermarking[C], Lecture Notes in Computer Science, 4283, Springer-Verlag, 2006. 422 – 432.

[10] 张宪海,杨永田. 基于脆弱水印的图像认证算法研究[J]. 电子学报,2007,35(1):34 – 39.
Zhang Xianhai, Yang Yongtian. Image authentication scheme research based on fragile watermarking[J]. Acta Electronica Sinica, 2007, 35(1): 34 – 39. (in Chinese)

[11] X Zhang, S Wang. Statistical fragile watermarking capable of locating individual tampered pixels[J]. IEEE Signal Processing Letters, 2007, 14(10): 727 – 730.

[12] X Zhang, S Wang. Fragile watermarking scheme using a hierarchical mechanism[J]. Signal Processing, 2009, 89(4): 675 – 679.

[13] C Li. Digital fragile watermarking scheme for authentication of JPEG images[J]. IEE Proc.-Vis. Image Signal Process., 2004, 151(6): 460 – 466.

作者简介:

金喜子　女,1965 年 2 月出生于吉林龙井,1986 年毕业于东北师范大学数学系,2003 年获日本大阪教育大学教育学硕士学位. 现为东北师范大学计算机学院副教授,并在吉林大学计算机科学与技术学院攻读博士学位. 主要研究领域:数字水印、数字图像处理等.

E-mail: jinxz865@nenu.edu.cn