ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.20, No.2, February 2009, pp.437−450
doi: 10.3724/SP.J.1001.2009.03184

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

对水印信息篡改鲁棒的自嵌入水印算法*

和红杰+, 张家树

(西南交通大学 信号与信息处理四川省重点实验室,四川 成都 610031)

Self-Embedding Watermarking Algorithm with Robustness against Watermark Information Alterations

HE Hong-Jie+, ZHANG Jia-Shu

(Sichuan Key Laboratory of Signal and Information Processing, Southwest Jiaotong University, Chengdu 610031, China)

+ Corresponding author: E-mail: Hehojie@126.com

He HJ, Zhang JS. Self-Embedding watermarking algorithm with robustness against watermark information alterations. *Journal of Software*, 2009,20(2):437−450. http://www.jos.org.cn/1000-9825/3184.htm

Abstract: To improve the quality of the reconstructed image in any tamper condition, this work proposes a self-embedding watermarking scheme with robustness against watermark alterations, and discusses the reasonability of the predefined threshold and the reliability of tamper detection. The proposed scheme firstly sets the least significant bit (LSB) of three-quarter pixels and two LSBs of the residual pixels in the original image to zero. And then the low-frequency feature image is obtained by quantizing the low-frequency wavelet coefficients of the original image content. The improved security watermark, which is the binary code of the scramble version of the low-frequency feature image, is embedded into the LSBs which were set to zero. While the image authentication, the proposed method is able to discriminate the malicious modifications from the mild distortions according to the predefined threshold to enhance the robustness against innocuous alterations such as watermark changes and channel noise. Theoretical analysis and simulation results show that the proposed scheme can discriminate the different modifications according to the predefined threshold no matter the embedded watermark in the host image is randomly or regionally tampered. The quality of the recovery image can be effectively improved due to the fact that the different reconstructed methods are adopted for the different tamper blocks.

Key words: fragile watermarking; self-embedding; discrete wavelet transform; scalar quantization

摘 要: 为提高自嵌入水印算法在任意篡改条件下的篡改恢复质量,提出一种对水印信息篡改鲁棒性的空域自嵌入水印算法,分析了算法中阈值选取的合理性和检测篡改的可靠性.该算法首先基于密钥将原始图像的最低位和1/4 次低位置零,通过对图像内容的小波低频系数实施均匀标量量化生成低频特征图像,将低频特征图像置乱加密后生成的二值编码嵌入原始图像的置零位;认证时通过设定的阈值识别图像内容被恶意篡改的图像块,从而提高自嵌入水印算法对水印信息篡改和信道噪声的鲁棒性.理论分析和仿真结果表明,无论水印信息被随机篡改还是区域

* Supported by the Program for New Century Excellent Talents in University of China under Grant No.NCET-05-0794 (新世纪优秀人才支持计划); the Application Basic Foundation of Sichuan Province of China under Grant No.2006 J13-10 (四川省应用基础研究项目); the Doctors Innovation Funds of Southwest Jiaotong University of China under Grant No.2007 (西南交通大学博士生创新基金)
Received 2007-04-02; Accepted 2007-09-19

篡改,算法均能根据阈值区分不同篡改并选用不同的方法对其进行篡改恢复,有效地提高了自嵌入算法在部分水印信息篡改时的恢复质量.

关键词: 脆弱水印;自嵌入;离散小波变换;标量量化

中图法分类号: TP309 **文献标识码**: A

基于数字水印的图像认证技术分为精确认证(hard authentication)和模糊认证(soft authentication)两类[1].其中,精确认证数字水印技术要解决的核心问题是鉴别数字图像的真实性(篡改检测)和定位图像被篡改的位置(篡改定位)[2−4],并由此推断图像被篡改的程度和方式[5−10].自嵌入水印算法[7−10]作为精确认证技术的一种,不仅能够定位图像被篡改的位置,而且能够近似恢复图像被篡改的内容,充分体现了基于数字水印的图像认证技术的优势.

在自嵌入水印算法中,为了得到较好的篡改恢复质量,要求基于图像内容生成的水印信息量越多越好,兼顾水印嵌入的不可见性,在图像的不重要位(least significant bit,简称 LSB)嵌入水印信息就成为自嵌入算法的首选[7−10].当含水印图像中嵌入的水印信息不被篡改时,即使图像内容被严重篡改,现有自嵌入水印算法也能有效恢复被篡改的图像内容.然而,当部分水印信息被改变(即使被改变的水印信息量很少)时,利用现有自嵌入算法得到的篡改恢复图像的质量会急剧降低.对含认证水印的数字图像而言,尽管攻击者不会故意破坏含水印图像中的水印信息,但要保证含水印图像中嵌入的水印信息完全不被改变也不太可能.一方面,通过网络传输的含水印数字图像不可避免地会受到信道噪声的影响,另一方面,攻击者在篡改图像内容时也不可能刻意保持水印信息不变以利于篡改恢复.因此,在确保自嵌入算法有效检测和恢复篡改的条件下,如何提高自嵌入算法对水印信息篡改的鲁棒性,即在少量水印信息改变时提高自嵌入算法的篡改恢复质量,就成为自嵌入认证水印算法走向实际应用必须解决的关键问题.

通过对现有自嵌入算法的深入研究发现,以下 3 个方面的问题是导致现有自嵌入水印算法对水印信息篡改脆弱的主要原因:

① 不能识别哪些图像块需要恢复:在自嵌入水印算法中,基于图像块内容生成的水印信息被嵌入在其他图像块的最低位,因此含水印图像中一个图像块(记作 A_i)被篡改,会导致该图像块 A_i 和另一个图像块 A_j (它的水印信息被嵌入在 A_i 的低位)同时被检测.此时,图像块 A_i 需要恢复,而图像块 A_j 不需要恢复.但是,现有分块自嵌入算法不能有效识别这两种图像块,从而导致一些真实的图像块被误判为篡改(虚警)并错误地执行恢复操作.

② 定位精度不高:算法以 8×8 图像块为单位生成水印信息,即算法的篡改定位精度为 8×8 个像素.

③ 对噪声敏感:为提高现有自嵌入水印算法抵抗伪造攻击[11−13]的能力,现有自嵌入水印算法一般采用 RSA,DES 等方法加密水印信息,这些加密方法具有较高的安全性,同时也增加了对水印信息篡改得敏感性.由 DES,RAS 加密算法的性质可知,密文中 1 比特信息的改变,对其解密得到明文中每个比特发生改变的概率约为 50%.因此,现有分块自嵌入水印算法中,1 比特水印信息的改变将导致一个 8×8 图像块被检测且不能对该图像块进行有效的篡改恢复.

为了提高自嵌入水印算法在任意篡改条件下的篡改恢复质量,本文从上述 3 个问题出发,提出一种对水印信息篡改鲁棒的自嵌入水印算法,并分析了该算法的阈值选取和检测篡改的可靠性.该算法利用一级小波分解的低频系数量化生成水印信息以提高自嵌入算法的定位精度,通过对低频特征图像置乱加密来保证算法的安全性,根据设定域值定位图像被恶意篡改的位置并对其进行篡改恢复,这些策略的有机结合大大提高了自嵌入水印算法对水印信息篡改和随机噪声的鲁棒性.理论分析和仿真结果表明无论水印信息被随机篡改(如信道噪声)还是被区域篡改(如恶意替换等),算法均能根据阈值识别出内容被篡改的图像块并对其进行有效的篡改恢复,从而提高了自嵌入算法的实用性.

1 离散小波变换(DWT)

小波分析(wavelet analysis)是一种日益获得广泛应用的信号分析方法,在信号分析、视频图像分析、数据

压缩等领域都有重要的应用,已成为 JPEG2000 压缩标准的核心技术.本文基于离散小波变换(discrete wavelet transform)良好的时-频特性生成水印信息,以提高自嵌入水印算法的篡改定位精度.

利用离散小波变换分析数字图像的一个数学模型[14]为:对于一个给定的数字图像 $\boldsymbol{A}_{m*n}$,图像中的像素构成一个数据集合:

$$A=\{a_{i,j}^0, 0\le i<m, 0\le j<n\} \tag{1}$$

必然存在一个函数 $f(x,y)\in L^2(R^2)$ 使 $f(x,y)=\sum_{i,j}a_{i,j}^0\phi_{i,j}(x,y)$,通过离散小波变换获得 L 层小波分解:

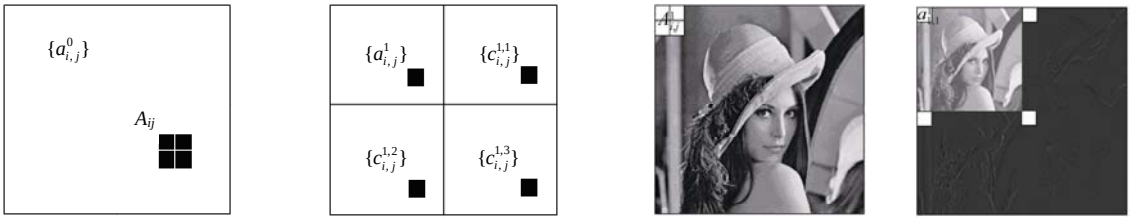
$$f(x,y)=\sum_{i,j}a_{i,j}^0\phi_{i,j}(x,y)=\sum_{i,j}a_{i,j}^L\phi_{i,j}^L(x,y)+\sum_{l,h}\sum_{i,j}c_{i,j}^{k,d}\varphi_{i,j}^{k,d}(x,y) \tag{2}$$

记为

$$DWT\{a_{i,j}^0, 0\le i<m, 0\le j<n\}\Rightarrow\left\{a_{i,j}^L, c_{i,j}^{k,d}, 0\le i<2^{-k}m, 0\le j<2^{-k}n, k=1,2,...,L, d=1,2,3\right\} \tag{3}$$

其中,$a_{i,j}^L$ 称为低频系数,$c_{i,j}^{k,d}$ 为不同方向的高频系数,m 和 n 表示原始图像的行列数.

图像的离散小波变换给出了图像在不同分辨率上的表示[14],每个 DWT 系数反映了图像在局部空域上的内容.图 1 给出了图像局部空域与一级 DWT 系数的映射关系及图像小波分解实例:(a)为空域图像块;(b)为与(a)图像块对应的一级 DWT 系数;(c)为 Lena 原始图像;(d)为 Lena 图像的一级小波分解.



(a) Spatial-Domain block (b) Corresponding DWT coefficients to (a) (c) Lena image (d) 1-Level DWT of Lena

(a) 空域图像块 (b) 与(a)对应的 DWT 系数 (c) Lena 图像 (d) Lena 的一级离散小波变换

Fig.1 Mapping between spatial block and DWT coefficients

图 1 空域块与 DWT 系数映射关系

需要说明的是,对不同的小波基[14,15],由于小波函数支集长度较大,在函数中心振幅之外还有一些幅度较小的余振,与图像同一局部空域相关的 DWT 系数一般会更多.严格说来,图 1 所示的映射关系并不成立,但对于紧支集小波来说,这种重叠面积较小且位于重叠区域内的 DWT 系数的幅值也较小,对图像块的影响相对较小,故图 1(a)所示的空域图像块可近似由图 1(b)中的 DWT 系数来表征.因此,可近似认为一级小波分解的低频系数 $a_{i,j}^1$ 仅与原始图像 2×2 的像素(i×2−1:i×2,j×2−1:j×2)有关,为描述方便,与低频系数 $a_{i,j}^1$ 对应的空域 2×2 的图像块记作 A_{ij}.在我们的算法中,小波低频系数用来生成待嵌入的水印信息,与低频系数对应的空域图像块越小,算法的篡改定位精度越好.因此,本文采用 DB1 小波基.

由图 1(c)和图 1(d)可以看出,小波变换与其他正交变换一样具有良好的去相关能力,变换后系数矩阵的主要能量集中在小波低频系数上,小波低频系数在水平和垂直方向都平滑逼近原始图像[15],因此,基于 $a_{i,j}^1$ 生成水印信息能够保存原始图像块 A_{ij} 的主要信息,利用这些信息能够有效地检测篡改并近似地恢复被篡改图像块的内容.

2 自嵌入水印算法

2.1 水印嵌入

兼顾篡改定位、视觉特性、水印容量及图像的恢复质量等因素,本文选取图像二维一级小波分解的低频系数生成待嵌入的水印信息,对小波低频系数作 5 比特均匀标量量化(scalar quantization)来满足水印信息的嵌入

容量.为使空域中每个 2×2 图像块的特征随机嵌入在其他图像块的低有效位,采用置乱作为本算法的加密方法.图 2 为水印嵌入算法框图,详细步骤如下:

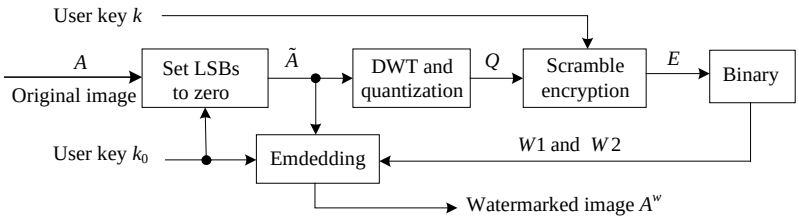


Fig.2 Flowchart of embedding watermark

图 2 水印嵌入算法框图

Step 1. 基于密钥 k_0 将原始图像 $A_{m\times n}$ 的部分低有效位置零生成图像内容 $\tilde{A}$.基于密钥 k_0 生成$(m/2)\times(n/2)$的整数随机矩阵$Z=\{z_{ij}\mid i=1,...,m/2, j=1,...,n/2\}$,然后将每个元素 z_{ij} 转换为一个 2×2 的图像块,得到基于密钥 k_0 生成的二值矩阵 $B_{m\times n}$,即

$$B_{ij}=\begin{cases}\begin{bmatrix}1&0\\0&0\end{bmatrix}, & \text{if } \mathrm{mod}(z_{ij},4)=0\\ \begin{bmatrix}0&1\\0&0\end{bmatrix}, & \text{if } \mathrm{mod}(z_{ij},4)=1\\ \begin{bmatrix}0&0\\1&0\end{bmatrix}, & \text{if } \mathrm{mod}(z_{ij},4)=2\\ \begin{bmatrix}0&0\\0&1\end{bmatrix}, & \text{if } \mathrm{mod}(z_{ij},4)=3\end{cases} \tag{4}$$

则图像内容 $\tilde{A}$ 为

$$\tilde{A}=\lfloor A/2\rfloor-\left(\mathrm{mod}\left(\lfloor A/2\rfloor,2\right)\right)\&B \tag{5}$$

其中,&代表按位与运算,mod()为整除取余,$\lfloor . \rfloor$ 为下取整.

Step 2. 生成低频特征图像 Q.对图像内容 $\tilde{A}$ 作二维一级小波分解(本文选取 DB1 小波基),对其低频系数 $\{\tilde{a}^1_{i,j}\}$ 作 5 比特均匀标量量化,量化区间个数 $N_p=2^5$,生成低频特征图像 $Q=\{q_{ij}|i=1,2,\ldots,m/2, j=1,2,\ldots,n/2\}$,

$$q_{ij}=\begin{cases}x, & \tilde{a}^1_{i,j}\in[\min+xp,\min+(x+1)p)\\ N_p-1, & \tilde{a}^1_{i,j}=\max\end{cases} \tag{6}$$

其中,max 和 min 分别为集合 $\{\tilde{a}^1_{i,j}\}$ 的最大值和最小值,$p=(\max-\min)/N_p$ 为均匀量化步长.

Step 3. 生成加密矩阵 E.基于密钥 k 将低频特征图像 Q 置乱(scrambling)生成置乱加密矩阵 E,即

$$E=S(Q,k) \tag{7}$$

其中,$S(.)$代表置乱加密函数.为保证水印算法的安全性,$S(.)$应满足置乱的随机性和较大的密钥空间,本文采用文献[4]的置乱加密方法.

Step 4. 生成两个待嵌入的二值水印图像 $W1$ 和 $W2$.对加密矩阵的每个元素 e_{ij} 进行二值编码,即 $e_{ij}=(b_{ij4}b_{ij3}b_{ij2}b_{ij1}b_{ij0})$,此处,$b_{ijy}\in\{0,1\}$, $y=0,1,2,3,4$ 且满足:

$$e_{ij}=\sum_{y=0}^{4}2^y b_{ijy} \tag{8}$$

则

$$W1_{ij}=\begin{bmatrix}b_{ij3}&b_{ij2}\\b_{ij1}&b_{ij0}\end{bmatrix}, W2_{ij}=\begin{bmatrix}b_{ij4}&b_{ij4}\\b_{ij4}&b_{ij4}\end{bmatrix} \tag{9}$$

显然,待嵌入的二值水印图像 $W1$ 和 $W2$ 均与原始图像的大小相同,不过 $W2_{ij}$ 中 4 个元素是相同的.

Step 5. 水印嵌入.将二值水印图像 $W1$ 和 $W2$ 嵌入图像内容 $\tilde{A}$ 生成含水印图像 A^W.

$$A^W = \tilde{A} \times 2 + (W2 \,\&\, B) \times 2 + W1 \tag{10}$$

其中,二值矩阵 $\boldsymbol{B}$ 是根据密钥 k_0 按式(4)计算得到的.

2.2 篡改检测及恢复

图 3 为篡改检测与恢复算法框图,详细步骤如下:

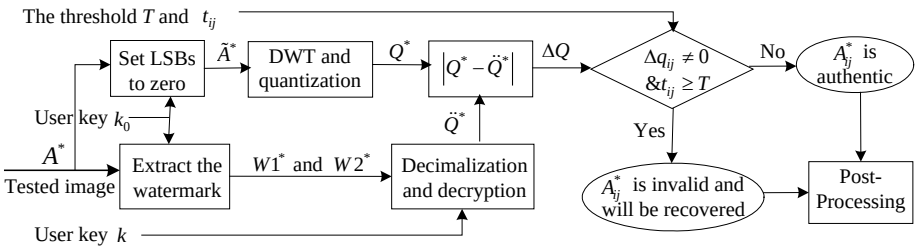


Fig.3 Flowchart of tamper detection and recovery

图 3 篡改检测及恢复框图

Step 1. 水印提取.根据密钥 k_0,按照式(4)生成二值矩阵 $\boldsymbol{B}$,基于 k_0 从被测图像中提取的水印信息为

$$\begin{cases} W1^* = \mathrm{mod}(A^*, 2) \\ W2^* = \left(\mathrm{mod}\left(\left\lfloor A^*/2 \right\rfloor, 2\right)\right) \,\&\, B \end{cases} \tag{11}$$

Step 2. 水印信息重构的低频特征图像 $\ddot{Q}^* = \left\{\ddot{q}_{ij}^* \mid i = 1,2,...,m/2,\ j = 1,2,...,n/2\right\}$.首先将 W_1^* 和 W_2^* 分为 2×2 的图像块,然后将每个 2×2 的图像块按照下面的方法转换为一个十进制数,生成水印信息恢复的加密矩阵 $E^* = \left\{e_{ij}^* \mid i = 1,2,...,m/2,\ j = 1,2,...,n/2\right\}$:

$$e_{ij}^* = d1_{ij} + d2_{ij} \tag{12}$$

其中,

$$\begin{cases} d1_{ij} = 8w1^*_{(i\times 2-1,\ j\times 2-1)} + 4w1^*_{(i\times 2-1,\ j\times 2)} + 2w1^*_{(i\times 2,\ j\times 2-1)} + w1^*_{(i\times 2,\ j\times 2)} \\ d2_{ij} = 16\times\left(w2^*_{(i\times 2-1,\ j\times 2-1)} + w2^*_{(i\times 2-1,\ j\times 2)} + w2^*_{(i\times 2,\ j\times 2-1)} + w2^*_{(i\times 2,\ j\times 2)}\right) \end{cases} \tag{13}$$

根据密钥 k 对 E^*解密(反置乱)生成保存在水印信息中的低频特征图像 $\ddot{Q}^*$,即

$$\ddot{Q}^* = S^{-1}(E^*, k) \tag{14}$$

其中,$S^{-1}(.)$代表置乱加密的反函数.

Step 3. 按照第 2.1 节的 Step 1 与 Step 2,计算得到被测图像的低频特征图像 $Q^* = \left\{q_{ij}^* \mid i = 1,2,...,m/2,\ j = 1,2,...,n/2\right\}$.

Step 4. 定义差值图像 $\Delta Q = \left|Q^* - \ddot{Q}^*\right|$,根据 ΔQ 和给定阈值 T(第 2.3 节讨论阈值的选取)识别内容被篡改的图像块,并对其进行相应的篡改恢复操作.设 t_{ij} 为差值图像中像素 Δq_{ij} 的 8 邻域(记作 $N_{\_8}(\Delta q_{ij})$)中包含非零点的个数(图像边界补 0),则:

① 篡改定位:如果 $\Delta q_{ij} \neq 0$ 且 $t_{ij} \geq T$,则判定 2×2 的图像块 A_{ij}^* 被篡改,否则,判定 2×2 的图像块 A_{ij}^* 没有被篡改.

② 篡改恢复:对上一步判定为篡改的图像块 A_{ij}^*,利用 $\ddot{q}_{ij}^*$ 对其进行篡改恢复.即令 A_{ij}^* 中 4 个像素的值都等于 $\ddot{q}_{ij}^* \times p + p/2$,其中 p 为均值量化步长.

Step 5. 后处理(post-processing):对所有 $\Delta q_{ij} \neq 0$ 的图像块 A_{ij}^*,分别计算 A_{ij}^* 的均值 $Ave(A_{ij}^*)$ 和与该图像块相邻

的 8 个图像块的平均值(记为 $Ave(N_{-8}(A_{ij}^*))$),如果 $\left|Ave(N_{-8}(A_{ij}^*))-Ave(A_{ij}^*)\right|\geq 10$,则 A_{ij}^* 中 4 个像素的值修改为 $Ave(N_{-8}(A_{ij}^*))$.

篡改检测与恢复算法中的后处理(Step 5)是为了消除恢复图像中可能存在的噪声点而设置的.当被测图像中有水印信息被篡改时,本文算法根据阈值 T 能够有效识别被篡改图像块并对其进行篡改恢复.然而,根据阈值检测篡改难免会出现虚警和漏警.也就是说,被测图像中可能存在真实图像块被误判为篡改(虚警),或虚假图像块没有被检测出来(漏警)(第 2.3 节将讨论算法的虚/漏警概率).此外,被测图像中也可能存在这样的图像块:其内容和保存在其他图像块中的水印信息同时被篡改.这些情况都会导致图像块不能被有效恢复,从而造成恢复图像中存在噪声点.因此,该步骤的使用能够进一步提高篡改恢复图像的质量,同时也在一定程度上弥补了算法对小区域篡改(如添加随机噪声等)漏警概率高的缺点(第 2.3 节指出的).下面,我们通过考察篡改检测算法的虚/漏警概率来讨论阈值 T 的选取.

2.3 域值分析

根据图像块 A_{ij}^* (i=1,2,…,m/2,j=1,2,…,n/2)是否被篡改,被测图像中的图像块可分为 H_0 和 H_1 两个集合,即如果图像块 A_{ij}^* 被篡改,则 $A_{ij}^*\in H_0$;否则, $A_{ij}^*\in H_1$.设 P_{H_0} 和 P_{H_1} 分别表示被测图像中图像块被篡改和没有篡改的比例,显然,

$$\begin{cases}P_{H_0}+P_{H_0}=1\\ H_0\cap H_1=\varnothing\end{cases} \tag{15}$$

其中,Φ为被测图像中所有图像块的集合.对给定的阈值 T,算法的篡改检测性能可以通过漏警概率(probability of false acceptance)$P_{fa}(T)$和虚警概率(probability of false rejection)$P_{fr}(T)$来衡量.本算法中,漏警概率是指图像块被篡改而算法未能检测出该篡改的概率;虚警概率是指图像块没有被篡改,而篡改检测算法误判该图像块被篡改的概率.由篡改检测与恢复算法的 Step 4 可知,本文篡改检测算法的漏警概率和虚警概率分别为

$$P_{fa}(T)=P\left\{\Delta q_{ij}=0\mid A_{ij}^*\in H_0\right\}+P\left\{\Delta q_{ij}\neq 0\mid A_{ij}^*\in H_0\right\}P\left\{t_{ij}<T\mid A_{ij}^*\in H_0\right\} \tag{16}$$

$$P_{fr}(T)=P\left\{\Delta q_{ij}\neq 0\mid A_{ij}^*\in H_1\right\}P\left\{t_{ij}\geq T\mid A_{ij}^*\in H_1\right\} \tag{17}$$

理想情况下,阈值 T 应使虚警概率和漏警概率均趋向于 0,下面从概率论的角度讨论虚/漏警概率与阈值的关系.为此,我们首先给出概率论中的一个基本定理.

定理 1[16]. 如果随机变量 t 服从参数为(n,p)的二项分布,则随机变量 t 小于 T 的概率为

$$P(t<T)=\sum_{i=0}^{T-1}C_n^i p^i(1-p)^{n-i} \tag{18}$$

其中, C_n^i 代表从 n 个元素中任取 i 个的组合数.

利用本文算法生成的含水印图像中,每个含水印图像块可分为图像内容和嵌入的水印信息两部分,即 $A_{ij}^*=\tilde{A}_{ij}^*+\ddot{A}_{ij}^*$,该图像块相应的水印信息嵌入在其他图像块的低有效位 $\ddot{A}_{i'j'}^*$.设 $P_{C|H_0}$,$P_{L|H_0}$ 和 $P_{W|H_0}$ 分别表示 $A_{ij}^*\in H_0$ 时 $\tilde{A}_{ij}^*$,$\ddot{A}_{ij}^*$和 $\ddot{A}_{i'j'}^*$ 发生变化的概率,设 $P_{C|H_1}$,$P_{L|H_1}$ 和 $P_{W|H_1}$ 分别表示 $A_{ij}^*\in H_1$ 条件下 $\tilde{A}_{ij}^*$,$\ddot{A}_{ij}^*$ 和 $\ddot{A}_{i'j'}^*$ 发生变化的概率.显然,

$$P_{C|H_1}=P_{L|H_1}=0 \tag{19}$$

在可恢复水印算法中,基于图像内容生成的水印信息经过置乱加密后才嵌入图像块的低有效位,因此无论图像块是否被篡改,其相应水印信息被篡改的概率都近似相同且与被测图像中低位被篡改的概率成正比,因此,被测图像中每个图像块的相应水印信息被改变的概率 P_W 相等且等于

$$P_W=P_{W|H_0}=P_{W|H_1}=P_{L|H_0}P_{H_0}+P_{L|H_1}P_{H_1}=P_{L|H_0}P_{H_0} \tag{20}$$

下面分 4 种情况来讨论 $q_{ij}^*\neq\ddot{q}_{ij}^*$ 的概率($\tilde{A}_{ij}^*$ 与 $\ddot{A}_{i'j'}^*$ 分别表示含水印图像块的内容和水印信息,q_{ij} 表示原始图像生成的量化值):

(1) $\tilde{A}_{ij}^*$ 和相应的水印信息 $\ddot{A}_{i'j'}^*$ 都没有改变.由水印算法可知,如果 $\tilde{A}_{ij}^*$ 和 $\ddot{A}_{i'j'}^*$ 都没有被篡改,则 $q_{ij}^*=\ddot{q}_{ij}^*=q_{ij}$,即

$$P_{00}\left\{q_{ij}^{*}=\ddot{q}_{ij}^{*}\right\}=0 \tag{21}$$

(2) $\tilde{A}_{ij}^{*}$ 不改变而 $\ddot{A}_{i'j'}^{*}$ 改变.$\ddot{A}_{i'j'}^{*}$ 改变意味着 $\ddot{q}_{ij}^{*}\neq q_{ij}$,而$q_{ij}^{*}=q_{ij}$,因此图像块内容不变而相应水印信息改变时一定有$q_{ij}^{*}\neq\ddot{q}_{ij}^{*}$,即

$$P_{01}\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\right\}=1 \tag{22}$$

(3) $\tilde{A}_{ij}^{*}$ 改变而 $\ddot{A}_{i'j'}^{*}$ 没有改变.$\ddot{A}_{i'j'}^{*}$ 没有改变意味着 $\ddot{q}_{ij}^{*}\neq q_{ij}$.$\tilde{A}_{ij}^{*}$ 改变时,其相应的量化值 q_{ij}^{*} 可能相等也可能不等,如果 $q_{ij}^{*}=q_{ij}$,则意味着篡改前、后的图像块差别不大(第 3.2 节分析图像块内容篡改量与 $\ddot{q}_{ij}^{*}\neq q_{ij}$ 的关系),因此从保证图像块真实性的角度,可以认为

$$P_{10}\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\right\}\approx 1 \tag{23}$$

(4) $\tilde{A}_{ij}^{*}$ 和 $\ddot{A}_{i'j'}^{*}$ 都改变.此时, q_{ij}^{*} 与 $\ddot{q}_{ij}^{*}$ 相同的概率与其取值空间有关.因为 $q_{ij}\in[0,31]$,因此,

$$P_{11}\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\right\}=1-\binom{32}{1}\left(\frac{1}{32}\right)^{2}=1-1/32=31/32 \tag{24}$$

设 $P_{U|H_0}$ 和 $P_{U|H_1}$ 分别表示 $A_{ij}^{*}\in H_0$ 和 $A_{ij}^{*}\in H_1$ 条件下$\Delta q_{ij}\neq 0$($q_{ij}^{*}\neq\ddot{q}_{ij}^{*}$)的概率,由全概率公式可知:

$$\begin{aligned}P_{U|H_0}&=P\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\mid A_{ij}^{*}\in H_0\right\}\\&=\left(1-P_{C|H_0}\right)\left(1-P_{W|H_0}\right)P_{00}\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\right\}+\left(1-P_{C|H_0}\right)P_{W|H_0}P_{01}\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\right\}+\\&\quad P_{C|H_0}\left(1-P_{W|H_0}\right)P_{10}\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\right\}+P_{C|H_0}P_{W|H_0}P_{11}\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\right\}\\&\approx\left(1-P_{C|H_0}\right)P_{W|H_0}+P_{C|H_0}\left(1-P_{W|H_0}\right)+P_{C|H_0}P_{W|H_0}\left(31/32\right)\\&=P_W+P_{C|H_0}-33P_WP_{C|H_0}/32\end{aligned} \tag{25}$$

相应地,

$$\begin{aligned}P_{U|H_1}&=P\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\mid A_{ij}^{*}\in H_1\right\}\\&=\left(1-P_{C|H_1}\right)\left(1-P_{W|H_1}\right)P_{00}\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\right\}+\left(1-P_{C|H_1}\right)P_{W|H_1}P_{01}\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\right\}+\\&\quad P_{C|H_1}\left(1-P_{W|H_1}\right)P_{10}\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\right\}+P_{C|H_1}P_{W|H_1}P_{11}\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\right\}\\&=P_{W|H_1}P_{01}\left\{q_{ij}^{*}\neq\ddot{q}_{ij}^{*}\right\}\\&=P_W\end{aligned} \tag{26}$$

显然, $P_{U|H_0}$ 和 $P_{U|H_1}$ 均与被测图像中水印信息被篡改的量有关, $P_{U|H_0}$ 还与图像块被篡改时图像内容被篡改的可能性有关.

假设被测图像的某区域被篡改,设该区域与整个图像的比例为ρ且篡改区域中的像素被随机篡改(即像素的每个 bit 独立且其是否变化的可能性相同),在此条件下,

$$\begin{cases}P_{H_0}\approx\rho\\P_{L|H_0}=1-P_{=0}\left(5,0.5\right)=1-1/2^{5}=31/32\\P_{C|H_0}=1-P_{=0}\left(4\times 8-5,0.5\right)=1-1/2^{27}\approx 1\end{cases} \tag{27}$$

将式(27)代入式(20)得:

$$P_W=P_{L|H_0}P_{H_0}=31\rho/32 \tag{28}$$

将式(27)分别代入式(25)和式(26)得:

$$P_{U|H_0}=P_W+P_{C|H_0}-33P_WP_{C|H_0}/32=1-31\rho/32^{2} \tag{29}$$

$$P_{U|H_1}=P_W=31\rho/32 \tag{30}$$

对任意一个图像块 $A_{ij}^{*}\in H_0$,假定其 8 邻域 $N_{-8}(A_{ij}^{*})$ 中每个图像块均被篡改(即忽略了篡改边界的情况),该

条件下 t_{ij} 服从参数为 $\left(8,P_{U|H_0}\right)$ 的二项分布,由定理 1 可知:

$$P(t_{ij}<T\mid A_{ij}^{*}\in H_0)=\sum_{t=0}^{T-1}C_8^t\left(P_{U|H_0}\right)^t\left(1-P_{U|H_0}\right)^{8-t} \tag{31}$$

相应地,

$$P(t_{ij}<T\mid A_{ij}^{*}\in H_1)=\sum_{t=0}^{T-1}C_8^t\left(P_{U|H_1}\right)^t\left(1-P_{U|H_1}\right)^{8-t} \tag{32}$$

将式(29)和式(31)代入式(16)得到本文篡改检测算法的漏检概率为

$$\begin{aligned}P_{fa}(T)&=P\left\{\Delta q_{ij}=0\mid A_{ij}^{*}\in H_0\right\}+P\left\{\Delta q_{ij}\neq 0\mid A_{ij}^{*}\in H_0\right\}P\left\{t_{ij}<T\mid A_{ij}^{*}\in H_0\right\}\\&=\left(1-P_{U|H_0}\right)+P_{U|H_0}\sum_{t=0}^{T-1}C_8^t\left(P_{U|H_0}\right)^t\left(1-P_{U|H_0}\right)^{8-t}\\&=31\rho/32^2+\left(1-31\rho/32^2\right)\sum_{t=0}^{T-1}C_8^t\left(1-31\rho/32^2\right)^t\left(31\rho/32^2\right)^{8-t}\end{aligned} \tag{33}$$

将式(30)和式(32)代入式(17)得到本文篡改检测算法的漏检概率为

$$\begin{aligned}P_{fr}(T)&=P\left\{\Delta q_{ij}\neq 0\mid A_{ij}^{*}\in H_1\right\}P\left\{t_{ij}\geq T\mid A_{ij}^{*}\in H_1\right\}\\&=P_W\left(1-P\left\{t_{ij}<T\mid A_{ij}^{*}\in H_1\right\}\right)\\&=P_W\left(1-\sum_{t=0}^{T-1}C_8^t\left(P_{U|H_1}\right)^t\left(1-P_{U|H_1}\right)^{8-t}\right)\\&=(31\rho/32)\left(1-\sum_{t=0}^{T-1}C_8^t(31\rho/32)^t(1-31\rho/32)^{8-t}\right)\end{aligned} \tag{34}$$

由式(33)和式(34)可以看出,本文算法的虚/漏警概率与阈值 T 和被测图像中被篡改区域的比例 ρ 有关.图 4 分别给出了 ρ=0.2,0.3,0.4 时,本文篡改检测算法的虚/漏警概率随阈值变化的曲线,图 4 中横坐标为域值 T,纵坐标为虚/漏警概率 P_{fr}/P_{fa}.由图 4 可以看出,虚/漏警概率随篡改比例 ρ 的增大而增大,但是,ρ 对漏警概率的影响较小.漏警概率 P_{fa} 随阈值 T 的增大而增大,虚警概率 P_{fr} 随阈值 T 的增大而减小.考虑到攻击者对认证水印信息的篡改量不会太大[12],兼顾虚警概率和漏警概率,本文取阈值 T=5.

需要说明的是,即使 T 为 0,算法的漏警概率仍然不为 0,这是由 $P_{U|H_0}$ 的值小于 1 造成的,可以通过增加图像块中嵌入的水印信息量来减少算法的漏警概率.同时,上述推导给出的漏警概率是在图像块的 8 邻域均被篡改的条件下得到的,因此,对篡改区域边界和随机篡改,算法的漏警概率将增大.但是,由于本文算法的图像块较小(2×2),使得漏警概率略大于 0 也是可以接受的,而且篡改检测与恢复算法的 Step 5 能够在一定程度上弥补该缺点.

3 性能分析

3.1 不可见性

脆弱水印要求加入的水印不可觉察,为了衡量水印图像与原始图像之间的差别,定义峰值信噪比 PSNR (peak signal-to-noise ratio)为

$$PSNR=10\log 10\left[\frac{255\times 255}{\frac{1}{m\times n}\sum_{i=1}^{m}\sum_{j=1}^{n}[A(i,j)-A^{w}(i,j)]^2}\right] \tag{35}$$

该算法将水印嵌入到图像的最低位和 1/4 次低位,因为原始图像的最低位和 1/4 次低位的每个比特是独立的,所以 $[A\{i,j\}-A^{w}(i,j)]^2$ 的数学期望为

$$E\left(\left[A(i,j)-A^{w}(i,j)\right]^2\right)=\frac{3}{4}\times\left(\frac{1}{2}+\frac{0}{2}\right)+\frac{1}{4}\times\left(0\times\frac{4}{16}+1\times\frac{6}{16}+4\times\frac{4}{16}+9\times\frac{2}{16}\right)=1 \tag{36}$$

原始图像与水印图像峰值信噪比的数学期望为

$$E(PSNR)=10\log 10\left[\frac{255\times 255}{\frac{1}{m\times n}\sum_{i=1}^{m}\sum_{j=1}^{n}E([A(i,j)-A^{w}(i,j)]^{2})}\right]=48.1308\text{dB} \tag{37}$$

可见,在最低位和 1/4 次低位嵌入水印可满足脆弱水印的不可见要求.

3.2 篡改量与量化值变化的关系

在讨论 $P_{U|H_0}$ 和 $P_{U|H_1}$ 时,我们忽略了图像块内容改变而相应量化值不变的情况.下面我们结合量化算法,讨论不引起量化值变换($q_{ij}^{*}=q_{ij}$)的篡改对图像语义的影响.

按式(6)的量化方法将区间[min,max]划分为 N_p 个互不相交的子集,设

$$\begin{cases}E_x=\left\{\tilde{a}_{i,j}^{1}\mid \tilde{a}_{i,j}^{1}\in[\min+xp,\min+(x+1)p)),x=0,1,...,N_p-2\right\}\\ E_{N_p-1}=\left\{\tilde{a}_{i,j}^{1}\mid \tilde{a}_{i,j}^{1}\in\left[\min+(N_q-1)p,\max\right]\right\}\end{cases} \tag{38}$$

显然,

$$\begin{cases}E_0\cup E_1\cup...\cup E_{N_q-1}=[\min,\max]\\ E_{r_1}\cap E_{r_2}=\varnothing,\forall r_1\neq r_2\end{cases} \tag{39}$$

因此, $E_0,E_1,...,E_{N_q-1}$ 可看作样本空间[min,max]的一个划分.

事件 B_Δ 为低频系数 $\tilde{a}_{i,j}^{1}$ 改变Δ(Δ≥0)时,量化值发生改变(即 $q_{ij}^{*}\neq q_{ij}$)的概率.由全概率公式可知,

$$P(B_\Delta)=\sum_{x=0}^{N_p-1}P(B_\Delta\mid E_x)P(E_x) \tag{40}$$

假设图像内容的低频系数 $\tilde{a}_{i,j}^{1}$ 在[min,max]上服从均匀分布,易知

$$P(E_x)=P\left(\tilde{a}_{i,j}^{1}\mid \tilde{a}_{i,j}^{1}\in[\min+xp_x,\min+(x+1)p)\right)=\frac{p}{\max-\min} \tag{41}$$

$P(B_\Delta/E_x)$ 为 $\tilde{a}_{i,j}^{1}\in E_x$ 条件下, $\tilde{a}_{i,j}^{1}$ 改变Δ时 $q_{ij}^{*}\neq q_{ij}$ 的概率.下面我们来求此概率.

设篡改后的小波低频系数 $\tilde{a}_{i,j}^{1\,*}=\tilde{a}_{i,j}^{1}\pm\Delta$, 在 $\tilde{a}_{i,j}^{1}\in E_x$ 即 $\tilde{a}_{i,j}^{1}\in[\min+xp,\min+(x+1)p)$ 的条件下,当 $\tilde{a}_{i,j}^{1\,*}\geq\min+(x+1)q$ 或 $\tilde{a}_{i,j}^{1\,*}<\min+xq$ 时, $q_{ij}^{*}\neq q_{ij}$,假设小波系数变大或变小的概率相等,可得:

$$\begin{aligned}P(B_\Delta\mid E_x)&=0.5P\left(\left(\tilde{a}_{i,j}^{1\,*}\geq\min+(x+1)p\right)\mid E_x\right)+0.5P\left(\left(\tilde{a}_{i,j}^{1\,*}<\min+xp\right)\mid E_x\right)\\ &=0.5P\left(\left(\tilde{a}_{i,j}^{1}+\Delta\geq\min+(x+1)p\right)\mid E_x\right)+0.5P\left(\left(\tilde{a}_{i,j}^{1}-\Delta<\min+xp\right)\mid E_x\right)\\ &=0.5P\left(\left(\tilde{a}_{i,j}^{1}\geq\min+(x+1)p-\Delta\right)\mid E_x\right)+0.5P\left(\left(\tilde{a}_{i,j}^{1}<\min+xp+\Delta\right)\mid E_x\right)\\ &=0.5\times(\Delta/p)+0.5\times(\Delta/p)\\ &=\Delta/p\end{aligned} \tag{42}$$

将式(41)和式(42)的结果代入式(47)得:

$$\begin{aligned}P(B_\Delta)&=\sum_{x=0}^{N_p-1}P(B_\Delta/E_x)P(E_x)\\ &=\sum_{x=0}^{N_p-1}\frac{\Delta}{p}\times\frac{p}{\max-\min}\\ &=\frac{N_p\times\Delta}{\max-\min}\\ &=\frac{\Delta}{p}\end{aligned} \tag{43}$$

其中,p=(max−min)/N_p 为量化步长.

因此,当图像块的内容被篡改,小波低频系数的改变量为Δ时, $q_{ij}^* \neq q_{ij}$ 的概率为

$$P(B_\Delta)=\begin{cases}\Delta/p, & \Delta \le p\\ 1, & \Delta > p\end{cases} \tag{44}$$

图 5 给出了小波低频系数改变量与篡改检测概率的关系图.实验中采用 DB1 小波基,分别以 256×256 的 Lena 和 Woman 灰度图像为测试对象.对每一幅测试图像篡改 30 次,每次图像内容被篡改的区域为 200×180,则图像内容被篡改的总像素数为 200×180×30,统计并分别计算出当小波低频系数改变量为Δ时,其量化值被改变的比率.仿真结果分别如图 5 中的 Lena 和 Woman 曲线所示,图中仅给出Δ∈[1,11]时,篡改量与量化值变化率之间的关系曲线,Δ≥11 时量化值变化率恒为 1.

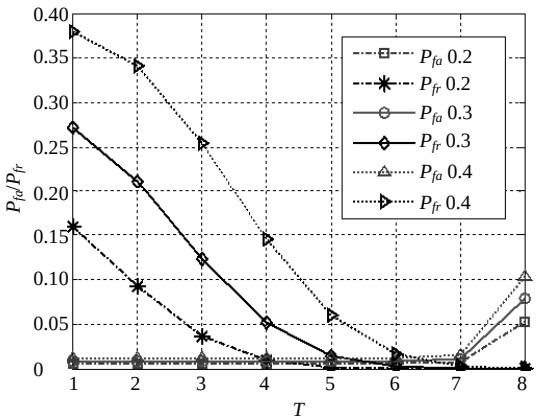


Fig.4 $P_{fr}(T)$ and $P_{fa}(T)$ curves for different ρ

图 4 ρ不同时的 $P_{fr}(T)$ 和 $P_{fa}(T)$曲线图

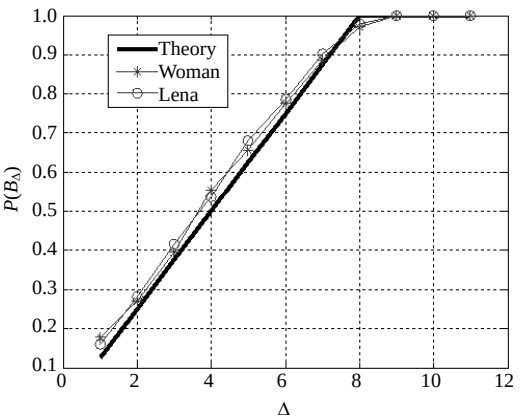


Fig.5 Relationship between $P(B_\Delta)$ with Δ

图 5 Δ与 $P(B_\Delta)$之间的关系

理论上,含水印图像的内容其有效位为 7bits,故含水印图像内容的取值范围为 0~2^7−1,使用 DB1 小波基对其进行小波变换得到低频系数的取值范围为 0~254,因此均值量化步长 p≤8.当小波低频系数改变量Δ≤p 时,根据式(44)计算可得图 5 中的“Theory”曲线.由图 5 可以看出,实验值和理论推导值相符合.

为衡量不改变量化值的篡改对被测图像语义的影响,我们用式(35)定义的峰值信噪比来衡量篡改前、后低频系数的差别:

$$PSNR=10\log 10\left[\frac{254\times 254}{\frac{1}{m\times n}\sum_{i=1}^{m}\sum_{j=1}^{n}\Delta^2}\right]\ge 10\log 10\left[\frac{254\times 254}{\frac{1}{m\times n}\sum_{i=1}^{m}\sum_{j=1}^{n}8^2}\right]=30.0349\text{dB} \tag{45}$$

显然,不改变量化值的篡改对被测图像语义的改变不大.

上述理论分析和实验结果表明,当图像内容的篡改量小于量化步长时, $q_{ij}^* \neq q_{ij}$ 的概率与篡改量成正比;当图像内容的篡改量大于或等于量化步长时, $q_{ij}^* \neq q_{ij}$ 的概率为 1.这既保证了篡改检测的可靠性,又提高了算法对偶然篡改的鲁棒性.

4 实验仿真

为了验证本文算法的定位精确度和对水印信息篡改的鲁棒性,利用 Matlab 对大量图像进行仿真,下面给出部分仿真结果.仿真结果中的被测图像均为灰度图像.

4.1 低频压缩图像

基于小波低频系数的量化生成水印,不仅可以保存原始图像的大量信息,而且可以保持原始图像的时-频特

性.图 6 给出 Lena 和 Car 原始图像及其相应的低频特征图像.从图 6 中可以看出,低频特征图像在水平和垂直方向都平滑逼近原始图像,即使图 6(c)中像“车牌号”这样的细节,低频特征图像也能保持原始图像的意义.

(a) Original image of Lena (a) Lena 原始图像 (b) Compressed low-frequency image of Lena (b) Lena 低频压缩图像 (c) Original image of Car (c) Car 原始图像 (d) Compressed low-frequency image of Car (d) Car 低频压缩图像

Fig.6 Compressed low-frequency image of different images

图 6 不同图像的低频压缩图像

4.2 篡改区分及恢复

为了验证本文算法在部分水印信息篡改时的篡改定位及篡改恢复能力,图 7 给出了区域篡改情况下本文算法与文献[9]算法的篡改定位及篡改恢复比较,其中(a)为原始图像;(b)为利用本文算法生成的含水印图像,与原始图像的峰值信噪比(PSNR)为 48.14dB,与理论分析结果 E(PSNR)=47.16dB 基本一致,验证了本文算法能够满足脆弱水印不可见性的要求.使用 Photoshop 编辑软件,直接在含水印图像上添加一个酒杯,篡改图像与含水印图像的峰值信噪比为 24.86dB,该篡改图像为图 7(c),记为篡改_1.图 7(d)为第 2 个篡改图像,记为篡改_2.篡改_2 与篡改_1 的篡改区域相同,但是仅仅修改该区域最低位,与含水印图像的峰值信噪比为 63.76dB.显然,篡改_1 被严重篡改,而篡改_2 与含水印图像的语义相同.

图 8 分别给出本文与张鸿宾[9]算法对篡改_1 和篡改_2 图像的篡改检测与恢复结果.对篡改_1 图像,本文算法的篡改检测与恢复结果分别为图 8 中(a)和(b).(b)不仅准确地定位出了篡改_1 图像被篡改的位置,而且被篡改的图像内容也有效地恢复出来.该恢复图像与含水印图像的峰值信噪比为 63.76dB.图 8(e)和图 8(f)是张鸿宾等人[9]算法的篡改检测与篡改恢复结果.可以看出,在篡改区域外存在被误检的图像块(被误检图像块在图像中的分布与密钥有关,本图为密钥等于 14 和 137 时的篡改检测和恢复结果),因此得到的恢复图像中也存在被误恢复的图像块.图 8(f)与含水印图像的峰值信噪比为 20.90dB.

(a) Original image (a) 原始图像 (b) Watermarked image (b) 含水印图像 (c) Tamper_1 (c) 篡改_1 (d) Tamper_2 (d) 篡改_2

Fig.7 Watermarked and two tampered images

图 7 含水印图像与两个篡改图像

对篡改_2 图像,本文算法的篡改检测与恢复结果分别为图 8 中(c)和图 8(d),张鸿宾等人[9]算法的篡改检测与篡改恢复结果为图 8 中(g)和图 8(h)所示.正如我们所希望的,我们的算法能够识别出篡改_2 图像的内容没有被严重篡改,因此得到的篡改恢复图像与篡改_2 相同.与之相对应,张鸿宾等人[9]的算法不能识别水印信息篡改,从而导致一些图像被误检测,并用被篡改的水印信息恢复真实图像块的内容,从而严重影响恢复图像的质量.

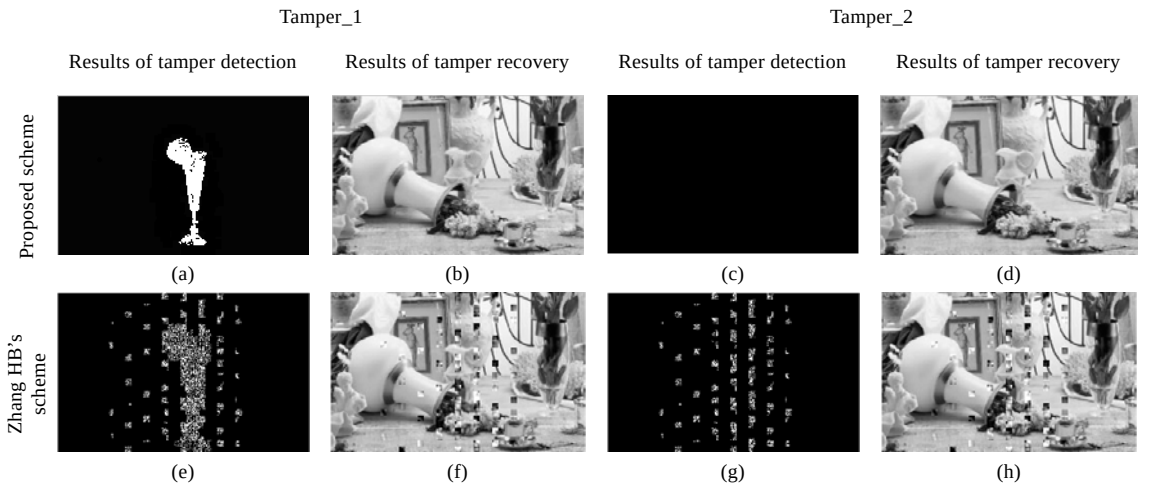


Fig.8 Comparison of detection and recovery results for two tampered images

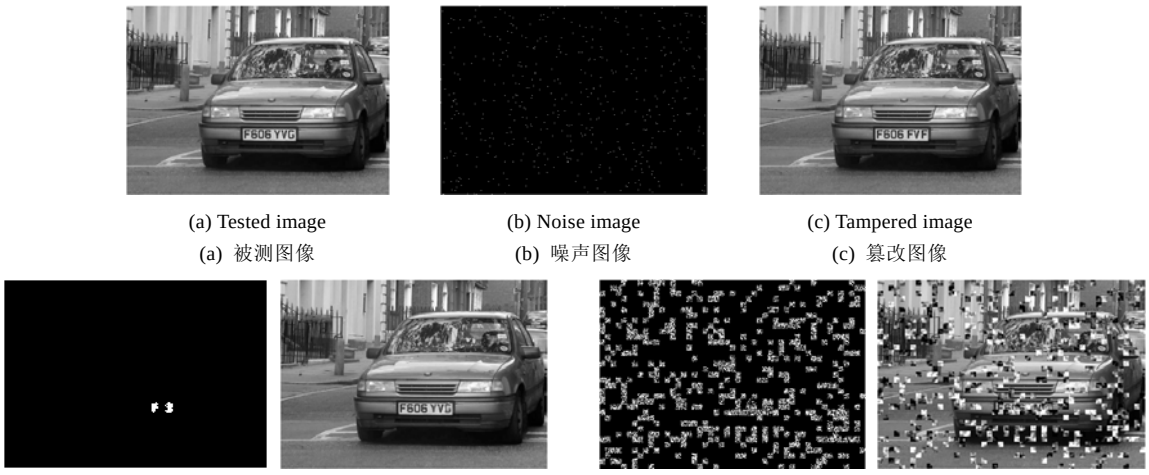
图 8 两个不同篡改图像检测与恢复结果比较

由图 8 的实验结果可以看出,本文算法的定位精度高于文献[9]的定位精度.更重要的是,本文算法根据给定的阈值,能够区分图像内容和水印信息篡改,对图像内容篡改的图像块进行篡改恢复,水印信息篡改的图像块不进行篡改恢复,大幅度提高了被测图像中有水印信息篡改时的篡改恢复质量.相反地,现有分块自嵌入水印算法不能区分图像内容和水印信息篡改,因此,算法对图像内容和水印信息篡改的图像块都进行篡改恢复,致使篡改恢复图像中存在一些被误恢复的图像块,出现方块效应,严重降低了算法的恢复质量.

4.3 随机噪声

为了验证本文算法抵抗随机噪声攻击的能力,图 9 给出了被噪声污染且包含小区域篡改的检测与恢复结果,其中,(a)为被测图像,(b)为噪声图像.对图 9(a)生成的含水印图像实施如下篡改:首先在含水印图像上添加随机噪声(如图(b),噪声比例为 0.004 7),并将车牌号“F606YVG”修改为“F606FVF”.篡改后的图像如图 9(c)所示,该篡改图像与含水印图像的峰值信噪比为 35.1985dB.图 9(d)为本文算法的篡改检测与篡改恢复结果,图 9(f)为张鸿宾等人[9]算法的篡改检测与篡改恢复结果.由图示可以看出,本文算法的能将恶意篡改(即使像“车牌号”这样小区域)与随机噪声区分开,使得篡改恢复操作仅对恶意篡改的图像块实施,从而得到高质量的篡改恢复图像.相比之下,由于现有分块自嵌入水印算法很难将小区域的恶意篡改与随机噪声区分开,从而对被检测到的图像块都进行篡改恢复,使得篡改恢复图像中出现较严重的方块效应,难以得到质量较好的篡改恢复图像.

从图 7~图 9 的仿真结果可以看出,本文算法在含水印图像中的部分水印信息被篡改时(无论随机篡改还是区域篡改),算法均能根据阈值识别出被恶意篡改的图像块并对其进行有效的篡改恢复,有效地提高了自嵌入算法的对水印信息篡改的鲁棒性.



(a) Tested image (b) Noise image (c) Tampered image

(a) 被测图像 (b) 噪声图像 (c) 篡改图像

(d) Detection and recovery results by our scheme (e) Detection and recovery results in Ref.[9]

(d) 本文算法的检测和恢复结果 (e) 文献[9]的检测和恢复结果

Fig.9 Tamper detection and reconstructed results of randomly tamper

图 9 随机篡改时的篡改检测与恢复结果

5 结 论

本文提出一种对水印信息篡改鲁棒的自嵌入脆弱水印算法,讨论了该算法的阈值选取和检测篡改的可靠性.本文基于一级小波分解的低频系数生成水印信息,将 DWT 低频系数生成的量化图像置乱加密后嵌入图像的低有效位平面.认证时,通过设定域值识别出内容被恶意篡改的图像块并对其实施篡改恢复.与现有自嵌入水印算法相比,本文具有以下优点:

(1) 提高了自嵌入水印算法的定位精度;

(2) 提高了自嵌入水印算法对信道噪声和偶然篡改的鲁棒性;

(3) 大幅度提高了被测图像中有水印信息篡改时的篡改恢复图像的质量.

如何提高可恢复水印算法对 JPEG 压缩的鲁棒性是我们下一步的研究内容.

本文工作已获专利授权(专利号:ZL 2004 1 0081400.6).

References:

[1] Zhu BB, Swanson MD, Tewfik AH. When seeing isn't believing. IEEE Signal Processing Magazine, 2004(3):40−49.

[2] Suthaharan S. Fragile image watermarking using a gradient image for improved localization and security. Pattern Recognition Letters, 2004,25(16):1893−1903.

[3] Izquierdo E, Guerra V. An ill-posed operator for secure image authentication. IEEE Trans. on Circuits and Systems for Video Technology, 2003,13(8):842−852.

[4] He HJ, Zhang JS, Tai HM. A secure fragile watermarking scheme for image authentication. In: Cheung YC, ed. Proc. of the 2006 Int'l Conf. on Computational Intelligence and Security. New York: IEEE, 2006. 1180−1185.

[5] He HJ, Zhang JS, Tian L. A fragile watermarking scheme with discrimination of tampers on image or watermark. Acta Electronica Sinica, 2005,33(9):1557−1561 (in Chinese with English abstract).

[6] Kundur D, Hatzinakos D. Digital watermarking for telltale tamper proofing and authentication. Proc. of the IEEE, 1999,87(7): 1167−1180.

[7] Fridrich J, Goljan M. Images with self-correcting capabilities. In: Proc. of the ICIP'99. Kobe, 1999. 792−796.

[8] Lin PL, Huang PW, Peng AW. A fragile watermarking scheme for authentication with localization and recovery. In: Proc. of IEEE the 6th Int'l Symp. on Multimedia Software Engineering (ISMSE 2004). IEEE Computer Society, 2004. 146−153.

[9] Zhang HB, Yang C. Tamper detection and self recovery of image using self-embedding. Acta Electronica Sinica, 2004,32(2): 196−199 (in Chinese with English abstract).

[10] He HJ, Zhang JS. Chaos-Based scramble self-embedding watermarking algorithm. Journal on Communications, 2006,27(7):80−87 (in Chinese with English abstract).

[11] Holliman M, Memon N. Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes. IEEE Trans. on Image Processing, 2000,3(9):432−441.

[12] Fridrich J, Goljan M, Memon N. Cryptanalysis of the Yeung-Mintzer fragile watermarking technique. Journal of Electronic Imaging, 2002,11(4):262−274.

[13] HE HJ, Chen F. On the security of the self-embedding watermarking scheme. Acta Electronica Sinica, 2007,35(3):557−562 (in Chinese with English abstract).

[14] Zhang XD, Lu GD, Feng J. Fundamentals of Image Coding and Wavelet Compressing: Principles, Algorithms and Standards. Beijing: Tsinghua University Press, 2004. 235−270 (in Chinese).

[15] Romberg JK, Choi H, Baraniuk RG. Bayesian tree-structured image modeling using wavelet-domain hidden Markov models. IEEE Trans. on Image Processing, 2001,10(7):1056−1068.

[16] Li YQ. The Theory of Probability and Statistics. Beijing: National Defense Industry Press, 2001. 54−84 (in Chinese).

附中文参考文献:

[5] 和红杰,张家树,田蕾.能区分图像或水印篡改的脆弱水印方案.电子学报,2005,33(9):1557−1561.

[9] 张鸿宾,杨成.图像的自嵌入及篡改的检测和恢复算法.电子学报,2004,32(2):196−199.

[10] 和红杰,张家树.基于混沌置乱的分块自嵌入水印算法.通信学报,2006,27(7):80−87.

[13] 和红杰,陈帆.自嵌入水印算法的安全性分析.电子学报,2007,35(3):557−562.

[14] 张旭东,卢国栋,冯健.图像编码基础和小波压缩技术——原理、算法和标准.北京:清华大学出版社,2004.235−270.

[16] 李裕奇.概率论与数理统计.北京:国防工业出版社,2001.54−84.

和红杰(1971－),女,河南宝丰人,博士生,讲师,主要研究领域为信息隐藏,数字水印技术.

张家树(1965－),男,博士,教授,博士生导师,主要研究领域为数字图像工程,非线性信号分析与处理,通信理论,信息对抗技术.