用于图像认证的变容量恢复水印算法

陈 帆[1)] 和红杰[2)] 王宏霞[1)]

[1)](西南交通大学信息安全与国家计算网格实验室 成都 610031)

[2)](西南交通大学四川省信号与信息处理重点实验室 成都 610031)

摘 要 兼顾水印嵌入容量和安全性,提出一种水印容量可变的数字图像可恢复水印算法.该算法提取2×2图像块特征生成变容量恢复水印——平滑块6比特,纹理块12比特.图像块的恢复水印基于密钥随机嵌入在其它图像块的低有效位,通过比较图像块特征与相应恢复水印重构的块特征并结合邻域特征判定图像块的真实性.变容量恢复水印用尽可能少的比特数保存足够的图像块信息,仅被嵌入一次且同时用于篡改检测与恢复,不仅有效降低了水印嵌入容量,而且提高了算法抵抗恒均值攻击的能力.实验仿真结果表明,该算法得到的含水印图像和恢复图像的质量好,且能有效抵抗拼贴攻击、恒均值攻击等已知伪造攻击.

关键词 数字图像;恢复水印;篡改检测;水印嵌入容量;变容量

中图法分类号 TP391 **DOI号**: 10.3724/SP.J.1016.2012.00154

Variable-Payload Self-Recovery Watermarking Scheme for Digital Image Authentication

CHEN Fan[1)] HE Hong-Jie[2)] WANG Hong-Xia[1)]

[1)](*Information Security and National Computing Grid Laboratory, Southwest Jiaotong University, Chengdu* 610031)

[2)](*Sichuan Key Laboratory of Signal and Information Processing, Southwest Jiaotong University, Chengdu* 610031)

Abstract This paper proposes an image self-recovery fragile watermarking scheme with variable watermark embedding capacity. Both watermark embedding payload and security are taken into account. For each block of size 2×2 pixels, the features are extracted to generate the recovery watermark with different length——6 bits for a plain block and 12 bits for a texture block. The recovery watermark of an image block is inserted into the less significant bit planes of the other block based on the secret key. The validity of a block is determined by comparing the features computed by the block content with the ones reconstructed by the corresponding recovery watermark incorporating with its neighbor characteristic. The variable-capacity recovery watermark contains the adequate information of image block to as few bits as possible. The recovery watermark is inserted in the original image only once, and used to both tamper detection and tamper recovery in the proposed scheme. These strategies make the watermark embedded payload as low as possible and the ability against the constant-average attack to be improved. Experimental results show that the proposed scheme not only has the better quality of the watermarked and recovered images, but also resists the known counterfeiting attack such as the collage attack and the constant-average attack.

Keywords digital image; self-recovery watermarking; tamper detection; watermark payload; variable-payload

收稿日期:2011-02-21;最终修改稿收到日期:2011-05-09.本课题得到国家自然科学基金(60970122,61170226)、中央高校基本科研业务专项基金(SWJTU09CX039,SWJTU10CX09)以及教育部博士点专项基金(20090184120021)资助.**陈 帆**,男,1971年生,博士研究生,副教授,中国计算机学会(CCF)会员,主要研究方向为多媒体数据安全、数字水印技术.E-mail:mrchenfan@126.com.**和红杰**,女,1971年生,博士,副教授,主要研究方向为数字图像处理、信息隐藏.**王宏霞**,女,1973年生,博士,教授,主要研究领域为信息隐藏、数字水印技术和智能信息处理.

1 引　言

随着信息技术，尤其是网络和数码成像技术的发展，数字图像成为人们获取与交换信息的主要来源和信息传播的重要载体之一. 图像数字化存储和各种图像处理软件的出现使数字图像的编辑、修改等处理变得简单. 数字图像处理技术提高了图像的显示质量. 如果篡改和伪造数字图像被用于新闻媒体、法庭证据、科学发现等领域，对社会的诚信、政府的公信力和科学的真实性等可能带来严重的负面影响. 因此，如何检测和鉴别数字图像的真实性和完整性已成为近年来国内外的前沿性研究课题，不仅具有重要的学术价值，更具有重要的社会意义和广泛的应用前景.

基于数字水印的数字图像真实性鉴定应满足以下几点：① 不可见性. 人的感知系统察觉不到嵌入水印后的数字图像与原始图像的不同，一般用含水印数字图像的峰值信噪比(PSNR)来衡量；② 安全性. 理论上，以 100％的概率检测出对数字图像恶意的改变；③ 盲检测. 对数字图像认证不需要原始图像；④ 篡改检测与定位. 检测并定位数字图像被篡改区域(图像块/像素)；⑤ 篡改恢复. 近似恢复篡改区域的原始内容，为推断篡改方式和攻击者的目的提供依据. 篡改恢复对基于数字水印的数字图像真实性鉴别技术提出更高的要求.

为提高篡改恢复质量，水印算法需准确判定图像块的真实性. 因为篡改块未被检测或真实图像块被误判为篡改，将导致恢复图像中存在未被恢复或被错误恢复的图像块，从而降低图像的恢复质量. 为降低图像块与其恢复水印同时被篡改的可能性，Fridrich 等人[1]指出图像块的恢复水印不能嵌入在图像块自身，而是嵌入在其它图像块中，不过这使篡改检测变得相对困难[2]. 为解决可恢复水印算法的篡改检测问题，Lin 等人[3]提出了一种分层检测可恢复水印算法，除为每个 2×2 图像块生成 6 比特恢复水印(块均值)外，还附加生成 2 比特的认证水印. 图像块的 2 比特认证水印嵌入在图像块自身的低位，用来检测该图像块的真实性. 附加认证水印的思想被许多研究者采用，如文献[4-7]，然而附加认证水印无疑增大了水印嵌入容量，且认证水印的块独立特性使其易受拼贴攻击(collage attack)[8]. 为此，文献[9-10]基于密钥伪随机生成图像块恢复水印在宿主图像中的嵌入位置，通过比较恢复水印的一致性并结合邻域块的分布特性判定图像块的真实性. 在不增加水印嵌入容量的前提下，实现可恢复水印算法的篡改检测.

提高图像恢复质量还需解决同步篡改问题(tampering coincidence problem)[6]，即如何有效恢复同步篡改块(即内容和相应恢复水印同时被破坏的篡改块)的问题. 文献[9-10]无法对“同步篡改块”进行有效恢复，尽管其篡改检测优良，但算法的篡改恢复质量不高. 为有效恢复同步篡改块，研究者提出将恢复水印多次重复嵌入[4-5]在图像中，当一个版本的恢复水印被篡改时，用另外一个版本的恢复水印对篡改块进行恢复. 此外，文献[6-7]利用共享机制解决同步篡改块的恢复问题，其实质是通过增加冗余信息来实现. 如文献[7]将 64 比特块特征扩展成为 160 比特的参考比特(reference bits). 多次重复嵌入和共享机制等策略能提高可恢复水印算法的恢复质量，但算法水印嵌入容量的增加也会降低含水印图像的质量.

此外，恢复水印生成方法也是影响篡改恢复质量和算法安全性的重要因素. 现有生成恢复水印最常用的方法有 2 种：① 8×8 图像块重要 DCT 系数的量化编码[1,7,10]，其缺点是定位精度较低. ② 2×2 图像块的平均值[3-4,9]，定位精度高，但易受恒均值攻击(constant-average attack)[11]. 假设 2 个图像块 $B1=(30,30,30,30)$ 和 $B2=(1,1,59,59)$，它们的平均值相同，因此在利用 2×2 图像块均值生成恢复水印的算法中它们的恢复水印是相同的. 如果用图像块 $B1$ 替换块 $B2$，通过比较恢复水印检测块真实性的算法[9]不能检测该篡改(即不能抵抗恒均值攻击)，而利用附加认证水印检测块真实性的算法[3-4]得到的恢复图像的质量不高，因为恢复水印仅能重构 $B2$ 的均值，而无法重构 $B2$ 的纹理特征. 如果图像块恢复水印除保存块均值特征外还保存块纹理特征，则算法的恢复质量和安全性都会得到提高. 不过，这无疑会增加水印容量，且对类似 $B1$ 这样的平滑块也无需纹理特征. 本文提出的算法思想是对平滑和纹理图像块分别生成不同长度的恢复水印，同时实现恢复水印的变长嵌入和提取，从而用尽可能少的水印容量提高算法的恢复质量和安全性. 目前尚未见到变容量恢复水印算法方面的研究报道.

综上所述，现有可恢复水印算法的水印嵌入容量都是固定的，篡改恢复质量和安全性提高大多依赖于水印嵌入容量的增加. 兼顾水印嵌入容量和安全性，本文提出一种水印嵌入容量可变的恢复水印

算法，该算法首先提取 2×2 图像块特征生成变容量恢复水印信息——平滑图像块 6 比特，纹理图像块 12 比特，然后将图像块的恢复水印基于密钥随机嵌入到其它图像块的低有效位. 认证时通过比较块特征与相应恢复水印重构特征并结合邻域特征判定图像块的真实性. 变容量恢复水印在尽可能多保存原始图像信息的同时使水印容量最少，另外，本文算法的恢复水印仅被嵌入一次且同时用于篡改检测和篡改恢复，既提高了算法抵抗拼贴攻击的能力，又降低了水印嵌入容量，从而提高了含水印数字图像的质量. 为提高篡改恢复质量，首先利用提取的有效恢复水印修复篡改块，然后利用有效邻域像素修复同步篡改块. 与最近文献[4]和[6]的实验比较结果表明，本文算法的水印嵌入容量最小，且能有效抵抗拼贴攻击、恒均值攻击等已知伪造攻击.

2 变容量恢复水印

本文提出的变容量恢复水印算法分为水印生成与嵌入、水印提取与篡改检测、篡改恢复 3 个部分.

2.1 水印生成与嵌入

图 1 示出恢复水印的生成与嵌入过程，包括以下 5 个步骤.

(1) 分块并分类. 将大小为 $2m\times 2n$ 的原始图像 X 分为 $m\times n$ 个互不重叠的 2×2 图像块，按从上至下、从左至右对图像 X 的块编号：$X=\{X_i\mid i=1,2,\cdots,N\}$，其中 $N=m\times n$ 为图像块个数. 大小为 2×2个像素的图像块表示为 $X_i=\{x_{i1},x_{i2},x_{i3},x_{i4}\}$. 如果图像块 X_i 中 2 个最大像素高 5 位之和与 2 个最小像素高位 5 位之和的差不超过 3，则认为图像块 X_i 为平滑图像块，否则为纹理图像块.

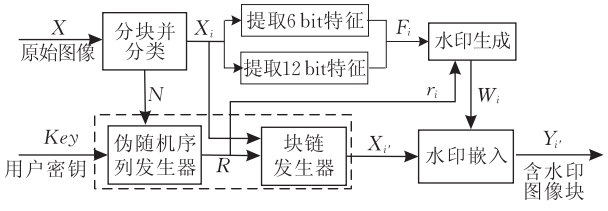


图 1 变容量恢复水印生成与嵌入

(2) 伪随机序列与块链生成. 基于用户密钥 Key 生成长度为 N 的实值伪随机序列 $R=\{r_i\mid i=1,2,\cdots,N\}$，根据 R 生成块链 $\{(X_i,X_{i'})\mid i,i'\in[1,N]\}$ 以确定每个图像块的水印嵌入位置，其中 i' 为 R 中第 i 个元素在 R 中的索引位置，详见参考文献[9].

(3) 特征提取. 对每个图像块 X_i，生成 v 比特块特征 $F_i=\{f_{i1},f_{i2},\cdots,f_{iv}\}$. $f_{i2}\sim f_{i6}$ 为图像块 X_i 高 5 位平均值的二进制编码，即

$$f_{i2}\sim f_{i6}=\left(\left\lfloor\frac{1}{4}\sum_{j=1}^{4}\lfloor x_{ij}/8\rfloor\right\rfloor\right)_B \tag{1}$$

其中 $(\cdot)_B$ 表示整数的二进制编码. 如果图像块为平滑图像块，$v=6$ 且 $f_{i1}=0$，否则，$v=12$ 且 $f_{i1}=1$. 图 2 给出了非平滑图像块的 6 种类型及相应的子类编码，其中黑色表示 2×2 图像块中最大 2 个像素的位置；$f_{i7}\sim f_{i9}$ 为子类编码；$f_{i10}\sim f_{i12}$ 为 2 个最大值像素之和与其它 2 个像素之和的差均匀量化的二值编码，即

$$f_{i10}\sim f_{i12}=\left(\left\lfloor\frac{1}{8}((\lfloor x_{i1'}/8\rfloor+\lfloor x_{i2'}/8\rfloor)-(\lfloor x_{i3'}/8\rfloor+\lfloor x_{i4'}/8\rfloor))\right\rfloor\right)_B \tag{2}$$

式(2)中的 4 个像素满足 $x_{i1'}\geq x_{i2'}\geq x_{i3'}\geq x_{i4'}$.

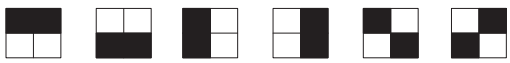


图 2 非平滑图像块分类及子类编码

(4) 恢复水印生成. 利用伪随机序列 R 的第 i 个随机数 r_i 生成二值伪随机序列 $B_i=\{b_{ij}\mid j=1,2,\cdots,12\}$，加密图像块特征 F_i 生成恢复水印 $W_i=\{w_{i1},w_{i2},\cdots,w_{iv}\}$，

$$w_{ij}=f_{ij}\oplus b_{ij},\ j=1,2,\cdots,v \tag{3}$$

其中，$\oplus$ 为异或操作.

(5) 水印嵌入. 对每个图像块 X_i，由步(2)生成的块链得到其映射块 $X_{i'}$. 把图像块 X_i 的恢复水印 W_i 嵌入到其映射块 $X_{i'}$ 的低位生成含水印图像块 $Y_{i'}$：

如果 X_i 为非平滑图像块，将 12 比特恢复水印 $w_{i1}\sim w_{i12}$ 嵌入 $Y_{i'}$，

$$Y_{i'}=\lfloor x_{i'j}/8\rfloor\times 8+4w_{i(j+8)}+2w_{i(j+4)}+w_{ij} \tag{4}$$

如果 X_i 为平滑图像块，将 6 比特恢复水印 $w_{i1}\sim w_{i6}$ 嵌入 $Y_{i'}$，

$$Y_{i'}=\begin{cases}\lfloor x_{i'j}/4\rfloor\times 4+2w_{i(j+4)}+w_{ij}, & j=1,2\\ \lfloor x_{i'j}/2\rfloor\times 2+w_{ij}, & j=3,4\end{cases} \tag{5}$$

2.2 水印提取与篡改检测

图像 Y^* 为被测图像，水印提取是水印嵌入的逆过程，篡改检测是判定每个图像块的真实性. 用篡改检测矩阵 $\boldsymbol{T}=(t_i\mid i=1,2,\cdots,N)$ 表示检测结果，$t_i=1$ 表示被测图像块 Y_i^* 被篡改，$t_i=0$ 表示被测图像块 Y_i^* 是真实的.

(1)分块、伪随机序列和块链生成. 按 2.1 节中的“水印生成与嵌入”的步(1)和(2)把 Y^* 划分为 2×2 的图像块 $Y^*=\{Y_i^*\mid i=1,2,\cdots,N\}$ 并基于密钥 Key 生成随机序列 $R=\{r_i\mid i=1,2,\cdots,N\}$ 和块链 $\{(Y_i^*,Y_{i'}^*)\mid i,i'\in[1,N]\}$.

(2)块特征计算与重构. 对每个待测图像块 Y_i^*，按 2.1 节中的“水印生成与嵌入”的步(3)计算图像块特征 F_i^*，同时根据从其映射块 $Y_{i'}^*$ 低位提取恢复水印 W_i^* 重构块特征 F_i^L，

$$F_i^L=W_i^*\oplus B_i \tag{6}$$

其中，B_i 为基于伪随机序列 R 的第 i 个随机数 r_i 生成的长度为 12 的二值伪随机序列，从待测图像块 Y_i^* 的映射块 $Y_{i'}^*$ 低位提取的水印信息 $W_i^*=\{w_{ij}^*\mid j=1,2,\cdots,12\}$ 如式(7)所示，

$$w_{ij}^*=\begin{cases}\mathrm{mod}(y_{i'j}^*,2), & j=1,2,3,4\\ \mathrm{mod}(\lfloor y_{i'j}^*/2\rfloor,2), & j=5,6,7,8\\ \mathrm{mod}(\lfloor y_{i'j}^*/4\rfloor,2), & j=9,10,11,12\end{cases} \tag{7}$$

(3)特征比较. 对每个待测图像块 Y_i^*，比较高位计算的块特征 F_i^* 与由其恢复水印重构的块特征 F_i^L 并生成比较矩阵 $\boldsymbol{D}=\{d_i\mid i=1,2,\cdots,N\}$，

$$d_i=\begin{cases}0, & \sum_{j=1}^{c}2^jf_{ij}^*=\sum_{j=1}^{c}2^jf_{ij}^L\\ 1, & \text{其它}\end{cases} \tag{8}$$

其中，$c=6(1+f_{i1}^*)$，即平滑图像块比较 F_i^* 与 F_i^L 前 6 个比特，非平滑图像块比较 F_i^* 与 F_i^L 所有 12 个比特.

(4)篡改检测. 首先根据比较矩阵 $\boldsymbol{D}$ 生成其邻域特征矩阵 $\boldsymbol{\Delta}=\{\delta_i\mid i=1,2,\cdots,N\}$(边界不作处理，详见文献[9])，

$$\delta_i=\sum d_j,\ j=i\pm1,i\pm n,i+n\pm1,\ i-n\pm1 \tag{9}$$

然后根据矩阵 $\boldsymbol{D}$ 和 $\boldsymbol{\Delta}$ 生成初态篡改检测矩阵 $\boldsymbol{T}^0=(t_i^0\mid i=1,2,\cdots,N)$，

$$t_i^0=\begin{cases}1, & (d_i=1)\&(\delta_i\geq\delta_{i'})\\ 0, & \text{其它}\end{cases} \tag{10}$$

最后，修正初态篡改检测矩阵 $\boldsymbol{T}^0$ 并生成篡改检测矩阵 $\boldsymbol{T}=(t_i\mid i=1,2,\cdots,N)$，

$$t_i=\begin{cases}1, & \tau_i\geq3\\ t_i^0, & \text{其它}\end{cases} \tag{11}$$

其中 $\tau_i=\sum t_j^0,\ j=i\pm1,i\pm n,i\pm n\pm1$.

2.3　篡改恢复

如果被测图像中存在被判定为篡改的图像块，则对被测图像依次执行以下 2 步操作得到篡改恢复图像 Y^R.

(1)特征恢复. 对判定为篡改的图像块 Y_i^*，如果其映射块 $Y_{i'}^*$ 是真实的，则用式(5)重构的块特征 F_i^L 对 Y_i^* 进行恢复，得到修复图像块 Y_i^R，

$$Y_i^R=\begin{cases}\Gamma^{-1}(F_i^L), & t_i=1\text{ 且 }t_{i'}=0\\ Y_i^*, & \text{其它}\end{cases} \tag{12}$$

其中，$\Gamma^{-1}()$ 为图像块特征编码的逆过程. 并生成同步篡改块的标示矩阵 $\boldsymbol{L}=(l_i\mid i=1,2,\cdots,N)$，

$$l_i=\begin{cases}1, & t_i=1\text{ 且 }t_{i'}=1\\ 0, & \text{其它}\end{cases} \tag{13}$$

(2)邻域恢复. 在上步恢复的结果上，如果 $l_i=1$，即篡改块 Y_i^* 为同步篡改块，则用与图像块 Y_i^R 相邻 12 个像素中的有效像素的均值修正图像块 Y_i^R 中的每一个像素.

3　性能分析与比较

本文从鉴定数字图像真实性的实际需要出发，用以下 5 个指标来衡量可恢复水印算法的性能：① 水印嵌入容量. 单位像素嵌入的比特数(bit per pixel，bpp)；② 含水印图像质量. 与原始图像的峰值信噪比(Peak Signal-to-Noise Ratio，PSNR)；③ 篡改检测性能. 漏警概率和虚警概率；④ 篡改恢复质量. 篡改恢复图像与含水印图像的 PSNR；⑤恢复水印信息量. 水印重构图像与原始图像的 PSNR. 下面给出本文、Lee 等人[4]和 Zhang 等人[6]的实验仿真与比较结果. 由于文献[6]中第 2 个算法(Hierarchical self-embedding scheme)的性能更优，因此本文仅给出与文献[6]第 2 个算法的比较结果.

3.1　水印嵌入容量与水印信息量

表 1 给出了本文、Lee 等人[4]和 Zhang 等人[6]的水印嵌入容量和含水印图像质量比较结果. 由表 1 可以看出，文献[4]和[6]的水印嵌入容量固定为 3 bpp，本文算法的水印容量是可变的，表中给出数字图像的水印嵌入容量的变化范围为 1.8～2.62 bpp，图像越平滑，水印嵌入容量越小. 与之相对应，文献[4]和[6]对不同图像生成的含水印图像的 PSNR 相似，分别约为 40.5 dB 和 38 dB. 文献[4]生成的含水印数字图像的 PSNR 比文献[6]高约 2.5 dB，这是由于文献[4]在水印嵌入时采用了“平滑”技术. 由于不同数字图像的水印嵌入容量不同，本文算法生成的含水印数字图像的 PSNR 变化较大，介于 39～43 dB 之间，图像越平滑，生成含水印图像的质量越好，由表 1 看出本文的含水印图像的质量整体上明显优于文献[4]和[6].

表 1 水印嵌入容量和含水印图像质量

图像	水印嵌入容量/bpp			含水印图像的质量/dB		
	本文	Lee 等人[4]	Zhang 等人[6]	本文	Lee 等人[4]	Zhang 等人[6]
Lena	1.80	3	3	42.54	40.74	37.67
Beach	1.80	3	3	42.93	40.73	37.92
Airplane	1.80	3	3	42.14	40.70	37.35
Mona Lisa	1.81	3	3	42.75	40.75	37.97
Man	1.94	3	3	42.00	40.72	37.93
Napoleon	2.05	3	3	41.39	40.77	37.95
Goldhill	2.01	3	3	40.96	40.73	37.55
Barbara	2.14	3	3	40.60	40.73	37.76
Fingerprint	2.62	3	3	39.00	40.73	37.71

恢复水印包含原始图像的信息量(简称“水印信息量”)越多,相同条件下算法的篡改检测概率越大,篡改恢复质量也越高.理想情况是用尽可能少的比特保存尽可能多的图像信息.兼顾算法定位精度和抗恒均值攻击的能力,本文在提取 2×2 图像块均值的同时,对纹理图像块增加“细节”编码,从而保存更多原始图像的信息,即增加了水印信息量.相应地,根据水印信息重构图像的质量也会更好.为定量地比较算法生成恢复水印的信息量,表 2 给出了恢复水印重构图像与原始图像的 PSNR.文献[4]的恢复水印是 2×2 图像块的均值编码,本文在均值编码的基础上,添加了对纹理图像块的细节编码,因此,对不同的数字图像,本文恢复水印重构图像的质量都优于文献[4].由表 2 看出,本文算法的恢复水印重构图像的 PSNR 高于文献[4]约 5 dB 以上,有的甚至高达 9 dB(如 Napoleon).文献[6]对 8×8 图像块分层编码,8×8 图像块的编码长度为 167(102+45+20)比特,尽管图像块的特征编码比特数多于本文算法,但恢复水印重构图像的质量并没有优势.由表 2 可以看出,文献[6]恢复水印重构图像的 PSNR 仅 Goldhill 图像的高于本文算法约 1 dB,其它图像的 PSNR 都低于本文算法,有的甚至高达 10 dB(如 Fingerprint).

表 2 恢复水印重构图像与原始图像的 PSNR 比较

图像	PSNR		
	本文	Lee 等人[4]	Zhang 等人[6]
Lena	33.82	28.11	32.75
Beach	35.65	27.54	31.30
Airplane	33.35	26.87	29.07
Mona Lisa	34.91	25.55	26.56
Man	33.56	27.37	31.23
Napoleon	33.34	24.14	25.97
Goldhill	33.02	27.41	34.00
Barbara	30.43	24.26	27.56
Fingerprint	30.06	24.22	20.10

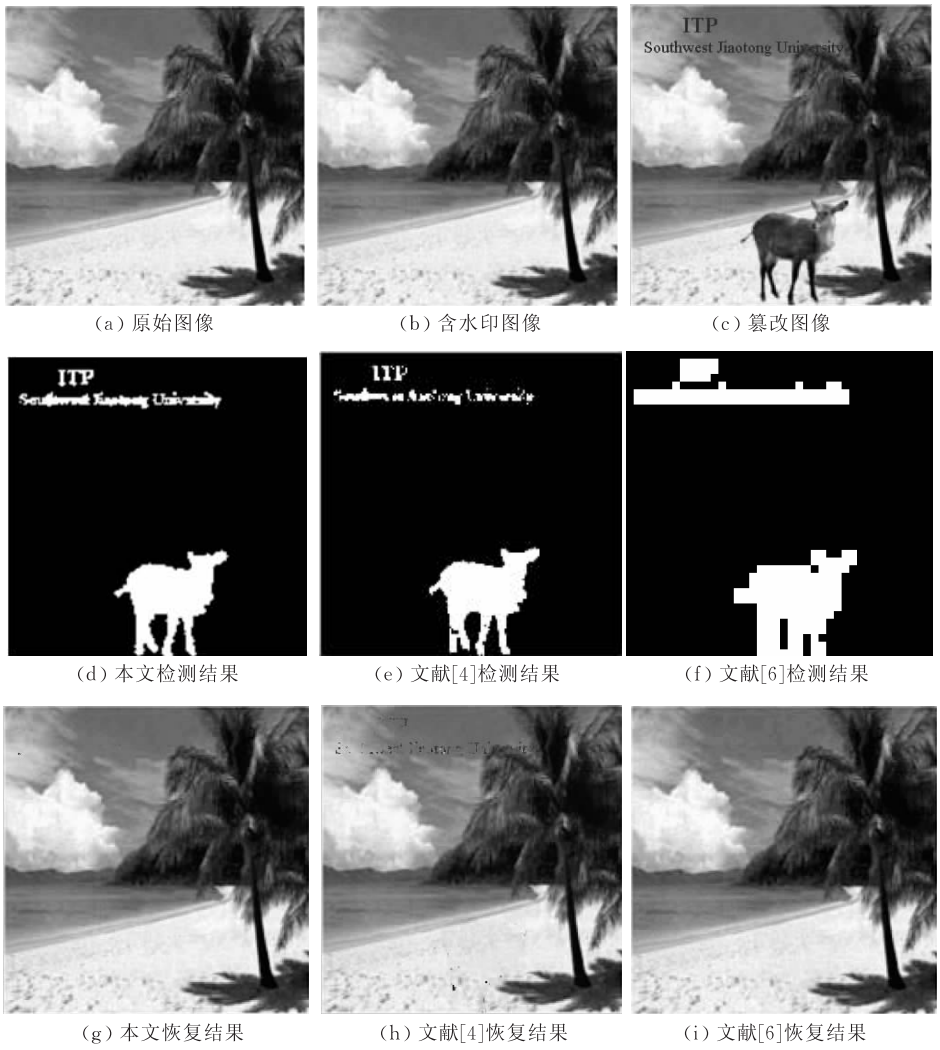
总之,文献[4]和[6]的水印嵌入容量是固定的,而本文的水印嵌入容量是可变的,数字图像越平滑,水印嵌入容量越少,相应地生成含水印数字图像的质量也越好.由于在水印嵌入时没有引入冗余信息,本文的水印嵌入容量低于文献[4]和[6],而恢复水印信息量总体也优于文献[4]和[6],为提高算法的安全性和篡改恢复质量提供了保障.

3.2 篡改检测与篡改恢复

现有恢复水印算法都能检测一般篡改,其差别主要体现在定位精度、篡改检测性能和篡改恢复图像质量的高低.图 3 给出一般篡改条件下,本文、文献[4]和[6]的篡改检测与篡改恢复的结果.其中,图 3(a)是大小为 328×328 像素的 Beach 原始灰度图像,图 3(b)为利用本文算法生成的含水印数字图像,其 PSNR 为 42.93 dB,图 3(c)为篡改图像,其中包含以下篡改:① 上方添加文字“ITP Southwest Jiaotong University”;② 下方添加一只小鹿.篡改像素的比例约为 7.78%.图 3(d)、(e)和(f)分别为本文、文献[4]和[6]的篡改检测结果;图 3(g)、(h)和(i)分别为本文、文献[4]和[6]的篡改恢复图像.

利用 2×2 像素的均值生成恢复水印其定位精度较高,却不能抵抗文献[11]提出的恒均值攻击.本文提出的变容量恢复水印扩大了水印信息量,也提高了算法抵抗恒均值攻击的能力.图 4 给出恒均值攻击条件下,本文、文献[4]和[6]的篡改检测与篡改恢复结果.其中,图 4(a)是大小为 256×256 像素的 Lena 原始灰度图像,图 4(b)为利用本文算法生成的含水印数字图像,图 4(c)为篡改图像,其中,大小 140×140 像素的正方形区域遭到恒均值攻击,篡改比例约为 30%.图 4(d)、(e)和(f)分别为本文、文献[4]和[6]的篡改检测结果,图 4(g)、(h)和(i)分别为本文、文献[4]和[6]的篡改恢复结果.

拼贴攻击[8]是对脆弱水印算法威胁最大的攻击之一,现有很多脆弱水印算法不能有效抵抗该攻击.图 5(a)和(b)分别是利用相同密钥生成的含水印 Napoleon 和 Mona Lisa 灰度图像,大小均 208×328



(a) 原始图像　(b) 含水印图像　(c) 篡改图像

(d) 本文检测结果　(e) 文献[4]检测结果　(f) 文献[6]检测结果

(g) 本文恢复结果　(h) 文献[4]恢复结果　(i) 文献[6]恢复结果

图 3　一般篡改时的篡改检测与篡改恢复性能比较

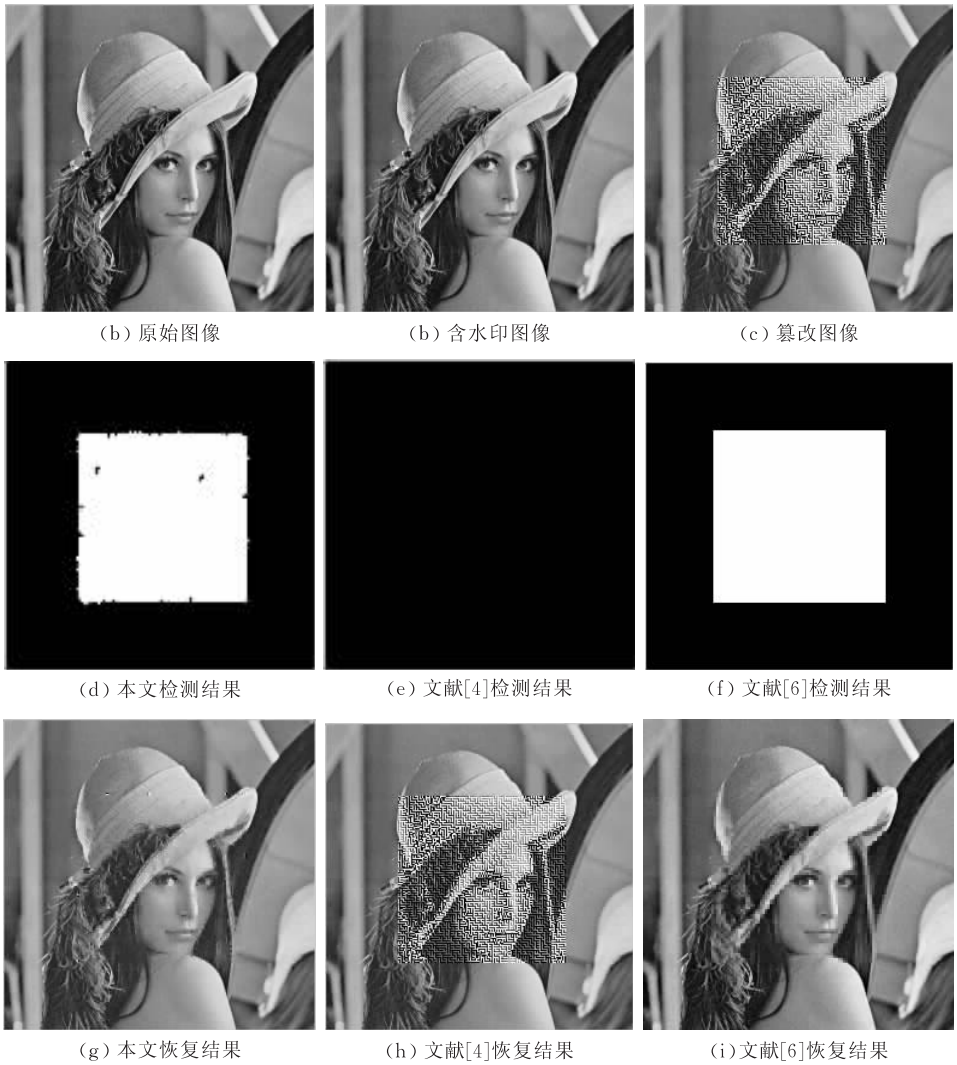
像素.将含水印 Napoleon 图像显示的头部区域替换为含水印 Mona Lisa 图像的相同区域(拼贴攻击),得到的篡改图像如图 5(c)所示,篡改比例约为 18%.图 5(d)、(e)和(f)分别为本文、文献[4]和[6]的篡改检测结果,图 5(g)、(h)和(i)分别为本文、文献[4]和[6]的篡改恢复结果.

为定量比较算法的篡改检测和恢复性能,表3给出了上述 3 种篡改方式下,本文、文献[4]和[6]的篡改检测性能(漏警概率和虚警概率)和篡改恢复质量(篡改恢复图像与原始图像的 PSNR).由图 3～图 5 的仿真结果可以看出,本文和文献[4]定位篡改的基本单位是 2×2 像素,文献[6]是 8×8 像素.为统一衡量标准,文献[6]漏警概率和虚警概率也以 2×2 像素为单位计算.

表 3　篡改检测性能与篡改恢复质量比较

Test	漏警概率/%			虚警概率/%			篡改恢复图像/dB		
	本文	文献[4]	文献[6]	本文	文献[4]	文献[6]	本文	文献[4]	文献[6]
一般篡改	0.3	7.9	0	0.39	0.41	5.19	43.63	33.32	45.24
恒均值攻击	0.21	100.0	0	0.94	0	3.69	30.81	11.20	28.46
拼贴攻击	2.03	98.4	86.67	0.13	0	2.27	33.82	17.26	17.72

结合图 3 和表 3 可以看出,一般篡改条件下,文献[6]的漏警概率最低,但虚警概率较高,这主要是由于文献[6]的图像块大小为 8×8 像素.文献[4]的虚警概率较低(0.41%),而漏警概率较高(7.9%),这是由于文献[4]仅用 2 比特认证水印检测图像块的真实性,导致篡改边界(尤其是添加的文字)存在较多没被检测的图像块.本文算法的漏警概率和虚警概率都较低,分别为 0.3%和 0.39%.由于漏检的



(b) 原始图像 (b) 含水印图像 (c) 篡改图像

(d) 本文检测结果 (e) 文献[4]检测结果 (f) 文献[6]检测结果

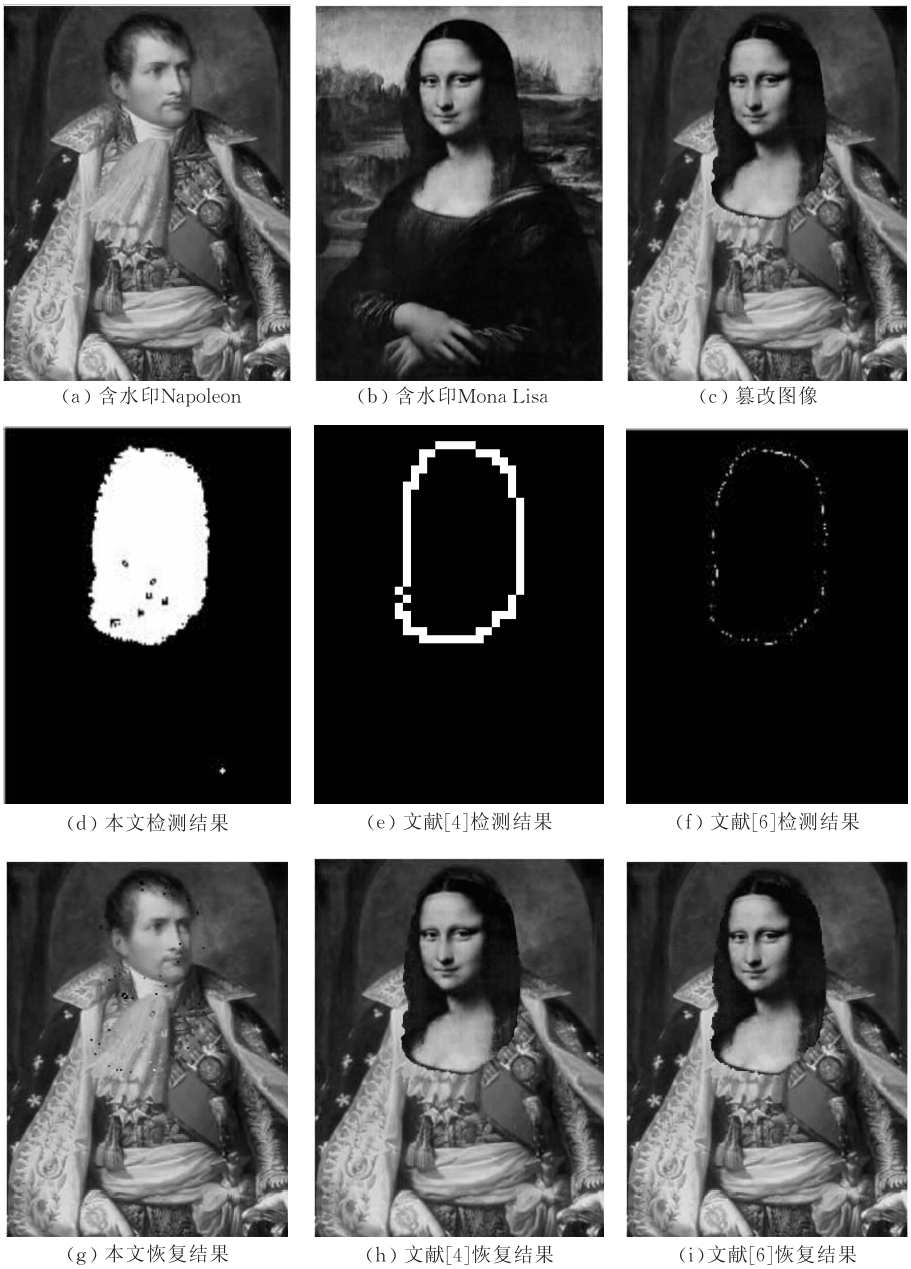
(g) 本文恢复结果 (h) 文献[4]恢复结果 (i)文献[6]恢复结果

图 4 恒均值攻击时的篡改检测与篡改恢复结果比较

篡改块不执行篡改恢复,文献[4]在一般篡改下的恢复质量不高.相应地,文献[6]的漏警概率为 0,所有被篡改的图像块都能进行篡改恢复,且由于篡改比例小,三层编码都能得到有效恢复,因此文献[6]的恢复质量最好.由于本文算法的漏警概率不为零,导致此条件下本文的恢复质量略低于文献[6].本文、文献[4]和[6]的恢复图像的 PSNR 分别为 43.63 dB、33.33 dB 和 45.24 dB.显然一般篡改条件下,本文算法能精确定位篡改且能得到质量较好的恢复图像.

结合图 4 和表 3 可以看出,恒均值攻击条件下,文献[4]的漏警概率和虚警概率分别为 100% 和 0%.由于没有检测到篡改块,文献[4]算法不执行篡改恢复操作,说明文献[4]不能抵抗恒均值攻击.文献[6]能有效抵抗恒均值攻击,其漏警和虚警概率分别为 0% 和 3.69%,然而由于篡改比例较大(约 30%)导致文献[6]第三层的一些编码比特不能正确恢复,导致恢复图像的质量略低,恢复图像与含水印图像的 PSNR 为 28.46 dB.本文算法的漏警和虚警概率分别为 0.21% 和 0.94%,其篡改恢复图像与原始图像的 PSNR 为 30.81 dB,说明本文算法能有效抵抗恒均值攻击.

结合图 5 和表 3 可以看出,拼贴攻击条件下,文献[4]和[6]的漏警概率较高,分别为 98.4% 和 86.67%.从图 5(e)和(f)可以看出,文献[4]和[6]仅检测出拼贴区域的边界,拼贴区域的内部通过认证.文献[4]和[6]之所以能检测出拼贴区域边界的图像块,主要是因为拼贴区域不是以图像块为单位实施的篡改.由于漏检的篡改块不执行篡改恢复操作,导致文献[4]和[6]的篡改恢复质量较低,它们恢复图像的 PSNR 分别为 17.26 dB 和 17.72 dB,说明文献[4]和[6]都不能有效抵抗拼贴攻击.本文算法的漏警概率和虚警概率分别为 2.03% 和 0.13%,其恢复图像的 PSNR 为 33.82 dB,说明本文算法能有效抵



(a) 含水印Napoleon　(b) 含水印Mona Lisa　(c) 篡改图像

(d) 本文检测结果　(e) 文献[4]检测结果　(f) 文献[6]检测结果

(g) 本文恢复结果　(h) 文献[4]恢复结果　(i) 文献[6]恢复结果

图 5　拼贴攻击时的篡改检测与篡改恢复性能比较

抗拼贴攻击.

本文算法中的恢复水印容量是可变的,数字图像越平滑,生成的恢复水印量越少.在满足保存适量块信息的前提下降低了恢复水印的容量,从而提高了含水印图像的质量.同时,本文算法具有较高的安全性和篡改恢复质量.在一般篡改、恒均值攻击、拼贴攻击下,都能精确定位篡改位置、高概率检测篡改块并得到高质量的恢复图像.

4　结束语

现有数字图像可恢复水印算法的水印嵌入容量都是固定的,其篡改恢复质量和安全性的提高,主要依赖于水印嵌入容量的扩大,而水印嵌入容量的扩大降低了含水印图像的质量.兼顾水印嵌入容量、安全性和篡改恢复质量,本文根据图像块自身特性生成变容量恢复水印,并从以下几个方面来提高算法的各项性能:(1)根据图像块的平滑或纹理特性,生成容量可变的恢复水印信息;(2)实现变长恢复水印的嵌入、提取和篡改检测;(3)无需增加认证水印信息且恢复水印仅嵌入一次,降低水印嵌入容量.与文献[4]和[6]的实验比较结果表明,本文算法的定位精度为 2×2 像素、水印嵌入容量可变、含水印数字图像的质量更好且能有效抵抗已知的伪造攻击,

如拼贴攻击、恒均值攻击等. 如何进一步降低本文算法的漏警概率、篡改检测性能分析将是我们下一步的研究内容.

参 考 文 献

[1] Fridrich J, Goljan M. Images with self-correcting capabilities//Proceedings of the IEEE International Conference on Image Processing Proceedings (ICIP'99). Kobe, Japan, 1999, 3: 792-796

[2] He Hong-Jie. Digital image secure authentication watermarking algorithms and their performance analysis of statistical detection[Ph. D. dissertation]. Southwest Jiaotong University, Chengdu, 2009(in Chinese)
(和红杰. 数字图像安全认证水印算法及其统计检测性能分析[博士学位论文]. 西南交通大学, 成都, 2009)

[3] Lin P L, Hsieh C K, Huang P W. A hierarchical digital watermarking method for image tamper detection and recovery. Pattern Recognition, 2005, 38(12): 2519-2529

[4] Lee T Y, Lin S D. Dual watermark for image tamper detection and recovery. Pattern Recognition, 2008, 41(11): 3497-3506

[5] Yang C W, Shen J J. Recover the tampered image based on VQ indexing. Signal Processing, 2010, 90(1): 331-343

[6] Zhang X, Wang S, Qian Z, Feng G. Reference sharing mechanism for watermark self-embedding. IEEE Transactions on Image Processing, 2011, 20(2): 485-495

[7] Qian Z, Feng G, Zhang X, Wang S. Image self-embedding with high-quality restoration capability. Digital Signal Processing, 2011, 21(2): 278-286

[8] Fridrich J, Goljan M, Memon N. Cryptanalysis of the Yeung-Mintzer fragile watermarking technique. Journal of Electronic Imaging, 2002, 11(2): 262-274

[9] He Hong-Jie, Zhang Jia-Shu, Chen Fan. A self-recovery fragile watermarking scheme for image authentication with superior localization. Science in China Series F-Information Sciences, 2008, 51(10): 1487-1507

[10] He Hongjie, Zhang Jiashu, Chen Fan. Adjacent-block based statistical detection method for self-embedding watermarking techniques. Signal Processing, 2009, 89(8): 1557-1566

[11] Chang C, Fan Y H, Tai W L. Four-scanning attack on hierarchical digital watermarking method for image tamper detection and recovery. Pattern Recognition, 2008, 41(2): 654-661

CHEN Fan, born in 1971, Ph. D. candidate, associate professor. His research interests include multimedia security and digital watermarking.

HE Hong-Jie, born in 1971, Ph. D., associate professor. Her research interests include digital image processing and information hiding.

WANG Hong-Xia, born in 1973, Ph. D., professor. Her research interests include information hiding, digital watermarking and intelligent information processing.

Background

Digital image authenticity is greatly threatened in that it is not difficult to modify or forge the image content without leaving detectable traces by publicly available image processing software packages. Fragile watermarking is to achieve digital image content authentication by imperceptibly embedding additional information into the host image. Self-recovery fragile watermarking not only has the capability of indicating where image contents have been tampered with, but also recovers the original content of the corrupted regions of the image. Self-recovery watermarking techniques for image authentication usually partition the image into blocks of the same size. The recovery watermark of an image block is a compressed version (features) of it and embedded into the less significant bits of the other distant block. This makes the self-recovery watermarking algorithm difficult to detect and localize the possible tampering. In the most existing self-recovery watermarking schemes, the validity of an image block was determined by additional authentication data embedded in the same block. Additional authentication data increase watermark payload of the self-recovery system and makes these self-embedding schemes vulnerable to the collage attack. To address this problem, our research group proposed an adjacent-block based statistical detection method to detect the validity of image block in the self-recovery watermarking techniques. This work was published in Journal of Signal Processing. To improve the localization accuracy, our research group proposed a self-recovery fragile watermarking scheme with superior localization published in Science in China Series F-Information Sciences. Furthermore, the existing self-recovery watermarking schemes generally resolved the tampering coincidence problem by embedding some redundant information in the host image. By inserting the redundant information, although the quality of recovery image is improved, the quality of watermarked image is decreased due to the increased watermark payload. The primary objective of this work is to take into account both the watermark embedding capacity and security of self-recovery watermarking schemes. The 6 bits and 12 bits features are generated for the smooth and rough blocks of 2×2 pixels, respectively. The watermark payload of each block is variable to make the distortion caused by watermark insertion as small as possible, and all bits in the watermark payload are used to both tamper detection and tamper recovery.