基于精细认证和迭代补偿机制的图像认证与恢复算法

程宝田 倪蓉蓉 赵耀

（北京交通大学信息科学研究所，北京，100044）

摘 要：针对基于分块思想的图像定位算法定位精度不高和恢复性能不好的问题，本文提出了一种基于精细认证和迭代补偿机制的图像认证与恢复算法。在认证算法中，包含基本认证和精细认证，融合基本认证和精细认证这两次认证的结果，并根据连通篡改区域中像素个数的多少进行优化，得到定位精度较高的定位结果。在恢复算法中，我们考虑到可嵌入的信息容量，提出利用迭代补偿机制恢复图像中被篡改的像素。理论分析和实验结果表明：本文提出的算法对于局部发生篡改的图像，能提高其定位的精度，而且恢复的性能也得到了提高。

关键字：基本认证；精细认证；连通篡改区域；迭代补偿机制

中图分类号：TP391 **文献标识码：**A

Refining-Localization-Based Image Authentication and Recovery Scheme using Iterative Compensation Mechanism

Cheng Baotian, Ni Rongrong, Zhao Yao

(Institute of Information Science, Beijing Jiaotong University, Beijing, 100044, China)

Abstract: To enhance the accuracy of tamper localization and the performance of recovery, the paper proposes refining-localization-based image authentication and recovery scheme using iterative compensation mechanism. The authentication algorithm contains basic authentication and refining authentication. It fuses the twice authentication results and take some post-processing measures based on connected tampered region to obtain the accuracy authentication result. According to embedding capacity, the recovery scheme uses iterative compensation mechanism to recover the tampered pixels. Theoretical analysis and simulation results show that the proposed algorithm can improve the accuracy of tamper localization and the probability of recovering the tampered pixels.

Key words: basic authentication; refining authentication; connected tampered region; iterative compensation mechanism

引言

[1]多媒体技术的飞速发展和功能强大的媒体处理工具给我们的生活带来了很大的便利，但是同时也带来了信息安全的问题。目前，非法的篡改已经引起了严重的危害。传统的信息加密方法已满足不了我们的要求，所以迫切的需要一种新的技术手段解决多媒体信息安全的问题。数字水印技术[1]就是在这样的背景下应运而生的，为解决多媒体信息安全的问题提供了一种新的思路，已经成为一个新的研究热点。

保护数字图像的完整性与真实性是保护多媒体信息安全的一个重要方面。目前，针对数字图像的完整性与真实性的数字水印算法已经出现了很多，大致上可以分为基于单像素的数字图像认证算法和基于分块的数字图像认证算法两类。基于单像素[2-5]的数字图像认证算法根据每个单独的像素生成认证信息，并嵌入到该像素中。在认证过程中，通过比较根据当前像素重新计算的参考信息与从该像素中提取的信息是否一致来判断像素是否发生篡改。但是这种方法

[1]基金项目：国家杰出青年科学基金（61025013），国家自然科学基金（61073159），中央高校基本科研业务费专项资金（2012JBM042）。

存在一个缺陷，即由于受到可嵌入信息容量的限制，不同的像素可能生成相同的认证信息，这样就会造成误判的发生。基于分块[6-8]的数字图像认证算法是将图像分成规则或不规则的图像块，然后根据各个图像块生成各自图像块的认证信息，并嵌入到自身块或匹配块中。在认证的过程中，通过比较根据当前图像块重新计算的参考信息与从该图像块或其匹配块中提取的信息是否一致来判断当前图像块是否发生篡改。基于分块算法的篡改定位是块级别的，也就是说即使块中只包含有一个篡改的像素，整个块都会被认为是篡改的。

基于水印的图像认证可以借助嵌入的信息实现篡改图像的恢复，最常用的技术手段就是将代表图像块主要特征的信息作为恢复水印，嵌入到该图像块的匹配块中。当图像块发生篡改时，就可以利用保存在其匹配块中的恢复信息对其进行恢复。但是当其匹配块也发生篡改时，就不能对图像块进行恢复。为了避免图像块与其匹配块出现篡改碰撞的情况，张新鹏等考虑不再使用匹配块技术，而是提出了一种基于参考共享机制[9]的恢复水印算法。该算法将代表图像内容的信息分散地嵌入到整幅图像中，当图像发生局部篡改时，图像篡改区域的恢复信息不会全部丢失，仍然会存在部分恢复信息，我们就可以利用这些恢复信息对篡改图像进行恢复。

针对上面提到的问题，本文提出了一种基于精细认证和迭代补偿机制的图像认证与恢复算法。认证算法包含基本认证和精细认证，基本认证可以将包含篡改像素的所有图像块都定位出来，而通过图像置乱技术，精细认证就有可能使得包含在已做篡改标记的图像块中的一些真实的像素通过认证。融合这两种认证的结果，并根据连通篡改区域中像素的个数优化融合的结果，得到定位精度较高的定位结果。而在恢复算法中，本文借鉴参考共享机制的思想，提出了一种基于迭代补偿机制的恢复水印算法，提高了篡改信息恢复的概率，使得图像的恢复性能得到提高。

1 基于精细认证和迭代补偿机制的图像认证与恢复算法

针对基于分块思想的图像篡改定位算法定位精度不高和恢复性能不好的问题，本文提出了一种基于精细认证和迭代补偿机制的图像认证与恢复算法。该算法在认证过程中可以提高图像篡改定位的精度；在恢复过程中，可以增加篡改信息恢复的概率，提高篡改图像恢复的性能。

1.1 置乱函数

假设一段信息含 M 比特，用 I_i 表示第 i 比特信息，i=1,2，...,M。然后根据一个随机的密钥生成一个从 1 到 M 的随机序列，表示为 R，R_i 表示序列中第 i 个数。利用序列 R，建立信息各比特之间一对一的对应关系，I_1 与 I_{R_1} 对应，I_2 与 I_{R_2} 对应，...,以此类推，I_i 与 I_{R_i} 对应。当信息需要置乱时，只需将信息 I_i 调整到信息 I_{R_i} 的位置即可。当所有的信息都经过上述处理之后，一段置乱的信息就生成了。以上过程，表示成一个函数的形式，即置乱函数 $S(\bullet)$。

当需要恢复置乱的信息时，我们根据生成置乱的信息的方法，只需要对信息位置做一个相反的调整，即将信息 I_{R_i} 调整到信息 I_i 所在的位置即可。将这一过程表示成函数的形式，即反置乱函数 $S^{-1}(\bullet)$。

1.2 迭代补偿机制

迭代补偿机制通过迭代运算，将第一组方程求出的解作为一种补偿，带入第二组方程中进行求解，紧接着，将第二组方程求出的解作为一种补偿，再次带入第一组方程中进行求解，将此时求出的解再作为补偿，带入第二组方程进行求解。重复上述过程，一直迭代进行计算，直到未知数的个数始终保持不变为止。这样的机制我们定义为迭代补偿机制。

1.3 水印的嵌入

我们用 X 表示大小为 N_1*N_2(假定 N_1 和 N_2 都是 8 的倍数)的原始图像,像素的总数为 N。

1.3.1 恢复水印的嵌入

恢复水印用于对篡改像素进行恢复，是通过对图像所有像素的高5位信息进行矩阵运算得到的。图像的高 5 位信息包含了图像的大部分内容，而由于受到水印可嵌入容量的限制，恢复水印信息的长度不能太长。这里我们使用矩阵运算生成恢复信息，相当于对像素的高 5 位信息进行了压缩。恢复水印的构造和嵌入过程流程图如图 1 所示。

恢复水印的构造和嵌入过程主要是通过以下四个步骤完成的：

步骤 1：将原始图像 X 的最低三位有效位置零，用于嵌入水印信息。最低三位有效位置零后生成的图像表示成 $\overline{X}$ ，然后提取图像 $\overline{X}$ 的所有像素的高 5 位信息，得到一个长度为 $5N$ 的信息序列 H 。

步骤 2：根据密钥 K_1，使用置乱函数 $S(\bullet)$ 对信息序列 H 进行置乱，然后将置乱后的序列分成 $N/2$ 组，每组 10 比特信息。用[$a_{m,1}$ $a_{m,2}$ $\cdots$ $a_{m,10}$]表示第 m 组信息。对每一组信息，使用公式(1)进行矩阵运算，得到 1 比特的恢复信息。当所有的分组都经过该运算后，得到一个长度为 $N/2$ 的恢复信息 R_1。

$$r_1^m=[a_{m,1}\ \ a_{m,2}\ \ \cdots\ \ a_{m,10}]\bullet B,\ m=1,2,\ldots,N/2 \tag{1}$$

其中矩阵 B 是一个大小为 10*1 的全 1 矩阵，公式(1)中的运算是模 2 运算，即对矩阵相乘的结果进行模 2，得到所需要的结果。下文中的矩阵运算采用与此相同的操作。

步骤 3：根据密钥 $K_2(K_1\neq K_2)$ ，同样的，使用置乱函数 $S(\bullet)$ 对信息序列 H 进行置乱，然后将置乱后的序列分成 $N/2$ 组，每组 10 比特信息。用[$b_{m,1}$ $b_{m,2}$ $\cdots$ $b_{m,10}$]表示第 m 组信息。对于每一组信息，使用公式(2)进行矩阵运算，得到 1 比特恢复信息。当所有的分组都经过该运算后，得到一个长度为 $N/2$ 的恢复信息 R_2。

$$r_2^m=[b_{m,1}\ \ b_{m,2}\ \ \cdots\ \ b_{m,10}]\bullet B,\ m=1,2,\ldots,N/2 \tag{2}$$

步骤 4：组合恢复信息 R_1 和 R_2 得到一个长度为 N 的恢复信息 R，然后根据密钥 K_3，使用置乱函数 $S(\bullet)$ 对其进行置乱。并将置乱后的恢复信息嵌入到图像 $\overline{X}$ 的倒数第三位，得到嵌入恢复水印后的图像 X_r 。

1.3.2 认证水印的嵌入

认证水印用来实现对图像的篡改检测与定位。哈希函数对于输入是非常敏感的，对于输入的任何改变，其输出结果都会发生变化。本文利用哈希函数这种高敏感性来构造认证水印。算法中包含两种认证水印，基本认证水印和精细认证水印。基本认证水印是根据未置乱的图像生成的，而精细认证水印是根据置乱图像生成的。其中，置乱可能将包含在已做篡改标记的图像块中的真实的像素分到一个真实的图像块中，然后通过精细认证认证出来。

认证水印的构造和嵌入的过程分为两步：精细认证水印的构造和嵌入、基本认证水印的构造和嵌入。认证水印的构造和嵌入过程流程图如图 2 所示。

精细认证水印的构造和嵌入过程：首先根据一个密钥，利用置乱函数 $S(\bullet)$ 对图像 X_r 进行置乱得到置乱图像 X_r^s 。然后将其划分为互不重叠的 8*8 大小的图像块，用 $X_{r_i}^s$ (i=1,2,...,$N/64$)表示第 i 块。对于每个图像块，将块中的所有像素输入到哈希函数中生成一个 64bits 哈希信息，将该信息作为图像块的认证信息嵌入到自身块的次低有效位。当所有的图像块都处理完成之后，再根据反置乱函数对图像进行反置乱，得到含有精细认证水印的图像 X_{rw} 。

基本认证水印的构造和嵌入过程：将图像 X_{rw} 同样进行 8*8 大小的分块，对于每一个图像块，为了抵抗 VQ 攻击[10]，不仅仅要

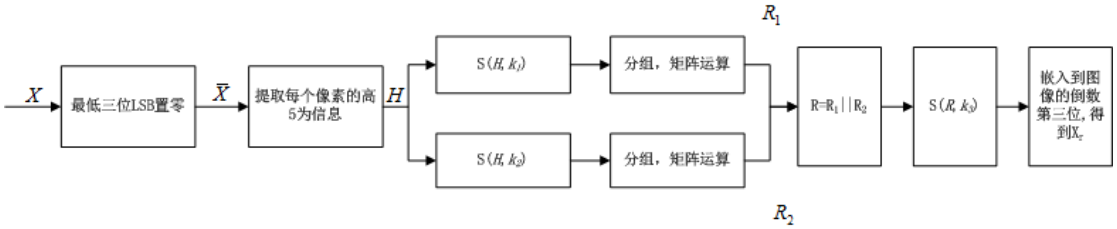


图 1 恢复水印的构造和嵌入过程

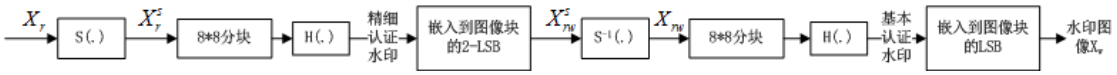


图 2 认证水印的构造和嵌入过程

将所有像素输入到哈希函数中，而且需要将图像块的位置索引和图像索引输入到哈希函数中生成一个 64bits 的哈希信息，该信息就是基本认证信息。将该信息嵌入到图像块的最低有效位，从而生成含水印图像 X_w。

1.4 图像的篡改检测与定位

在检测过程中，对于篡改的像素，用“1”做标记；对于未发生篡改的像素，用“0”做标记。认证过程的原理框图如图 3 所示：

认证过程包括基本认证和精细认证。基本认证能定位出包含篡改像素的所有图像块，精细认证能将已标记出的篡改图像块中的一些真实像素认证出来，这是因为图像置乱技术可能将原来包含在篡改块中的像素分到一个真实的块中。假设图像 Y 是待测图像，认证的过程由以下四个步骤完成。

步骤 1：基本认证。首先，我们对图像 Y 进行规则的 8*8 分块，用 $Y_i(i=1,2,3,\cdots,N/64)$ 来表示第 i 个图像块。对于每一个图像块 Y_i，提取图像块 Y_i 的最低有效位，形成提取序列 E_1^i。然后，将图像块 Y_i 中所有像素的最低有效位置零，再采用与水印嵌入过程中基本认证水印构造过程相同的方法，生成参考序列 C_1^i。最后，比较提取序列 E_1^i 与参考序列 C_1^i，当两者相同时，表示图像块 Y_i 未发生篡改；否则，该图像块发生了篡改。对于未发生篡改的图像块中的像素，用“0”标记；而对于发生篡改的图像块中的像素，用“1”标记。

当所有的图像块都经过上述认证之后，我们就得到了基本认证的结果 T_1。

步骤 2：精细认证。对图像 Y，根据图像置乱函数 $S(\bullet)$ 对其进行像素置乱，得到置乱图像 Y^s；同样的，对图像进行规则的 8*8 分块，表示成 $Y_i^s(i=1,2,...,N/64)$。对图像块 Y_i^s，首先提取图像块 Y_i^s 的次低有效位，形成提取序列 E_2^i。然后将图像块 Y_i^s 中所有像素的最低两位有效位置零，再采用与水印嵌入过程中精细认证水印构造过程相同的方法，生成参考序列 C_2^i。最后，比较提取序列 E_2^i 和参考序列 C_2^i，当两者相同时，表示图像块 Y_i^s 未发生篡改；否则，该图像块发生了篡改。同样的，对于未发生篡改的图像块中的像素，用“0”标记；而对于发生篡改的图像块中的像素，用“1” 标记。当所有的图像块都经过精细认证之后，将得到的认证结果通过反置乱函数 $S^{-1}(\bullet)$ 得到精细认证结果 T_2。

步骤 3：认证结果的融合。对于图像中的每一个像素，在基本认证和精细认证过程中都做了标记，那么只需将两次标记的结果进行“与”操作就可以得到整幅图像融合的结果，表示成 T。

步骤 4：后处理过程。首先基于四邻域寻找图像中存在的连通的篡改区域，并统计这些连通区域中像素的个数。然后将统计的个数按降序排列组成一个数组，并计算数组中相

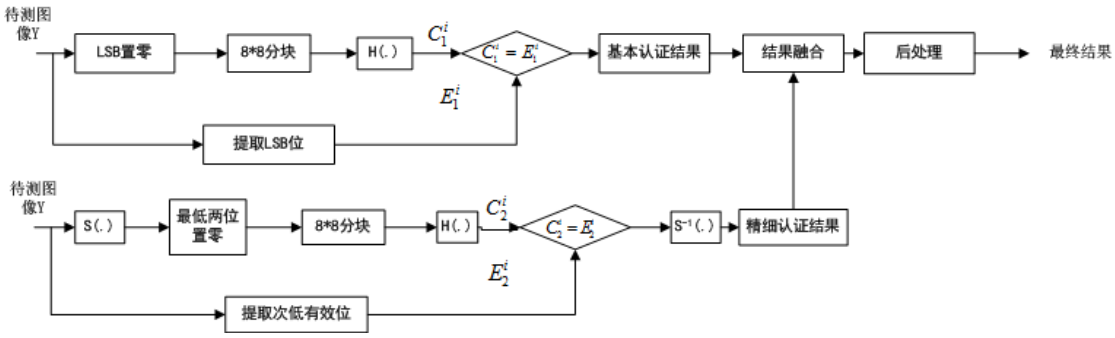


图 3 图像认证过程

邻元素的差值，当第一次出现某两个相邻的元素的差值大于等于 5 时，将此时较大的那个数作为阈值。当连通的篡改区域中像素的个数小于阈值时，将这个连通区域中的像素标记为“0”。

1.5 图像的篡改恢复

经过图像认证和篡改定位，所有的像素都已经被标记为篡改或未篡改。接下来，需要恢复篡改的像素。恢复过程如下：

步骤 1：对于篡改的每个像素，用八个相同的篡改标记表示八位比特信息，作为新的待测图像 Y。提取待测图像 Y 的所有像素的倒数第三位，得到一个长度为 N 的信息序列。根据密钥 K_3，使用反置乱函数对提取的信息序列进行反置乱得到信息序列 R_e。

步骤 2：提取待测图像 Y 的所有像素的高 5 位信息，得到一个长度为 $5N$ 的信息序列，表示成 A。

步骤 3：根据密钥 K_1，使用置乱函数，对序列 A 进行置乱得到 A^{s_1}，然后将置乱后的序列分成 $N/2$ 组，每组 10 比特信息，其中第 m 组序列表示为 $A_m^{s_1}$。取信息序列 R_e 的前 $N/2$ 比特，表示成 R_{e_1}。$r_{e_1}^m$ 表示 R_{e_1} 的第 m 个元素。当 $r_{e_1}^m$ 代表的信息是篡改标记的话，不做任何处理；否则可以建立(2)中的等式：

$$r_{e_1}^m - A_m^{s_1,R} \bullet B^R = A_m^{s_1,T} \bullet B^T \quad (2)$$

其中，$A_m^{s_1,R}$ 代表的是序列 A^{s_1} 第 m 组中未发生篡改的信息，而 $A_m^{s_1,T}$ 代表的是序列 A^{s_1} 第 m 组中发生篡改的信息，也就是那些需要恢复的信息，B^R 和 B^T 都是全 1 矩阵。

这样，我们就可以建立方程组一。

步骤 4：根据密钥 K_2，使用置乱函数，对序列 A 进行置乱得到 A^{s_2}，然后将置乱后的序列分成 N/2 组，每组 10 比特信息，其中第 m 组序列表示为 $A_m^{s_2}$。取信息序列 R_e 的后 $N/2$ 比特，表示成 R_{e_2}。$r_{e_2}^m$ 表示 R_{e_2} 的第 m 个元素。当 $r_{e_2}^m$ 代表的信息是篡改标记的话，不做任何处理；否则我们可以建立(3)中的等式：

$$r_{e_2}^m - A_m^{s_2,R} \bullet B^R = A_m^{s_2,T} \bullet B^T \quad (3)$$

其中，$A_m^{s_2,R}$ 代表的是序列 A^{s_2} 第 m 组中未发生篡改的信息，而 $A_m^{s_2,T}$ 代表的是序列 A^{s_2} 第 m 组中发生篡改的信息，也就是那些我们需要恢复的信息，B^R 和 B^T 都是全 1 矩阵。

这样，我们就可以建立方程组二。

步骤 5：对于两个方程组，我们使用迭代补偿机制进行求解，从而实现对篡改信息的恢复。解出的比特序列信息就是图像中被篡改的像素的高 5 位信息，从而实现了篡改图像的恢复。

2 实验结果

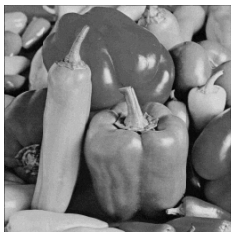
为了测试我们提出的算法的有效性和可行性，如图 4(a)-(c)所示，我们使用大小均为 256*256 的 car、peppers、couple 作为测试图像。根据本文提出的算法嵌入水印，生成的含水印图像如图 4(d)-(f)所示，它们的 PSNR 分别是：38.04dB、37.96dB 和 37.83dB。


(a) 原始图像 car


(b) 原始图像 peppers


(c) 原始图像 couple

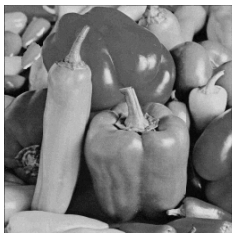

(d)含水印的图像


(e) 含水印的图像

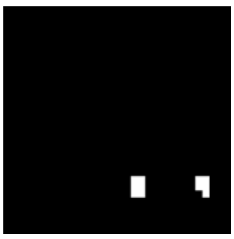

(f) 含水印的图像

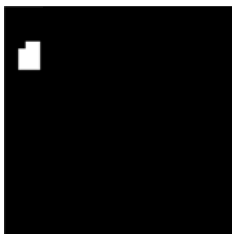
图 4 原始图像和嵌入水印后的图像

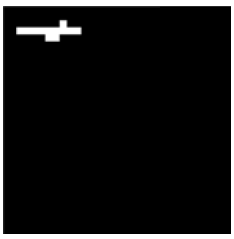

(a)篡改的图像 car

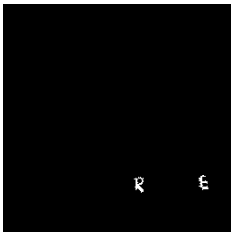

(b)篡改的图像 peppers

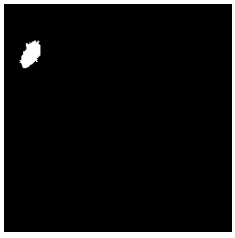
(c)篡改的图像 couple

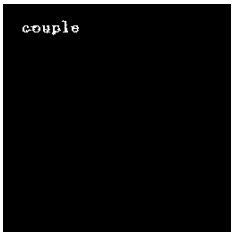

(d)图像 car 定位的结果


(e)图像 peppers 定位的结果


(f)图像 couple 定位的结果

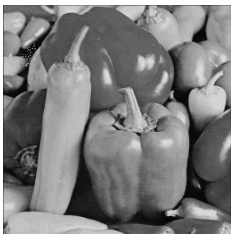

(g)图像 car 定位的结果


(h)图像 peppers 定位的结果



(i)图像 couple 定位的结果

图 5 篡改图像及图像篡改定位的结果

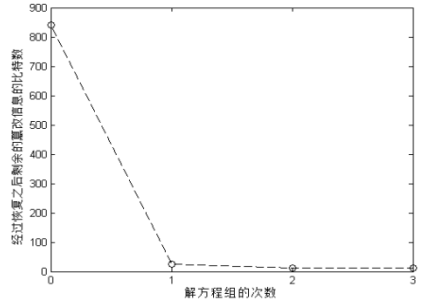




(a)图像 car 恢复图像　　(b)图像 peppers 恢复图像　(c)图像 couple 恢复图像

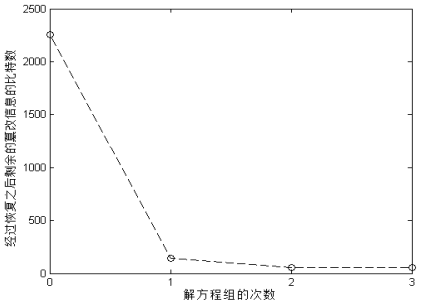
图 6 图像的篡改恢复图像

针对三幅含水印图像进行篡改。对图像“car”，如图 5(a)所示我们将车牌上的字母“P”改成“R”，“F”改成“E”；对图像“peppers”,如图 5(b)所示我们在图像添加了一个辣椒；对图像“couple”，如图 5(c)所示我们在图像中添加文字“couple”。图 5(d)-(f)，显示的是和红杰在[5]中提出的算法定位的结果，图 5(g)-(i)显示的是本文提出的算法定位的结果，比较两种方法定位出得结果，分析得出本文提出的认证算法在一定程度上提高了篡改定位的精度，定位出篡改区域的边缘更清晰。图 6(a)-(c)显示的是篡改恢复的图像，它们的 PSNR 分别是 37.98dB、35.50dB 和 36.68dB。

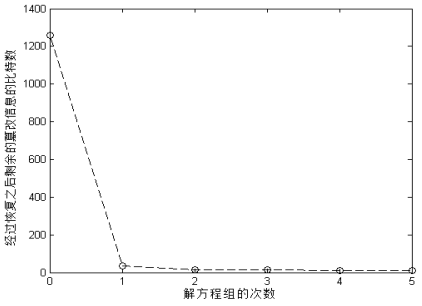
这里统计在恢复过程中，篡改信息的比特数变化情况，由于计算较复杂性，所以只统计方程中只有一个未知数的情况，统计的结果如图 7(a)-(c)所示。分析图 7(c)，根据本文提出的篡改检测与定位算法，统计出被篡改信息的比特总数为 1260 比特。经过解方程组一剩余篡改信息的比特数位 35 比特，将方程组一解出的结果代入方程组二继续求解，剩余篡改信息的比特数位 14 比特，再将方程组二求出的解代入方程组一进行求解，剩余的篡改信息的比特数位 13 比特，接着再将求出的解代入方程组二求解，剩余篡改信息比特数为 12 比特，将解出的结果再代入方程组一，篡改信息的比特数保持不变。从整个过程，可以看出随着迭代的进行，篡改信息的比特数在逐渐的减少，这是由于迭代补偿机制的应用增加了篡改信息恢复的概率，从而提高了篡改图像恢复的性能。



(a)car 图像中篡改信息比特数



(a)　peppers 图像中篡改信息比特数



(b)　couple 图像中篡改信息比特数

图 7 恢复过程中篡改信息比特数变化情况

3 结论

本文提出了一种基于精细认证和迭代补偿机制的图像认证与恢复算法。在认证的过程算法中包含基本认证和精细认证。由于图像置乱技术，精细认证可能将包含在已做篡改标记的图像块中的真实的像素分配到一个

真实的图像块中而通过认证。结合基本认证和精细认证，通过与操作，并采用本文规定的连通域优化规则对进行与操作之后的结果进行优化，得到定位精度较高的认证结果。在恢复的过程中，借鉴参考共享机制的思想，提出了一种迭代补偿机制，在一定程度上提高了篡改信息恢复的概率。

参考文献

[1] 周四清，余英林. 数字图像水印技术及其应用 [J]. 数据采集与处理，2001,16(3)：353-357.

[2] He Hongjie, Zhang Jiashu, Tai Hengming. A wavelet-based fragile watermarking scheme for secure image authentication [C] // IWDW2006, NewYork: Springer Press, 2006: 422-431.

[3] Liu Hongtao, Shen Ruiming, Chung Fulai. Fragile watermarking scheme for image authentication [J]. Electronics Letters, 2003, 39(12): 898-900.

[4] Liu Shaohui, Yao Hongxun, Gao Wen, et al. An image fragile watermark scheme based on chaotic image pattern and pixel-pairs [J]. Applied Mathematics and Computation, 2007, 185(2): 869-882.

[5] Wang Shiming, Chen Weiche. A majority-voting based watermarking scheme for color image tamper detection and recovery [J]. Computer Standards and Interfaces, 2007, 29(5): 561-570.

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

[7] Li Chunlei, Wang Yunhong, Ma Bin, et al. A novel self-recovery fragile watermarking scheme based on dual-redundant-ring structure [J]. Computers and Electrical Engineering, 2011, 37(6): 927-940.

[8] He Hongjie，Chen Fan, Tai Hengming, et al. Performance analysis of a block-neighborhood based self-recovery fragile watermarking scheme [J]. IEEE Transactions on Information Forensics and Security, 2012, 7(1): 185-196.

[9] Zhang Xinpeng, Wang Shuozhong, Qian Zhenxing, et al. Reference Sharing Mechanism for Watermark Self-Embedding [J]. IEEE Transactions on Image Processing, 2011, 20(2): 485-495.

[10] Holliman M, Memon N. Counterfeiting Attacks on Oblivious Block-wise Independent Invisible Watermarking Schemes [J]. IEEE Transactions on Image Processing, 2000, 9(3): 432-441.

作者简介：程宝田(1985-)，男，硕士研究生，研究方向：数字水印，E-mail:10120381@bjtu.deu.cn；倪蓉蓉(1976-)，女，硕士生导师，副教授，研究方向：数字水印；赵耀(1967-)，男，博士生导师，教授，研究方向：图像编码、数字水印。