窜改认证与内容恢复分离的自嵌入脆弱水印算法*

聂雪莲，戴　青
（解放军信息工程大学 电子技术学院，郑州 450004）

摘　要：为了提高可恢复水印的认证精度和恢复质量，降低虚警率，提出了一种窜改认证与内容恢复分离的自嵌入水印算法。该算法以 2×2 的图像块为认证单位，将认证水印与恢复水印分别嵌入到自身图像块和偏移图像块中，使认证虚警率降为 0；同时提出一种基于线性函数和混沌映射的偏移值选取方案，使图像的恢复质量得到提高。理论分析和实验结果表明，该算法能在抵抗拼贴和量化攻击的同时，准确定位图像窜改并以较高的质量进行恢复。

关键词：脆弱水印；自嵌入；认证精度；虚警率；混沌
中图分类号：TP391　　**文献标志码**：A　　**文章编号**：1001-3695(2012)07-2696-04
doi:10.3969/j.issn.1001-3695.2012.07.081

Self-embedding fragile watermarking algorithm with tamper authentication and content recovery separated

NIE Xue-lian, DAI Qing
(*School of Electronic Technology, The PLA Information Engineering University, Zhengzhou* 450004, *China*)

Abstract: To improve the accuracy of authentication and the quality of reconstructed image, reduce the probability of false rejection, this paper proposed an authentication and recovery separated self-embedding watermarking algorithm. It defined 2×2 image block as an authentication unit. It embedded the authentication watermark and recovery watermark respectively into the block itself and the corresponding excursion block, reducing the probability of false rejection to 0. It proposed an offset value selection sheme based on linear function and chaos mapping to improve the quality of the reconstructed image. Theoretical analysis and simulation results show that the proposed algorithm can not only thwart the collage and VQ attack, but also locate tampered blocks accurately and restore it effectively.

Key words: fragile watermarking; self-embedding; accuracy of authentication; probability of false rejection; chaos

近年来，图像认证技术作为数字水印的一个重要分支，在数字媒体的信息安全领域得到了广泛应用。认证水印[1,2]主要用以检测图像内容的真实性和完整性，自嵌入水印作为认证水印的一种，它从图像本身提取出水印信息，然后嵌入到图像中以完成对窜改的认证，这种水印除了具有窜改检测和定位功能外，还可以实现图像的自我恢复，因而受到越来越多研究者的关注。

1999 年，Fridrich 等人[3]提出了基于 DCT 变换的自嵌入水印算法，该方法利用 JPEG 量化表对 8×8 图像块的高 7 位 DCT 系数进行量化编码，然后采用固定偏移值嵌入到偏移块的 LSB 中。文献[4]通过修改量化表和图像块的偏移值，对 Fridrich 的算法进行了改进，提高了图像的恢复质量，并增强了算法的安全性。文献[5]对文献[4]的窜改判定方法和偏移块选取方案加以改进，进一步提高了认证的准确性和算法的安全性。但是此类自嵌入算法[6,7]存在几个共同的问题：

a)算法将认证信息与恢复信息合二为一，将由图像块 x_i 生成的水印信息嵌入到偏移块 x_j 中，既用于窜改定位，又用于图像内容的恢复。检测时，若由图像块 x_i 生成的水印与 x_j 中提取的水印不相等，则判定图像块 x_i 被窜改。但是这两者不相等，有多种被窜改的可能，简单地判定图像块 x_i 被窜改会造成认证不准确，使虚警率增高。

b)水印嵌入时的图像块偏移值选取不够合理。

c)图像的认证精度较低，大多以 8×8 的图像块为最小认证单位。

针对上述问题，本文提出了一种将窜改定位和内容恢复分离的自嵌入脆弱水印算法，将认证精度提高到 2×2 的图像块。利用图像块内容生成认证信息和恢复信息，分别嵌入到自身图像块和偏移图像块中，采用图像的 ID 和图像块的位置索引确保认证水印的安全性；并提出一种线性函数和混沌系统相结合的偏移值选取方法来确定恢复信息的嵌入位置，减少图像块内容和水印信息同时被破坏的可能性，并对恢复信息已破坏的图像块，利用未被窜改的邻域信息进行恢复，使图像具有较高的恢复质量。

1　偏移值的选取

恢复水印只要嵌入在图像当中，就不可能保证它完全不被窜改，任何一种算法都只能尽可能地避免图像块内容和恢复信息同时遭到破坏，所以偏移值的选取至关重要，其选取应该满

收稿日期：2011-11-17；**修回日期**：2011-12-31　　**基金项目**：河南省科技厅科技攻关计划基金资助项目(102102210053)

作者简介：聂雪莲(1985-)，女，湖北宜昌人，硕士研究生，主要研究方向为数字图像处理(annie630_2011@126.com)；戴青(1963-)，男，副教授，硕导，主要研究方向为多媒体应用技术.

足以下三个条件:

a)偏移值的密钥空间要足够大,以保证偏移值的安全。

b)恢复水印的生成块和偏移块之间应该是一一对应的关系,即文献[8]提出的两个条件。

c)恢复水印的生成块与偏移块应该相距较远,这样才能降低图像块内容和恢复水印同时被修改的几率。

现有的块级认证算法大多采用固定偏移值[3]或者利用线性函数[4]来确定偏移值,虽然能满足条件 b)c),但其密钥空间太小,安全性不高。文献[5]提出了一种利用混沌序列稳定排序法来确定水印嵌入位置的方法:该方法首先采用混沌系统生成伪随机序列 $R=(r_1,r_2,\cdots,r_n)$,然后按照数值大小对 R 进行排序,产生有序序列 $R'=(r_{a_1},r_{a_2},\cdots,r_{a_n})$,根据各元素排序前后的位置关系得到一一映射 $T:i\to a_i$,即将图像块 i 的恢复水印嵌入到图像块 a_i 中。该方法具有较高的安全性,但不满足条件 c),因为各元素排序前后的位置关系无法确定,有可能排序后位置没有发生改变,也有可能与原来的位置相距很近。

本文提出一种线性函数与混沌映射相结合的方法来选取偏移块,具体描述如下:

a)将大小为 $M\times N$(M 和 N 不是 8 的倍数则补 0)的图像分成$\frac{M}{4}\times\frac{N}{4}$的互不重叠的图像块 $B_1,B_2,\cdots,B_{16}$,对每个图像块 $B_i(i=1,2,\cdots,16)$进行二次分割,得到大小为 2×2 的子块 $b_{i1},b_{i2},\cdots,b_{in}$($n$ 为 B_i 中子块的个数)。

b)对于图像块 $B_1,B_2,\cdots,B_{16}$,使用线性函数 $T(i)=(i+p)\bmod 16(6\leqslant P\leqslant 10)$来形成一个循环嵌入链,即将 B_i 中子块的水印信息嵌入到 $B_{T(i)}$ 的子块中,P 值过大和过小都会使 B_i 和 $B_{T(i)}$ 的空间距离太近。

c)利用混沌系统 Logistic 映射 $x_{k+1}^{(i)}=\mu_i x_k^{(i)}(1-x_k^{(i)})(i=1,2,\cdots,16,k=1,2,\cdots,n)$生成 16 个长为 n 的混沌序列 $S_1,S_2,\cdots,S_{16}$:

$$\begin{cases}\mu_i=\mu_1+(i-1)q_1 & \mu_1,\mu_1+15q_1\in(3.46,4)\\ k_1^{(i)}=k_1^{(1)}+(i-1)q_2 & q_2,k_1^{(i)}+15q_2\in(0,1)\end{cases}\tag{1}$$

其中:μ_i 和 $k_1^{(i)}$ 分别为 S_i 的系数和初值;偏移值密钥 key = $\{p,\mu_1,k_1^{(1)},q_1,q_2\}$。

d)对任意图像块 B_i,进行如下操作:将混沌序列 $S_i=\{s_1^{(i)},s_2^{(i)},\cdots,s_n^{(i)}\}$按大小关系进行排序,生成有序序列 $S'=\{s_{a_1}^{(i)},s_{a_2}^{(i)},\cdots,s_{a_n}^{(i)}\}$,则将 B_i 中子块 b_{ij}的恢复水印嵌入到子块 b_{ia_j}中($1\leqslant j\leqslant n$)。

该方法将图像所有的 2×2 子块分成 16 组,利用线性函数来控制每组子块之间的空间距离。对于每一组子块,采用基于混沌映射的方法来选择偏移块,形成了一一对应的关系,混沌系统的伪随机性和初值敏感性增大了算法的密钥空间,提高了算法的安全性。

2　水印算法

2.1　水印的生成与嵌入

图 1 显示了水印生成和嵌入的流程图,具体步骤如下:

1)水印的生成

(1)将大小为 $M\times N$ 的图像 I 最低位和次低位置零,生成图像 $\bar{I}$;然后将 $\bar{I}$ 分成 2×2 的互不重叠的图像块,并利用混沌序列对图像块进行编号。

(2)利用密钥 k_1 对 $\bar{I}$ 的灰度值进行 5 bit 非均匀标量量化,过程如下:

$$f(i,j)=k,\bar{I}(i,j)\in(8k+\delta_k,8(k+1)+\delta_{k+1})\tag{2}$$

其中:$i=1,2,\cdots,\frac{M}{2};j=1,2,\cdots,\frac{N}{2};\bar{I}(i,j)$为 $\bar{I}$ 在(i,j)处的灰度值;$f(i,j)$表示 $\bar{I}(i,j)$的量化值;$k=0,1,\cdots,31;\delta_0=\delta_{31}=0$,$\{\delta_k|k=1,2,\cdots,30\}$为基于密钥 k_1 的取值范围为[-4,4]的随机序列。

(3)对编号为 r 的图像块,设量化后的灰度值为$\begin{pmatrix}f_{r_1} & f_{r_2}\\ f_{r_3} & f_{r_4}\end{pmatrix}$,将 r 中的每个量化值转换为 5 位二进制,按 $f_{r_1}f_{r_2}f_{r_3}f_{r_4}$的顺序进行排列,产生 20 bit 的 0-1 序列 $m_r=m_{r1}m_{r2}\cdots m_{r20}$。

(4)将图像的 ID 和图像块编号 r 转换成二进制序列,连接 m_r、ID 和 r,产生 0-1 序列 $M_r=\{m_r|\mathrm{ID}|r\}$,以 3 bit 为单位对 M_r 进行分组(若不是 3 的倍数则补 0),并对各分组进行异或,生成认证水印 w_A。

(5)计算 $\bar{I}$ 中每个图像块灰度量化值的均值,即$\overline{f_r}=\frac{1}{4}\times\sum_{i=1}^{4}f_{r_i}$,$r$ 为图像块的编号,$f_{r_i}(i=1,2,3,4)$为图像块 r 中的灰度量化值,将$\overline{f_r}$转换成 5 bit 二进制,采用密钥 k_2 对二进制序列进行加密,生成恢复水印 w_R。

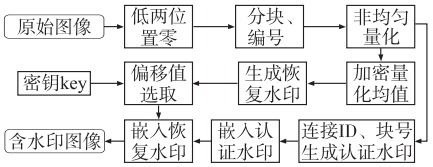


图1　水印生成及嵌入流程

2)水印的嵌入

(1)利用密钥 k_3 从 $\bar{I}$ 的每个图像块的低两位位平面中随机选取 3 个 bit,嵌入各图像块自身的认证水印,得到含认证水印的图像 I^A。

(2)将各图像块的恢复信息嵌入到其对应偏移块最低位和次低位剩余的 5 个 bit 位中,生成含水印图像 I^w。

2.2　窜改认证及图像恢复

设待检测图像为 I',窜改检测及内容恢复流程如图 2 所示。

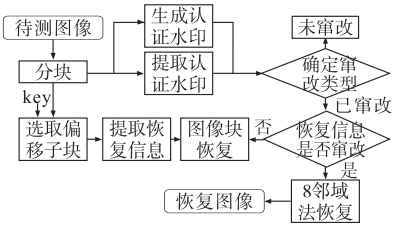


图2　认证及恢复流程

将图像 I'分成 2×2 的互不重叠的图像块,按照与生成水印相同的方法对图像进行编号,对任意的图像块 r,其认证及恢复过程如下:

a)利用密钥 k_3 从图像 I'的低两位中提取出图像块 r 的认证水印 w_{rA}和恢复水印 w_{rR}。

b)利用密钥 k_1 和 k_2 按照生成水印算法得到认证水印 w'_{rA}和恢复水印 w'_{rR}。

若 $w_{rA} = w'_{rA}$，则判定图像块 r 没有被窜改，通过认证；若 $w_{rA} \neq w'_{rA}$，则判定图像块 r 被窜改。

如果图像块 r 被窜改，再对图像块 r 的偏移子块 $r+k$ 进行判定：若 $w_{(r+k)A} = w'_{(r+k)A}$，则判定图像块 r 的恢复水印没有被破坏，利用恢复水印 w_{rR} 对其进行恢复：

$$F_{r_i} = w_{rR}^d \times 8 \quad i = 1,2,3,4 \tag{3}$$

其中：F_{r_i} 为恢复的图像块 r 的灰度值，w_{rR}^d 为 w_{rR} 的十进制值。若 $w_{(r+k)A} \neq w'_{(r+k)A}$，说明图像块 r 的恢复水印被窜改，则通过 r 的邻域块对其进行恢复：

$$F_{r_i} = \text{mean}(F_{r_R^n}) r_i^n \in N_8(r_i) \text{ 且 } r_i^n \notin r \tag{4}$$

其中：mean( ) 为均值函数；$N_8(r_i)$ 为像素点 r_i 的 8 邻域；r_i^n 为 r_i 的 8 邻域内未被窜改的像素点。对所有图像块完成以上操作，则得到窜改定位结果及恢复后的图像。

3 性能分析及实验仿真

3.1 水印的不可见性

为了衡量水印嵌入所引起的原始图像的失真度，定义峰值信噪比(PSNR)为

$$\text{PSNR} = 10 \log \left[\frac{255^2}{\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} [I(i,j) - I^w(i,j)]^2} \right] \tag{5}$$

本文算法将水印嵌入到原始图像的最低位和次低位，对于任一像素点，$I(i,j) - I^w(i,j)$ 的取值范围为 $\{0,1,2,3\}$，且四者的概率均为 1/4，由于图像中各像素点灰度值的改变都相对独立，所以 $[I(i,j) - I^w(i,j)]^2$ 的数学期望为 $E([I(i,j) - I^w(i,j)]^2) = \sum_{k=0}^{3} k^2 \times 1/4 = 3.5$，由此可以得到 PSNR 的数学期望值为 $E(\text{PSNR}) = 10\log \frac{255^2}{3.5} = 42.69$ dB，原始图像失真最大的情况是最低两个位平面中所有 bit 都发生了改变，即每个像素点灰度值的改变量都为 3，则 PSNR = 10 log $255^2/3^2$ = 38.59 dB，即用该算法嵌入水印后图像的 PSNR 一定不小于 38.59，大于文献[9]提出的含水印图像的峰值信噪比必须大于 35 dB 的标准。图 3 显示了嵌入水印前后的 Lena 图像(PSNR = 43.475)，从图 3(b)可以看出，水印具有很好的不可见性。

(a)原始图像 (b)含水印图像

图3 水印的不可见性

3.2 算法的窜改定位及恢复能力

为了测试本文算法的窜改定位和恢复能力，本文从标准图像库中选取了大量图像，并全部转换为 512 × 512 的 8 bit 灰度图像进行实验，下面是部分实验结果，认证图像中黑色表示通过认证，白色表示未通过认证。以 Lena 图像为例，对含水印图像进行如下窜改：a) 对图像进行局部修改，如图 4(a) 所示，图 4(b) 为利用本文算法进行窜改检测和恢复的结果；b) 在图像上添加文字，如图 4(c)，其认证结果和恢复图像如图 4(e) 所示。利用文献[4,5]的算法分别进行 a)b) 的窜改操作，从实验结果(如图 4(d)(e)) 可以看出，与文献[4,5]相比，本文算法的窜改定位精度更高、更准确，对图像的恢复效果更好。

(a)局部修改后的图像 (b)本文算法认证及恢复结果

(c)文献[4]认证及恢复结果 (d)添加文字后的图像

(e)本文算法认证及恢复结果 (f)文献[5]认证及恢复结果

图4 窜改认证及恢复实验结果

文献[4,5]的可恢复块级认证水印算法，为了抵抗量化攻击和有利于图像内容的自恢复，多是采用块相关的方法嵌入水印，即将基于图像块内容的水印信息嵌入到偏移图像块中，但是这类算法在对图像块进行窜改判定时会存在不确定性，使虚警率增高，造成认证不准确。本文算法将认证水印和恢复水印分离，将认证水印嵌入到自身图像块，若图像块没有发生窜改，检测时生成的水印信息和提取出的水印信息完全相同，虚警率为 0，因此可以清晰地界定图像被窜改的区域。在对图像进行恢复时，由于在选取恢复水印的嵌入位置时采用线性函数约束了每一对图像块之间的空间位置，降低了图像块内容和恢复信息同时被窜改的可能性，所以使图像恢复质量得到了提高；同时对于恢复信息被窜改的图像块，并不是直接放弃，而是通过像素点的 8 邻域中未被窜改的像素点对其进行恢复，进一步提高了恢复质量。表 1 显示了对标准图像库中一组不同图像进行 15% 中心剪切，并利用文献[4,5]和本文算法恢复后图像的 PSNR。从实验结果可以看出，本文算法的恢复质量好于文献[4]和[5]的算法。

表 1 图像恢复后的 PSNR

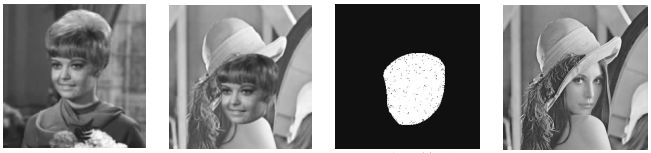
image PSNR	Lena	baboon	boat	peppers	camera
本文算法	39.07	39.13	39.10	39.31	38.94
文献[4]算法	35.21	35.09	35.32	35.77	34.85
文献[5]算法	37.68	37.64	37.79	37.53	37.31

3.3 算法的安全性

目前针对基于图像块认证水印的攻击方式主要分为内容攻击和量化攻击两大类。内容攻击是指保持水印信息不变，只改变图像内容的攻击方式。本文算法利用图像块本身的内容信息来生成水印，在发生内容攻击时，由待检测图像生成的水印信息会发生改变，即使保持嵌入在低位平面中的水印信息不变，也能准确地检测出图像内容的窜改。量化攻击是指攻击者通过搜集多幅由同一方案生成的含水印图像，并将图像块分成一系列等价类，然后在不同等价类中挑选图像块拼接成新的图像来通过认证，本文算法在生成水印时增加了原始图像的 ID 和图像块的位置信息，使认证信息的产生依赖于特定的图像块，因此可以很好地抵抗量化攻击。除此之外，密钥的搜索空间也直接影响着算法的安全性，对于密钥空间较小的算法，攻击者可以通过穷力搜索的方式来获取密钥。为了提高算法的安全性，本文在认证水印的生成过程和恢复水印的嵌入位置选

择上都使用了混沌系统进行加密。混沌系统具有极端的初值敏感性、伪随机性和不可预测性,单混沌序列的密钥空间key > 0.5×10^{16},而复合混沌序列的密钥空间更是高达 10^{46} 以上,因而攻击者难以通过蛮力攻击手段来破解系统密钥。图 5 和图 6 分别显示了算法对 Lena 图像进行内容攻击和量化攻击的实验结果。

对 Lena 图像进行内容攻击:利用 woman 图像(图 5(a))脸部区域的高六位替换 Lena 脸部图像块的高六位,保持低两位中的水印信息不变,得到内容窜改图像如图 5(b)所示,图 5(c)(d)分别显示了认证结果和恢复图像。从实验结果可以看出,本文算法能够较好地抵抗内容攻击,且恢复质量较好。



(a) woman含水印图像　(b) 内容攻击后的图像　(c) 认证结果　(d) 恢复图像

图5　量化攻击实验结果

对 Lena 图像进行量化攻击:图 6(a)是交换 Lena 含水印图像左眼和右眼区域图像块之后的图像,(b)为认证差值和恢复图像。从实验结果可以看出,本文算法能够很好地抵抗量化攻击,也能以较好的质量对图像进行恢复。

(a)量化攻击后的图像　(b)量化攻击认证和恢复结果

图6　量化攻击实验结果

4　结束语

本文提出一种窜改定位和内容恢复分离的自嵌入双水印算法,解决了现有带恢复功能的块级自嵌入水印算法认证精度低、虚警率高和恢复质量不理想的问题。该算法利用图像块内容生成认证比特,连接图像的唯一索引和图像块的位置信息共同生成认证水印,嵌入到自身图像块中,使水印算法的虚警率降为 0,且能够很好地抵抗量化攻击和拼贴攻击;同时通过对偏移值选取方案和恢复算法的改进,在增强水印算法安全性的同时,大大提高了图像内容的恢复质量。理论分析和实验仿真表明:该算法具有较好的水印不可见性,对窜改定位准确且精确度高,能够很好地抵抗量化、拼贴等一些常见的攻击,并对窜改具有很好的恢复质量。下一步将研究如何提高水印算法对随机噪声、JPEG 压缩等非恶意窜改的鲁棒性,使水印算法具有更广范的实用范围。

参考文献:

[1] MEENAKSHIDEVI P, VENKATESAN M, DURAISWAMY K. A fragile watermarking scheme for image authentication with tamper localization using integer wavelet transform[J]. Journal of Computer Science, 2009, 5(11): 831-837.

[2] CHAMLAWI R, KHAN A, USMAN I. Authentication and recovery of images using multiple watermarks[J]. Computers and Electrical Engineering, 2010, 36(3): 578-584.

[3] FRIDRICH J, GOLJAN M. Images with self-correction capabilities[C]// Proc of International Conference on Image Proceeing. [S. l.]: IEEE Press, 1999: 25-28.

[4] 张鸿兵,杨成. 图像的自嵌入及窜改的检测和恢复算法[J]. 电子学报, 2004, 32(2): 196-199.

[5] 和红杰,张家树. 基于混沌置乱的分块自嵌入水印算法[J]. 通信学报, 2006, 27(7): 80-86.

[6] HASAN Y M Y, HASSAN A M. Tamper detection with self-correction hybrid spatial-DCT domains image authentication technique[C]// Proc of IEEE International Symposium on Signal Processing and Information Technology. [S. l.]: IEEE Press, 2007: 369-374.

[7] LIEW S C, ZAIN J M. Tamper localization and lossless recovery watermarking scheme[C]//Proc of the 2nd International Conference on Software Engineering and Computer Systems. [S. l.]: Springer, 2011: 555-566.

[8] HE Hong-jie, ZHANG Jia-shu, WANG Hong-xia. Synchronous counterfeiting attacks on self-embedding watermarking schemes[J]. International Journal of Computer Science and Network Security, 2006, 6(1B): 251-257.

[9] SAMUEL S, PENZHORN W T. Digital watermarking for copyright protection[C]//Proc of the 7th AFRICON Conference in Africa. [S. l.]: IEEE Press, 2004: 953-957.