一类有效的脆弱型数字水印技术

张小华[1],孟红云[2],刘 芳[3],焦李成[1]
(1.西安电子科技大学雷达信号处理重点实验室,陕西西安 710071;
2.西安电子科技大学应用数学系;3.西安电子科技大学计算机学院,陕西西安 710071)

摘 要: 数字水印是多媒体数字产品版权保护和内容抗篡改的重要技术之一.本文在对以往脆弱型数字水印分析基础之上,给出了三种基于混沌系统的脆弱型数字水印技术,充分利用混沌系统对初值敏感和伪噪音等特性,使得相同子块在不同混沌状态下可能隐藏不同的水印信息,从而克服 Holliman 攻击和矢量拼贴攻击,有力的增强该脆弱型水印技术抵抗恶意攻击的能力.实验结果表明该算法在不破坏宿主图像视觉质量基础上,可精确地检测和定位对图像内容的局部恶意篡改,同时该算法很容易推广到其他数字媒体.

关键词: 鲁棒型数字水印;易碎型数字水印;混沌系统;内容防篡改
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2004) 01-0114-04

A New Kind of Efficient Fragile Watermarking Technique

ZHANG Xiao-hua[1],MENG Hong-yun[2],LIU Fang[3],JIAO Li-cheng[1]
(1. *State Key Laboratory of RSP, Xidian University, Xi'an, Shaanxi* 710071, *China*;
2. *Dept. of Applied Math., Xidian University*; 3. *Dept. of Computer, Xidian University, Xi'an, Shaanxi* 710071, *China*)

Abstract: Digital watermarking is a technique for inserting information (the watermark) into digital multimedia, which can be later extracted or detected for a variety of purposes including identification or authentication purposes. Based on analysis of the past fragile watermark techniques, three new efficient fragile watermark techniques are proposed, benefiting from the good properties of chaos, including the ease of their generation, their sensitive dependence on their initial condition and noise like, thus a same image sub block can embed different information at the same location and resist Holliman's attack. The scheme can detect any modification made to the image and indicate the specific locations that have been modified, because any modification would be reflected in a corresponding error in the extracted watermark.

Key words: robust watermarking; fragile watermarking; chaos system; content tampering resist

1 引言

数字水印是一门新兴的边缘学科,在通信、计算机、密码学等学科有着广泛应用前景.根据其对攻击的抵抗能力,数字水印可分为:鲁棒型水印、易碎型水印、半易碎性水印.鲁棒型水印主要在应用于数字图像的知识产权的保护,防止非法获取[1],对修改具有较强的鲁棒性.易碎型水印则主要应用在图像内容的完整性和可信性的验证[2~4],对修改具有较强的敏感性.半易碎数字水印则可区分偶然修改和恶意篡改,只对恶意篡改较为敏感.为了提高对恶意篡改的检测和定位的精度,往往采用块不相关水印技术[2~4].该类算法的优点在于算法简单,定位精度高,但存在安全缺陷.Fridrich[5]和 Holliman[6]等人分别独立地指出了这一类方法的安全缺陷并成功地实施了攻击.为克服这一缺陷,研究人员提出了很多改进算法.Fridrich[5]提出通过结合周围像素来确定嵌入的水印比特,从而引入基于图像的不确定性.Wong[6]等人提出另外一种解决方案,使用多个参数来控制嵌入的水印比特.本文基于对块不相关水印技术的深入分析,提出三种改进的脆弱型水印技术,不仅继承原算法的优点,同时提高了它们抵抗恶意攻击的能力.

2 块不相关水印技术

在隐藏水印信息 W 之前,将宿主信息 X 分为不相交的子块 $\{X_1, X_2, \cdots, X_n\}$,然后将水印信息 W_i 依据密钥 K 独立地嵌入到子块 X_i 中,结果记为 X'_i,其中 X'_i 完全取决于 X_i,W_i 和密钥 K,故统称此类方法为:块不相关水印技术,其隐藏过程可表示为:

$$X' = E_K(X, W) = E_K(X_1, W_1) \parallel E_K(X_2, W_2) \parallel \cdots \parallel E_K(X_n, W_n) \parallel \quad (1)$$

而水印信息的提取过程表示为:

$$W' = D_K(X') = D_K(X'_1) \parallel D_K(X'_2) \parallel \cdots \parallel D_K(X'_n) = W'_1 \parallel W'_2 \parallel \cdots \parallel W'_n \quad (2)$$

收稿日期:2002-05-23;修回日期:2003-07-01
基金项目:国家自然科学基金(No.60073053,No.60133010)

如果 $D_K(X'_i)=D_K(X'_j)$，则称 X'_i 和 X'_j 等价，这样 $\{X'_1,X'_2,\cdots,X'_n\}$ 就可被分为 m 个等价类集合 $\{C_1,C_2,\cdots,C_m\}$，m 为 W_i 可能取值的个数，一般取 m 为 2. 如果 X'_i 和 X'_j 同属于 C_k，将 X' 中的子块 X'_i 和 X'_j 相互替换，并不影响水印信息的提取结果. 所以在攻击者已知水印信息 W 或拥有多幅隐含相同的水印的可信图像时，很容易在保持所提取水印不变的情况下，恶意篡改图像 X'. 该算法之所以易受攻击就在于：

(1)每一子块的水印信息的隐藏和提取与其他子块并不相关，一个子块的改变并不影响其他子块水印信息的提取结果；

(2)在整个隐藏过程中等价类集合不发生变化或变化很少，在已知水印信息或密钥情况下，攻击者很容易获得等价类集合的大部分或全部元素；

(3)由于密钥和水印信息资源数量的有限性，如一部数字相机可能只拥有一个密钥和一种水印信息，从而导致所拍摄的不同图像在相同的位置隐含相同的水印.

为了提高该算法的抵抗攻击能力，可以采取如下措施：

(1)构造与宿主媒体 X 内容有关的数字水印信息，增加其多样性；

(2)增加密钥的个数. 如可基于该数字相机唯一密钥 K 和图像局部特征 F 构造合成密钥 $\hat{K}=Integrate(K,F)$，使得不同图像嵌入水印的密钥不同；

(3)增加获得等价类集合元素的难度，或者设计动态变化的等价类集合 $C(t)$；

(4)增加子块之间的相关性或增加子块的大小.

3　混沌动力系统

混沌现象是在非线性动力系统中出现确定性的、整体类似随机的过程，这种过程既非周期又不收敛，并且对初始值有极其敏感的依赖性.

一维离散非线性动力系统可定义如下：

$$x_{k+1}=\tau(x_k) \tag{3}$$

其中，$x_k\in V$，称之为状态，而 $\tau:V\to V$ 为混沌映射，作用在于由当前状态 x_k 切换下一个状态 x_{k+1}. 从一个初始状态 x_0 开始，反复调用映射 τ，就可得到一个混沌序列 x_k. 这一离散序列称为该离散时间动力系统的一条轨迹.

一类非常简单却被广泛研究的动力系统是 logistic 映射，其定义如下式：

$$x_{k+1}=\mu x_k(1-x_k) \tag{4}$$

其中，μ 称为分枝参数. 混沌动力系统的研究工作指出，当 $3.5699456\cdots<\mu\leqslant 4$ 时，logistic 映射处于混沌状态. 独立选取两个初始值 x_0 和 y_0，则离散序列的互相关函数为：

$$\begin{aligned}c(k)&=\lim_{N\to\infty}\frac{1}{N}\sum_{i=0}^{N-1}(x_i-\bar{x})(y_{i+k}-\bar{y})\\&=\int_0^1\int_0^1\rho(x,y)(x-\bar{x})(\tau^k(y)-\bar{y})\,dxdy\\&=0\end{aligned} \tag{5}$$

其中 $\rho(x,y)$ 为联合概率密度函数，显然 $\rho(x,y)=\rho(x)\rho(y)$，$\bar{y}$ 为序列 y_i 的均值，$\tau^k(y)$ 为以 y 为初始点进行 k 次迭代的结果. 而序列的自相关函数 ACF 则等于 delta 函数 $\delta(k)$. 以上特性表明，尽管混沌动力系统具有一定的确定性，但其遍历统计特性等同于白噪声，具有形式简单，初始条件的敏感性.

4　改进型数字水印算法

在一般情况下水印信息 W 与所要保护的媒体 X 没有关系，它可能是随机序列或二值图像或字符文字等. 在鲁棒型数字水印系统中，数字水印只是媒体所有者的身份标志，所以只与所有者有关. 而在脆弱型数字水印系统中，数字水印的作用在于证明宿主媒体的真实性和完整性，所以本文认为对于脆弱型水印系统，水印信息应能够反映宿主媒体的内容. 在本文中以原始图像二值化图像作为水印信息，由此所构造的水印信息具较强的抗攻击能力.

Yeung 和 Mintzer[2] 提出了一种脆弱型数字水印方法，主要思想为 如果 $D_K(X_i)=W_i$，则 $X'_i=X_i$，否则寻找满足下式的 X'_i：

$$\begin{cases}D_K(X'_i)=W_i\\\min\|X_i-X'_i\|\end{cases} \tag{6}$$

D_k 为水印提取函数，定义为 $D_K=(X'_i)=LUT(X'_i)$，其中 LUT 为依据密钥 K，所构造的查询表. 由于在隐藏过程中，LUT 并不发生变化，所以在已知 W 的情况下，攻击者很容易获得 LUT 的知识，实施恶意攻击.

4.1　基于图像特征的密钥合成算法

对于密钥和所嵌入的水印信息完全相同的情况下，同一位置可能会隐含相同的水印信息，所以攻击者交换两个可信图像同一位置的图像块则不会影响提取的水印信息. 为了提高算法的安全性，本文采用密钥合成算法，即基于唯一密钥 K 和当前图像局部特征 F，构造一个与图像内容有关的密钥 $\hat{K}$. 目的在于增加密钥的多样性，克服密钥数量的限制. 图像块局部特征 F 的选取，对算法的性能将起到决定性的作用，应满足以下条件：

(1)唯一性，要求对任意两幅图像，所产生的密钥相同的概率较小.

(2)稳定性，恶意篡改，可能导致图像特征变化，以致使产生不同的合成密钥 K'，导致完全不同的认证结果，希望用来提取局部特征的图像块遭遇攻击的概率应尽可能小.

由于本文以图像块的均值作为图像特征 F，所以要使唯一性高，则图像块的大小应尽可能地大，而为了减小受攻击的概率图像块的尺寸应尽可能的小，所以必须在唯一性和稳定性之间进行折衷. 本文取图像块大小为 8×8，至于图像块的选取是基于 K 随机选取的，当然图像块也可以是一些重要的区域. 由本文的第二部分讨论可知，混沌序列具有一定的确定性，形式简单，且对初始条件较为敏感等特性，所以本文取初始密钥 $K=(x_0,\mu)$，合成密钥 $\hat{K}=(x_F,\mu)$，其中 x_F 是由 x_0 经式(4) F 次迭代所得.

4.2　等价类动态变化算法

块不相关水印算法的一个重要的缺点，就是攻击者通过分析大量的可信图像，可获得查询表 LUT 的信息. 为了提高安全性，在 X_i 中隐藏 W_i 时，可依据状态信息 x_i 构造动态查

询表 LUT_i,这样在整个隐藏过程中同一个 X'_i 在状态 x_i 可能隐藏 0,而在状态 x_j 就有可能隐藏 1,LUT_i 不再是一个静态表,而变为一个随状态信息 x_i 而变的动态查询表,此时攻击者想得到关于查询表 LUT_i 的信息或等价类的信息,已变得完全不可能,除非得到状态信息 x_i. 具体算法为:

(1)基于初始密钥 $K=(x_0,\mu)$和当前图像 X 构造合成密钥 $\hat{K}=(x_F,\mu)$,并二值化处理原始图像 X 来构造水印信息 W;

(2)依据状态 x_i 构造动态查询表 LUT_i,将 W_i 嵌入到 X_i 中 如果 $D_{K_i}(X_i)=W_i$,则 $X'_i=X_i$,否则根据式(6)修改 X_i;

(3)切换到状态 $x_{i+1}=\mu x_i(1-x_i)$,继续隐藏 W_{i+1},如果已隐藏完所有的水印信息则结束. 提取算法较为简单,通过 $W_i=D_{K_i}(X'_i)=LUT_i(X'_i)$很容易求得水印信息 W.

4.3 嵌入信息动态变化算法

动态更新等价类算法,可大大增加获得等价类集合元素的难度,但每一次隐藏 W_i,都必须重新构造查询表 LUT_i,相对于原始算法计算量变大,解决的方法就是每隔 k 步更新一次查询表 LUT_i,k 越小抵抗攻击的能力就越强,计算量就越大;k 越大抵抗攻击的能力就越弱,计算量却越小. 另外一种方法就是让查询表 LUT_i 保持不变而让嵌入的信息发生动态变化. 我们知道攻击者之所以能得到等价类信息,是因为攻击者知道:$D(X'_i)=W_i$. 如果查询表 LUT 保持不变,而要求 $D(X'_i)=W_i\oplus V_i$,其中:

$$V_i=\Psi(x_i)=\begin{cases}0 & x_i<0.5\\ 1 & else\end{cases} \tag{7}$$

这样在整个隐藏过程中同一个 X'_i 在状态 x_i 可能隐藏 0,而在状态 x_j 下却有可能隐藏 1,攻击者只知道 W_i 是隐藏中 X'_i,但其中具体隐藏的是 0 还是 1 就无从而知. 因为:

$W_i=0$,当 $V_i=1$ 时,$D(X'_i)=1$,而当 $V_i=0$ 时,$D(X'_i)=0$;

$W_i=1$,当 $V_i=1$ 时,$D(X'_i)=0$,而当 $V_i=0$ 时,$D(X'_i)=1$;

所以该算法不仅抵抗能力强. 而且计算量小,具体算法如下:

(1)基于初始密钥 $K=(x_0,\mu)$和当前图像 X 构造合成密钥 $\hat{K}=(x_F,\mu)$,并二值化处理原始图像 X 构造水印信息 W;

(2)依据密钥 x_F 和 μ 构造静态查询表 LUT;

(3)依据状态 x_i 构造状态辅助信息位 V_i,将 $W_i\oplus V_i$ 嵌入到 X_i 中 若 $D_{K_i}(X_i)=W_i\oplus V_i$,则 $X'_i=X_i$,否则根据式(6)修改 X_i 为 X'_i;

(4)切换到状态 $x_{i+1}=\mu x_i(1-x_i)$,隐藏 W_{i+1},如果已隐藏完所有的水印信息则结束.

相应的提取函数为:$W_i=D(X'_i)\oplus V_i$,也很容易实现.

4.4 子块相关算法

由本文的第二节分析可知,子块不相关算法易受攻击的一个重要原因在于各个子块之间不相关,所以增加相关性也是提高此类算法抗攻击能力的一个重要措施,同时结合以上两个算法的优点设计子块相关算法就变为设计如下的嵌入函数和提取函数:

$$\begin{aligned}X'_i&=E_K(X_i,W_i\oplus f(X'_{i-1},X'_{i-2},\cdots,X'_{i-k},x_i))\\ W_i&=D_K(X'_i)\oplus f(X'_{i-1},X'_{i-2},\cdots,X'_{i-k},x_i)\end{aligned} \tag{8}$$

为了简单,我们设隐藏函数 E_K 和提取函数 D_K 和前两个算法一样,取 $k=3$,也就是说在 X_i 中隐藏 W_i 只与前面的三个子块 $X'_{i-1},X'_{i-2},X'_{i-3}$ 以及状态 x_i 有关,在本文中取:

$$f(X'_{i-1},X'_{i-2},X'_{i-3},x_i)=\begin{cases}0 & \theta(X'_{i-1})+\theta(X'_{i-2})>\theta(X'_{i-3})+\varphi(x_i)\\ 1 & \theta(X'_{i-1})+\theta(X'_{i-2})\leqslant\theta(X'_{i-3})+\varphi(x_i)\end{cases} \tag{9}$$

其中 $\theta(X'_{i-1})$ 可以是子块 X'_{i-1} 的均值或方差,也可以是由 X'_{i-1} 决定的其特征值. 假设其特征值取值范围为:$[Min_X,Max_X]$,而 $\varphi(x_i)$ 为由状态信息 xi 决定的一个临时状态值满足 $\varphi(x_i)=Min_X+x_i(Max_X-Min_X)\in[Min_X,Max_X]$. 具体嵌入算法只需上一嵌入算法的第 3 步更换为:依据状态 x_i 构造状态值 $\varphi(x_i)$,然后计算 $V_i=f(X'_{i-1},X'_{i-2},X'_{i-3},x_i)$,将 $V_i\oplus W_i$ 嵌入到 X_i 中 如果 $D_{K_i}(X_i)=W_i\oplus V_i$,则 $X'_i=X_i$,否则根据式(6)修改 X_i 为 X'_i.

该算法的优点在于增加各个子块之间的相关性,使得攻击者即使知道密钥 x_0 和 μ,也很难去伪造媒体信息,特别是 k 较大时 缺点在于随着 k 增大,篡改定位精度随之变差.

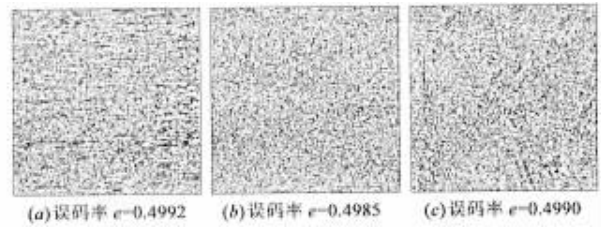
5 实验结果

为了证明算法的有效性,本文分别将给出的三个算法应用在标准图像“Camera Man”中. 本文取初始密钥 $K=(x_0=0.1234,\mu=4.0)$,其中图 1(a)为宿主图像,即保护对象. 图 2(b)为由宿主图像所构造的水印信息. 隐含水印信息的宿主图像的峰值信噪比 $PSNR$ 分别为:46.254、47.621、47.125.

(a)原始图像 (b)水印信息 (c)伪造图像

图 1

为了测试算法对密钥的敏感性,我们在隐藏时,取密钥 $K=(x_0=0.1234,\mu=4.0)$ 提取时,取密钥 $K'=(x_0=0.1235,\mu=4.0)$,所图提取的水印信息如图 2 所示,误码率分别为:0.4992、0.4985、0.4990. 由实验结果可知,即使密钥发生细微的变化,提取的水印信息也会发生较大的变化,而此时的误码率已非常接近 0.5,所提取的水印信息近似于随机噪音. 由此可见如果攻击者在不知道密钥情况下,想要成功检测水印信息是不可能的.



(a)误码率 e=0.4992 (b)误码率 e=0.4985 (c)误码率 e=0.4990

图 2

为了测试算法的对局部篡改的检测和定位能力,本文将隐含水印的信息的数字图像右边矩形建筑物(大小为:20 ×

38，左上角点坐标：(193，112))复制到图像的右边同样大小区域(对应左上角点坐标：(14，112))．图 3 为所提取的水印信息和当前图像(图 1(*c*))二值化图像的比较结果，黑色点代表颜色值相同的点，而白色点代表颜色值不相同的点．实验结果表明本文算法可以精确地检测和定位篡改的位置．

(*a*)第一种算法比较结果　(*b*)第二种算法比较结果　(*c*)第三种算法比较结果

图 3

6　结论

计算机网络为多媒体信息传输带来了方便，同时也带来挑战．数字水印技术为解决该问题带来了希望．本文正是在此背景下，基于对以往方法分析的基础之上，提出了三种基于混沌的脆弱型数字水印技术，其优点在于：有效地利用了混沌系统对初值敏感和伪噪音特性，使得攻击者即使在已知水印信息和隐藏位置的情况，也不可能去伪造一个含有相同水印信息的赝品媒体．相对于半易碎数字水印，本文算法的缺点在于不能有效的区分偶然攻击和恶意攻击，如何将本文算法推广到半易碎数字水印技术，是本文以后研究的一个重要方向．

参考文献：

[1] I J Cox，J Kilian，T Leighton，T Shamoon．Secure spread spectrum watermarking for multimedia[J]．IEEE Trans，1997，IP-6(12)：1673 – 1687．

[2] M Yeung，F Mintzer．An invisible watermarking technique for image verification[A]．Proc．Int．Conf．Image Processing[C]．Barbara California，1997．680 – 683．

[3] Wong P W．A public key watermark for image verification and authentication[A]．Proc．IEEE ICIP[C]．Chicago，USA，1998．425 – 429．

[4] J Zhao，E Koch．Embedding robust labels into images for copyright protection，Intellectual Property Rights New Technologies [A]．Proc．KnowRight'95 Conf[C]．California，1995．242 – 251．

[5] Matth Holliman，Nasir Memon，Counterfeiting attacks on oblivious block-wise independent invisible watermark schemes[J]．IEEE Trans，2000，IP-19(3)：432 – 441．

[6] Fridrich J，Goljan M，Memon N．Further attacks on Yeung-Mintzer watermarking scheme[A]．Proc．SPIE Security and Watermarking of Multimedia Contents[C]．San Jose，California，2000．428 – 437．

作者简介：

张小华　男，1974 年 8 月生于陕西，西安电子科技大学电路与系统专业博士研究生，主要从事图像处理、小波分析、进化算法方面研究．

孟红云　女，1975 年 5 月生于陕西，西安电子科技大学应用数学专业博士研究生，主要从事进化算法、数字水印方面研究．

刘　芳　女，1963 年生于湖南华容，西安电子科技大学教授，主要研究方向为人工智能、信息安全、电子商务．