基于图像特征的易损水印技术 [1]

钟 桦 刘 芳 焦李成

(西安电子科技大学雷达信号处理重点实验室 西安 710071)

摘 要 该文在 Yeung-Mintzer(1997) 的方法基础上提出了一种新颖的易损水印技术. 为解决原算法存在的安全问题, 提出利用基于图像特征的混沌密钥控制易损水印的嵌入. 理论分析和实验结果证明该算法可以有效地抵制由原算法所产生的攻击问题. 由于水印在空域逐个像素地进行嵌入, 所提出的算法具有良好的局部修改检测性能.

关键词 易损水印, 图像认证, 图像特征, 混沌密钥, 攻击

中图号 TN911.73, TP391.4

1 引 言

随着多媒体技术的飞速发展和日趋完善, 任何人都可以方便地对数字媒体, 如音乐, 视频或图像等进行修改. 这使得人们对数字媒体的完整性, 媒体内容的真实性产生怀疑. 为解决这一问题, 需要一种新的技术来保护数字媒体. 作为数字水印技术的一个分支, 认证水印 (authentic watermarking) 通过嵌入水印来达到这一目的. 到目前为止已经提出了很多种认证水印技术, 大体上分为易损水印和半易损水印两类. 半易损水印主要用于图像的内容认证. 由于一些图像处理操作, 例如 JPEG 压缩并不破坏图像的内容, 因此需要半易损水印对破坏图像内容的一些蓄意操作进行检测. 但是在某些应用中, 例如法律证据图像, 医疗图像等细节敏感图像, 甚至 JPEG 压缩也会破坏图像的细节. 易损水印正是针对这一类应用而设计的. 由于易损水印对任意修改都具有很高的敏感度, 因而可以应用于图像作为一个整体, 完整性不可破坏这一情形. 即使是非常轻微的修改, 易损水印都能够准确地检测到这一修改并精确定位.

早期的易损水印都是通过修改最不重要比特 (LSB) 来嵌入水印 [1,2] . 如果数字媒体遭到修改, 通过从 LSB 提取水印就可检测到这些修改. 这种方法的缺点是攻击者可以修改媒体的内容而保持其 LSB 不变, 从而使认证失败. Yeung-Mintzer 等 [3] 提出使用二值函数把灰度值 0 到 255 映射为 0 或 1, 并以一个二维标识作为水印, 修改图像的像素灰度使其映射值与标识中相应的比特相同从而嵌入易损水印. 该技术把水印的提取与整个灰度值联系起来而不局限于 LSB, 从而克服了以上问题. 另外, 这一技术还具有良好的局部检测性能, 可以把认证过程具体到每一个像素. 但是在文献 [4,5] 中均已指出这种技术具有致命的缺点, 即每一个像素中嵌入的水印比特具有确定性. 根据这一点, 攻击者可以轻易地伪造加水印图像来实施攻击. 针对这一缺点, Fridrich 等 [6] 提出了一种易损水印方案, 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. 这种方法的优点是可以抵制文献 [4,5] 中提出的攻击, 但是其局部检测性能下降.

本文提出的易损水印方案可以有效地抵制各种攻击, 并且保持了 Yeung-Mintzer 水印方案的局部检测性能. 在所提出的水印方案中, 所有图像均采用同一数字相机密钥 (camera key), 因此具有很好的可行性. 实验和分析结果证明了该算法的有效性和可行性.

[1] 2002-06-20 收到, 2002-12-30 改回
国家自然科学基金 (60133010,60073053) 资助项目

2 基于图像特征的易损水印

2.1 问题分析

我们先简要介绍一下 Yeung-Mintzer 的方法．以灰度图像为例，对于彩色图像可以进行类似的操作．首先根据密钥产生一个查询表 (LUT)．该密钥与图像无关．查询表对应一个二值映射函数 $f:\{0,1,\cdots,255\}\to\{0,1\}$，用于把灰度值 0 到 255 映射为 0 或 1．然后根据所嵌入的标识 L，逐个修改像素值使其灰度值所映射的比特与标识的相应比特相同，如 (1) 式所示．

$$L_{i,j}=f(g_{i,j}) \tag{1}$$

分析可知，如果采用同一密钥并嵌入相同的标识作为水印，则可能遭到以下攻击 [5]：

(1) 估计标识 L 和映射函数 f：显然每一幅加水印图像的特定位置 (i,j) 对应相同的标识比特，即 $L_{i,j}$．利用这一知识和足够多的加水印图像，就可以完全确定映射函数 f，从而也可以确定标识 L．一旦攻击者知道了标识 L 和映射函数 f，攻击者就可以任意地修改和伪造加水印图像．

(2) 拼贴攻击 (collage attack)：通过拼贴不同图像的局部块并保持各个块在图像中的位置不变来完成．拼贴后得到的图像即为伪造的加水印图像．拼贴攻击适用于任何局部嵌入且与图像无关的水印方案，即使映射函数与像素位置有关，这种攻击依然有效．

这两种攻击成功的关键是，不同加水印图像在图像的固定位置含有相同的嵌入信息．即使不知道某一位置确切的比特信息，仍然可以伪造任意的加水印图像 [4,5]．显然，为了抵制这种攻击，必须使局部位置的嵌入信息随图像而不同．一种最简单的想法是，对每一幅图像采用不同的密钥或不同的标识．这种想法对于一般的应用是可行的．但是应用于数字相机时则显然不实际．由于数字相机中的资源有限，需要用有限的密钥和标识，通常只用一个，来实现安全的水印算法．

在 Fridrich 等提出的易损水印方案中 [6]，首先通过相机密钥 K 产生块加密函数 E_K 的密钥．对每个像素 $g_{i,j}$，取像素块 $g_{i-u,j-v}$，$0\le u,\ v\le a-1$，其中 a 为整数 $(a=5)$．当嵌入标识比特 $L_{i,j}$ 时，修改像素 $g_{i,j}$ 使其满足

$$L_{i,j}=\mathrm{Parity}(E_K(g_1,\cdots,g_{a\times a})) \tag{2}$$

(2) 式右边表示对块中像素随机排序后加密函数的二值输出．在这一方案中，攻击者修改加水印图像的任一像素都必须考虑到对周围像素的影响．因此作者称从计算量上看这种易损水印方案是安全的．这种方案的缺点是，每一幅加水印图像相应位置的嵌入信息仍然相同．因此在本质上仍然为攻击者提供了机会；而且当某个像素被修改后，检测器会把周围的像素也看作被修改的区域，局部检测性能较 Yeung-Mintzer 的方法下降．

本文提出的易损水印方案可以有效地抵制以上两种攻击，而且可以公开标识．其根本思想是，为图像中嵌入的比特引入基于图像特征的不确定性．这通过产生一个基于图像特征的加密序列来实现．首先选取基于该图像的某些特征 $\{x_n,n=1,\cdots,N\}$，利用这些特征及相机密钥 K 合成混沌序列发生器的初始参数密钥 K'．对于混沌系统而言，K' 的微小变化都可能导致产生完全不同的一组序列．如果特征 $\{x_n,n=1,\cdots,N\}$ 随图像而变化，则可以得到不同的加密矩阵 $\{B_{i,j}\}$．攻击者将无法确定像素 (i,j) 处嵌入的比特，第一种攻击失败．由于不同加水印图像在相同位置嵌入的比特也可能不相同，第二种攻击失败．

2.2 图像特征的选取

本文易损水印算法的关键在于图像特征的选取．选取图像特征时应考虑到两个问题：

(1) 唯一性 对任意两幅图像，要求所产生的混沌加密序列不相关，从而保证水印嵌入对图像的依赖性。通常我们选取的图像特征主要是边缘，纹理，柱状图，块均值或特殊的像素点集等。在实际中，由于自然图像的差异以及自然环境的变化，数字相机摄得的两幅图像完全相同的概率非常小。因此一般的图像特征都能满足这一要求。

(2) 容错性 图像特征必须对局部修改具有一定的容错性。由于对加水印图像的局部修改可能导致图像特征部分遭破坏，致使产生不同的混沌密钥 K'。而密钥 K' 的微弱差别会产生完全不同的加密序列，使得水印检测器的输出结果是该图像完全遭破坏。这导致易损水印的局部修改检测功能失效。

本文提出一种简单的图像特征选取方法，利用一个像素集合 $\{x_n, n=1,\cdots,N\}$ 作为图像特征。集合中像素位置固定。在水印嵌入过程中保持该集合中的像素灰度值不变。其中像素位置的分布有两种：

(1) 像素集合 $\{x_n, n=1,\cdots,N\}$ 随机分布于整幅图像中。以此作为图像特征基本上可以反映该图像的内容特征，满足唯一性要求。当 N 增大时，依赖于图像的程度增强， K' 的不确定性也增加。但是随着 N 的增大，集合中的像素被破坏的几率增大，导致问题 (2) 的产生。

(2) 像素集合 $\{x_n, n=1,\cdots,N\}$ 集中于图像的某一局部块 (块内像素个数为 N)。对于 256 级灰度图像，该局部块共有 256^N 种。而在实际中任一图像的相应局部块中，其像素排列和灰度值完全相同的几率非常小，因此可以基本上满足唯一性要求。仅当该局部块遭修改时，才会导致易损水印的局部修改检测功能失效。

一般来说，攻击者主要是修改图像的某一部分. 在这一意义上，以随机分布的像素点作为图像特征更容易遭到破坏。因此本文采用后者，即利用局部像素块作为图像特征。同时该图像块应处于图像的重要内容处，例如图像中央，使得该局部块遭破坏时，图像的主要内容已遭破坏. 此时局部修改的检测已无必要. 从另一方面讲，攻击者也不敢修改图像特征区域，否则其攻击很容易使该图像被看作无效。 N 的合理选取也非常重要。较大的 N 意味着较高的安全性，但同时也意味着以降低局部检测性能为代价。因此必须在二者时间采取一个折衷。本文取图像中央的一个 8×8 的块，即 N=64 。

2.3 易损水印算法

本文提出的易损水印算法具体描述如下：

(1) 从图像中选取特征 $\{x_n, n=1,\cdots,N\}$ 。

(2) 利用相机密钥 K 及特征 $\{x_n, n=1,\cdots,N\}$ 合成混沌序列发生器的密钥 K' 。

$$f_1 : f_1(x_1, x_2, \cdots, x_N, K) \to K' \tag{3}$$

其中 f_1 是映射函数。

(3) 由 K' 及混沌序列发生器得到一个二值序列，重新排列成二值加密矩阵 $\{B_{i,j}\}$

(4) 设像素 $g_{i,j}$ 需嵌入标识比特 $L_{i,j}$，则修改 $g_{i,j}$ 使得

$$L_{i,j} = f(g'_{i,j}) \oplus B_{i,j} \tag{4}$$

其中 $g'_{i,j}$ 是修改后的像素， $\oplus$ 表示异或操作。对像素的修改应使 $g_{i,j}$ 与 $g'_{i,j}$ 距离最小。

(5) 保持图像特征 $\{x_n, n=1,\cdots,N\}$ 不变，完成所有像素的修改。

(6) 采用误差扩散算法提高加水印图像的感知质量 [3] 。

本文生成的查询表随机地把灰度值 0 到 255 映射为 0 或 1，并进行调整使得 3 个连续的灰度值不都映射为 0 或 1 。因此对像素的改动最多为 1，步骤 (6) 可以略去。

2.4 水印提取和图像认证

提取水印必须具有相机密钥 K .由于水印嵌入过程中并没有改动图像特征 $\{x_n, n=1,\cdots,N\}$, 因此可以按照上一节中相同的方法产生加密矩阵 $\{B_{i,j}\}$. 根据 $\{B_{i,j}\}$ 和映射函数 f 即可提取嵌入的标识. 提取过程逐个像素地进行. 通过主观观察提取的标识就可以看出图像所受到的修改. 也可将提取的标识与原始标识进行客观比较, 如存在不同的比特, 则在图像被修改的位置用灰度值 255 来表示, 即一个白点. 根据白点的多少和位置就可以判断该图像被修改的程度及局部位置.

3 安全性分析

算法安全性的前提是相机密钥 K 只能由安全部门和认证部门存取, 而且必须限制攻击者对检测器的访问. 目的是保证攻击者不可能获得有关相机密钥的信息, 而且不能通过反复测试的方法进行加水印图像的修改和伪造. 由于图像特征的唯一性, 以及混沌密钥 K' 的合成必须结合密钥 K, 图像特征的选取可以公开.

在提出的算法中, 决定算法安全性的两个重要因素分别是标识和映射函数. 相应的攻击可以分为 2 种情形: (1) 标识公开; (2) 映射函数 f 公开.

已知原始像素 $g_{i,j}$ 在嵌入水印后变为 $g'_{i,j}$. 对于情形 (1), 即标识 $L_{i,j}$ 公开, 攻击者可以得到的信息是

$$f(g'_{i,j}) \oplus B_{i,j} = L_{i,j} \tag{5}$$

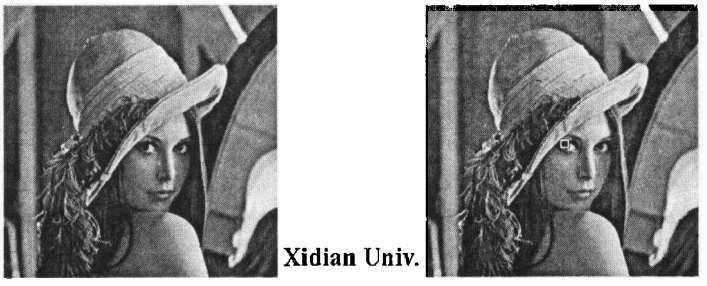
由于 $B_{i,j}$ 引入的不确定性, 攻击者只能进行随机的修改. 攻击成功的概率为 1/2 . 要想成功的伪造 M 个加水印像素, 成功的概率下降为 $(1/2)^M$. 显然 M 越大, 成功的概率越小. 攻击者可能会通过收集大量的加水印图像来估计映射函数 f . 由于每幅图像的 $B_{i,j}$ 对于攻击者来说都是随机的, 要想完全正确的估计映射函数 f 是不可能的. 但是对于情形 (2), 如果 $f(g'_{i,j})$ 已知, 尽管 $L_{i,j}$ 未知, 攻击者可以修改像素值为 $g''_{i,j}$, 使得 $f(g''_{i,j}) = f(g'_{i,j})$ 即可伪造任意的加水印图像. 因此应禁止情况 (2) 的发生.

4 实验结果

实验中使用的 256×256×8 的原始 Lena 图像和原始标识如图 1(a) 和图 1(b) 所示. 加水印图像如图 1(c) 所示, 其中白色正方形内的 8×8 方块是算法中的图像特征区域, 在加水印过程中保持不变. 可见加水印图像具有很高的感知质量, 其峰值信噪比 (PSNR) 约为 51.0dB .

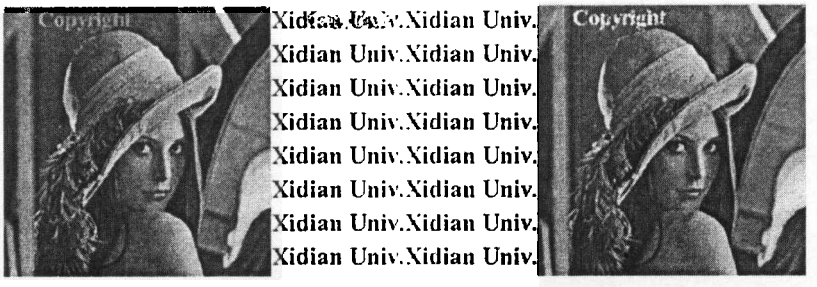
实验中对加水印图像的攻击可分为两种: (A) 修改位置与作为图像特征的局部块完全不重合; (B) 修改位置与作为图像特征的局部块部分重合. (A) 类型攻击如图 2(a) 所示, 其中 " Copyright " 的灰度值取自附近的像素灰度, 由图 2(a) 提取的标识以及图像修改的检测结果分别如图 2(b) 和图 2(c) 所示. 通过主观观察或与原始标识相比较均可看出, " Copyright " 所在区域被正确地检测到并定位.

当出现 (B) 类型攻击时, 根据算法可知, 这一攻击相当于修改了密钥 K' . 由于密钥不匹配, 检测器的检测结果将显示修改过的加水印图像已不具有可靠性. 这里, 我们把图 1(c) 中图像特征区域的某一个像素灰度值加 1, 修改结果如图 3(a) 所示. 检测结果如图 3(b) 和图 3(c) 所示. 可见提取的标识相当于随机噪声. 图 3(c) 中的白点即检测器输出的被修改位置, 类似于随机噪声, 这与以上的分析一致.



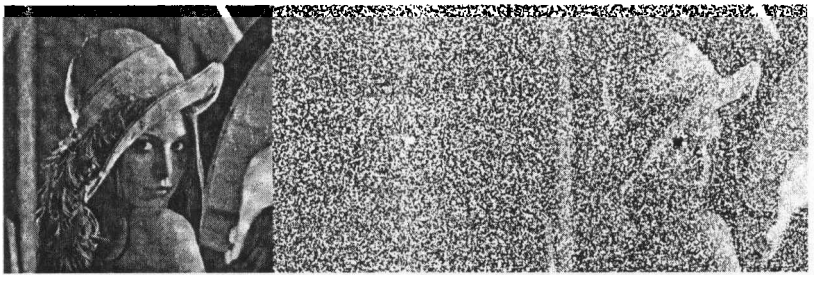
(a) 原始 Lena 图像 (b) 原始标识 (c) 加水印图像 (PSNR=51.0dB)

图 1



(a) A 类型攻击修改的加水印图像 (b) 提取的标识 (c) 图像中的修改位置

图 2



(a) B 类型攻击修改的加水印图像 (b) 提取的标识 (c) 图像中的修改位置

图 3

5 结 论

本文提出了一种新颖的空域易损水印技术。算法使用基于图像特征的混沌密钥，有效地解决了 Yeung-Mintzer 方法中存在的安全问题，并且保持了其良好的局部修改检测性能。实验结果表明该算法是有效的，具有简单，可行性好的特点。

参 考 文 献

[1] G. L. Friedman, The trustworthy digital camera: Restoring credibility to the photographic image, IEEE Trans. on Consumer Electronics, 1993, 39(4), 905–910.

[2] R. G. Schyndel, A. Z. Tirkel, C. F. Osborne, A digital watermark, In: Proc ICIP, Austin, Texas, 1994, vol.2, 86–90.

[3] M. Yeung, F. Mintzer, An invisible watermarking technique for image verification, In: Proc ICIP, Santa Barbara, California, 1997, 680–683.

[4] M. Holliman, N. Memon, Counterfeiting attacks for block-wise independent watermarking techniques, IEEE Trans. on Image Processing, 2000, 9(3), 432–441.

[5] J. Fridrich, M. Goljan, N. Memon, Further attacks on Yeung-Mintzer watermarking scheme, In: Proc. SPIE Photonic West, Electronic Imaging 2000, Security and Watermarking of Multimedia Contents, San Jose, California, January 24–26, 2000, 428–437.

[6] J. Fridrich, M. Goljan, A. C. Baldoza, New fragile authentication watermark for images, In: Proc ICIP, Vancouver, Canada, September 10-13, 2000, 446–449.

A NOVEL FRAGILE WATERMARK TECHNIQUE BASED ON IMAGE FEATURE

Zhong Hua　Liu Fang　Jiao Licheng

(*National Key Lab. for Radar Signal Processing, Xidian University, Xi'an* 710071, *China*)

Abstract In this paper, a novel fragile watermark technique is proposed, which is based on the method of Yeung-Mintzer's (1997). To resolve the existing security problem, a chaotic key which is based on image feature is used to control the embedding of fragile watermark. Theoretical analysis and experimental results demonstrate its effectiveness in resisting possible attacks. And because the watermark is embedded in pixel-wise order, the technique has good ability of detecting local alterations.

Key words Fragile watermarking, Image authentication, Image feature, Chaotic key, Attack

钟　桦：　男，1976 年生，博士生，主要研究领域为智能信息处理，信息隐藏，数字水印.

刘　芳：　女，1963 年生，副教授，主要研究领域为智能信息处理，模式识别，电子商务.

焦李成：　男，1959 年生，教授，博士生导师，主要研究领域为智能信息处理，非线性理论，数字水印.