抗拼贴攻击的二值图像认证算法

秦 瑶 [1] 和红杰 [1] 陈 帆 [1]

摘 要 为提高算法抵抗拼贴攻击的能力, 本文结合块相关思想提出一种二值图像认证算法. 该算法首先将每个图像块中的像素点分为 "可翻转" 和 "不可翻转" 两类, 将 "不可翻转" 像素点 Hash 生成水印信息, 用其替代相应映射块的 "可翻转" 像素点实现水印嵌入. 通过比较图像块重构水印及提取水印的一致性判断该图像块的真实性, 并结合各图像块邻域篡改特征进一步提高篡改检测性能. 实验结果表明, 该算法不仅提高了在替换、添加、删除攻击下的篡改检测性能, 而且能够有效抵抗拼贴攻击.

关键词 二值图像, 拼贴攻击, 块相关, 邻域篡改特征

引用格式 秦瑶, 和红杰, 陈帆. 抗拼贴攻击的二值图像认证算法. 自动化学报, 2014, **40**(6): 1184−1190

DOI 10.3724/SP.J.1004.2014.01184

Binary Image Authentication Watermarking Scheme Against Collage Attack

QIN Yao[1] HE Hong-Jie[1] CHEN Fan[1]

Abstract To improve the ability against collage attack, a watermarking scheme for binary image authentication is proposed by introducing dependence of blocks in this paper. In the proposed algorithm, all pixels in each block are divided into the flipped pixels and the unflipped ones. The watermark information of each block is generated by hashing the unflipped pixels of it, and embedded in the flipped pixels of the corresponding mapping block. The authenticity of each block is detected by comparing the consistency between the reconstructed watermark and the extracted one. The neighborhood tampering feature is used to improve the tamper detection performance. Experimental results show that the proposed algorithm can not only improve the performance of tamper detection in the general tampering such as replace, add, delete, etc., but also effectively resist collage attack.

Key words Binary image, collage attack, block correlation, neighborhood tampering feature

Citation Qin Yao, He Hong-Jie, Chen Fan. Binary image authentication watermarking scheme against collage attack. *Acta Automatica Sinica*, 2014, **40**(6): 1184−1190

随着数字成像技术的发展, 以及互联网的普及与应用, 使图像成为信息传播的重要载体. 同时图像处理与编辑软件使非专业人士也能对图像进行各种编辑与处理. 如果篡改或伪造的图像用于一些特殊领域会带来不良的社会影响, 如 "藏羚羊"、"广场鸽" 等新闻造假事件. 数字图像盲取证[1] 作为图像真伪和来源鉴别的技术之一, 近几年成为研究的热点. 研究者提出了针对复制粘贴、JPEG 压缩等篡改方式的盲取证算法[2−4]. 而反取证技术[5] 的发展对盲取证技术提出了更高的要求. 另一方面, 基于数字水印的图像主动取证技术[6] 通过事先在图像中嵌入水印信息来达到保护图像真实性、完整性等目的. 目前, 数字图像主动取证水印技术主要以灰度图像和彩色图像为研究对象[7−9], 二值图像具有占用空间少的突出优点, 在实际应用中, 如文档、支票、合同、证明、通知书以及办公自动化中的电子信函、公文、传真等文档, 常常以二值图像的方式保存, 因此, 对二值图像主动取证水印技术的研究, 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.

近年来许多学者对二值图像认证水印进行了大量研究, Tzeng 等[10] 将通过密钥产生与图像内容无关的水印信息随机嵌在图像中, 造成图像较为明显的失真. 这是由于二值图像中每个像素都具有特定含义, 若对其做任何不当改动都易被察觉, 因此选择修改哪些像素进行信息嵌入, 也就是说如何选择 "可翻转点" 尤为重要. Yang 等[11] 和 Xi 等[12] 分别通过考察翻转像素所引起的图像块连通性变化、光滑

收稿日期 2013-01-25 录用日期 2013-05-29
Manuscript received January 25, 2013; accepted May 29, 2013
国家自然科学基金 (61373180, 60970122), 中央高校基本科研业务专项基金资助 (SWJTU09CX039, SWJTU10CX09), 西藏自治区科技计划项目资助
Supported by National Natural Science Foundation of China (61373180, 60970122), the Central College Basic Scientific Research Business Special Fund (SWJTU09CX039, SWJTU10CX09), and Science and Technology Plan Projects of Tibet Autonomous Region
本文责任编委 周杰
Recommended by Associate Editor ZHOU Jie
1. 西南交通大学信息科学与技术学院 成都 610031
1. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031

性变化来选择可翻转点, Kwon 等[13] 对图像进行小波变换在 HH (High-High) 带选择可翻转系数, 这些方法虽实现了水印信息不可见, 但可嵌入信息量较少. 为提高水印嵌入容量, Wu 等[14]、Zhu 等[15] 和 Li 等[16] 分别通过设定阈值考察翻转中心像素所引起的图像块连通性和平滑性的变化、边界数的变化、像素扩展值的变化来选择对视觉影响较小的可翻转点作为水印嵌入位置.

另一方面, 检测并定位篡改区域是评价认证水印算法的重要指标. Wu 将图像置乱后实现分块信息嵌入, 混洗表的引入使算法无法定位篡改区域. Zhu 在每个图像块中嵌入 1 bit 水印信息, 由于容量较小导致算法篡改检测效果不理想. 以上算法完全基于块独立思想, 不能抵抗拼贴攻击[17], 存在安全隐患. 为提高抗拼贴攻击能力, Yang 将图像块分为两类, 其中合格子块的水印信息是通过自身及邻域生成的. 算法在拼贴图像块或其邻域块类别改变时, 能够检测到篡改但会造成邻域虚检, 若类别不变也只能检测到拼贴区域的边缘. Li 基于等级结构划分图像子块, 各等级子块进行独立的水印生成与嵌入. 算法低等级子块定位精度较高, 漏检率也高; 高等级子块漏检率较低, 但定位精度也随之下降. 也就是说, 若某等级子块小于拼贴篡改区域, 那该等级将无法抵抗拼贴攻击, 这时就必须依靠更高等级才能对篡改进行认证, 而这是以低定位精度为代价的. Kwon 等提出一种自适应改变图像块大小的水印算法, 这能改善因水印嵌入位置不足导致篡改检测能力下降的情况, 但分块不固定会造成定位精度较差, 同时算法只有在拼贴区域分块改变时, 才能检测到篡改, 即无法完全抵抗拼贴攻击.

综上, 针对现有算法不能有效抵抗拼贴攻击的问题, 本文提出一种能够抵抗拼贴攻击的二值图像认证算法. 算法将图像不重叠分块, 把基于图像块自身生成的水印信息嵌入到其他图像块来抵抗拼贴攻击. 通过比较重构及提取水印的一致性, 并结合邻域篡改特征判定图像块的真实性. 本文提出的算法在保证水印不可见性的前提下, 能有效抵抗拼贴攻击.

1　算法描述

本文首先通过实验对比分析现有可翻转点选择方法对水印容量和不可见性的影响, 然后结合可翻转点和图像块内容完成水印生成与嵌入, 最终实现水印提取和图像认证.

1.1　可翻转点的选择

与彩色图像或灰度图像相比, 二值图像视觉冗余信息较小, 若对像素做任何不当改动都易引起明显的修改痕迹, 因此在修改像素嵌入信息时必须考虑该像素的邻域. 对于二值图像认证算法中的可翻转点, 需满足两个必要条件:

1) 翻转可翻转点能够保证视觉不可见性;

2) 可翻转点个数能保证水印容量满足篡改检测要求.

为度量 3×3 图像块中心像素点的 “可翻转性”, Wu 等[14] 建立了中心像素点可翻转性查找表, 通过设置 5 个阈值将像素点的 “可翻转性” 分为 6 个等级. Zhu 等[15] 通过计算翻转中心点所引起的边界数变化设定了 4 个阈值, 把像素点的 “可翻转性” 分为 5 个等级. Li 等[16] 通过考虑图像块内各个像素点的方向将像素点的 “可翻转性” 通过设定 12 个阈值分为 13 个等级. 本文将使用这三种方法衡量 100 幅大小为 384×384 的二值图像中像素点的 “可翻转性”, 从中选择一种能够在嵌入较多信息量时, 仅造成较小视觉影响的判别方法.

图 1 为三种方法通过选择不同阈值确定的可翻转点个数. 横坐标表示可选择的阈值个数, 纵坐标为可翻转点个数. 其中 Li 方法选取 10 个阈值进行实验. 图 2 是翻转三种方法选择的可翻转点时不可见性测试, 用结构相似性 (Structural similarity, SSIM)[18] 来衡量.

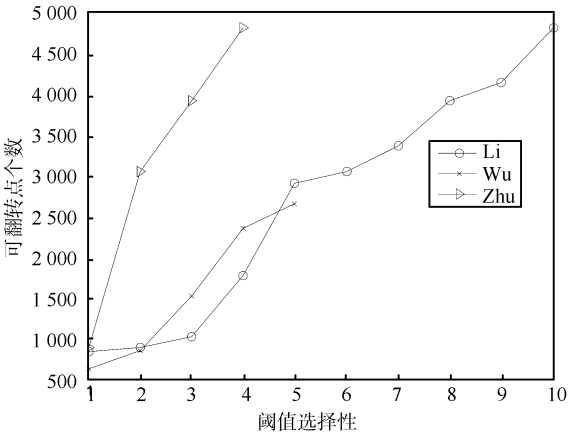


图 1　不同方法下的可翻转点个数

Fig. 1　Number of flipped pixels in different methods

从图 1 可以看出, Zhu 和 Li 选择的可翻转点个数范围都在 1 000 到 5 000 左右, 非常接近, 而 Wu 可选择的翻转点个数范围明显小于前两者. 从图 2 看出, 随着嵌入的信息量增多, 不可见性呈下降趋势. Zhu 和 Li 方法翻转像素点时 SSIM 值变化范围相近. 与此同时, Li 方法由于对像素点的可翻转性划分了更多的等级, 可翻转点个数和不可见性就具有了更多的选择, 这更满足实际应用的不同需求.

除此之外, Li 方法还可以考察更大图像块像素点的翻转性, 且不需要额外的存储空间, 故本文最终

使用 Li 方法来选择可翻转像素点.

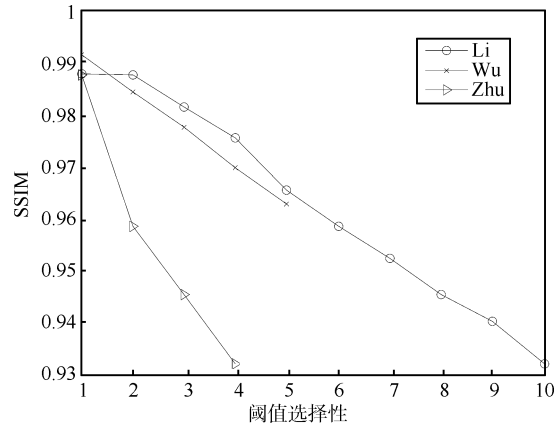


图 2 不同方法下的 SSIM

Fig. 2 SSIM in different methods

1.2 水印生成与嵌入

将图像块中 “不可翻转” 像素点 Hash 加密生成水印信息, 用其替代相应映射块的 “可翻转” 像素点完成水印嵌入, 水印生成与嵌入框图如图 3 所示.

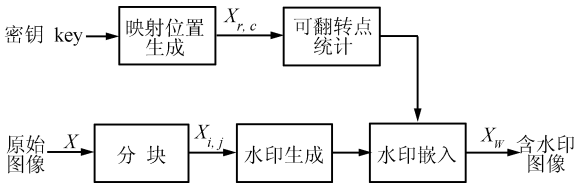


图 3 水印嵌入算法框图

Fig. 3 Schematic diagram of watermark embedding

步骤如下:

1) 分块: 将大小为 $m \times n$ 的原始图像 X 划分成大小为 $s_1 \times s_2$ 的图像块 $X_{i,j}$, 即 $X = \{X_{i,j}|i = 1,2,\cdots,h_1, j = 1,2,\cdots,h_2\}$, 其中 $h_1 = m/s_1$, $h_2 = n/s_2$, s_1 和 s_2 均为 3 的倍数, 即每个 $s_1 \times s_2$ 块中包含 $(s_1/3) \times (s_2/3)$ 个 3×3 块.

2) 映射位置生成: 基于混沌生成各图像块的映射位置[19]. 密钥 key $= \{k_1, k_2, k_3\}$, 其中, k_1, $k_2 \in (-1.5, 1.5)$, $k_3 \in (0, 1)$. 生成块相关标识矩阵 $E = \{E_{i,j}|i = 1,2,\cdots,h_1, j = 1,2,\cdots,h_2\}$, 设块 $X_{i,j}$ 的映射块为 $X_{r,c}$, 即基于块 $X_{i,j}$ 生成的水印信息嵌入到块 $X_{r,c}$, 则两图像块对应关系为

$$r = \begin{cases} \dfrac{E_{i,j}}{h_2}, & E_{i,j}|h_2 \\ \left[\dfrac{E_{i,j}}{h_2}\right] + 1, & \text{其他} \end{cases} \tag{1}$$

$$c = \begin{cases} h_2, & E_{i,j}|h_2 \\ E_{i,j} - h_2 \times \left[\dfrac{E_{i,j}}{h_2}\right], & \text{其他} \end{cases} \tag{2}$$

其中, $[\cdot]$ 为向下取整, $|$ 为整除. 分别用 A、B 表示图像块 $X_{i,j}$ 和 $X_{r,c}$.

3) 可翻转点统计: 把图像块 B 划分成不重叠的 3×3 块, 计算每块中心点的像素扩展差 (Pixel spread deviation, PSD)[16]. 图像块 B 中像素点记为 $B_{f,g}$, 其中, f, g 为各像素点在该块内坐标, $f = 1,2,\cdots,s_1$, $g = 1,2,\cdots,s_2$. 用标识矩阵 $P = \{P_{f,g}|f = 1,2,\cdots,s_1, g = 1,2,\cdots,s_2\}$ 记录图像块 B 中各像素点的可翻转性, 并统计该图像块中可翻转点的个数 L_B,

$$P_{f,g} = \begin{cases} 1, & \text{PSD}_{f,g} \le 1.5, (f+1)|3, (g+1)|3 \\ 0, & \text{其他} \end{cases} \tag{3}$$

其中, $P_{f,g} = 1$ 标识该像素点可翻转, $P_{f,g} = 0$ 标识该像素点不可翻转.

$$L_B = \sum_{f=1}^{s_1}\sum_{g=1}^{s_2} P_{f,g} \tag{4}$$

4) 水印生成 : 将图像块 A 中每个 3×3 块的中心点置 0 通过 Hash 生成长度为 L_B 的二值序列 $W1_A$, 中心点置 1 作为 Hash 函数的输入, 生成长度为 L_B 的二值序列 $W2_A$, 则基于图像块 A 生成的水印信息 W_A 为

$$W_A = W1_A \oplus W2_A = \{W_{A,k}|k = 1,2,\cdots,L_B\} \tag{5}$$

5) 水印嵌入 : 用图像块 A 生成的水印信息 W_A 代替图像块 B 中的可翻转点嵌入水印. 嵌入水印信息后图像块 B 中的各像素点需满足式 (6),

$$B_{f,g} = \begin{cases} W_{A,k}, & P_{f,g} = 1 \\ B_{f,g}, & P_{f,g} = 0 \end{cases} \tag{6}$$

其中, $k = 1,2,\cdots,L_B$. 依照光栅扫描顺序把所有图像块嵌入水印, 最终得到含水印图像 X_W.

1.3 水印提取与图像认证

算法实现了水印信息的盲提取, 具体步骤叙述如下:

1) 分块: 把待检测图像 Y 分成大小 $s_1 \times s_2$ 为的图像块 $Y_{i,j}$. 利用密钥 key* 生成块相关标识矩阵 E^*, 设由 E^* 判定图像块 $Y_{i,j}$ 的水印信息嵌入在图像块 $Y_{r,c}$ 中, 分别用 A^*、B^* 代表图像块 $Y_{i,j}$ 和 $Y_{r,c}$.

2) 水印的重构与提取: 通过同样的水印生成方法生成图像块 A^* 的信息序列, 即重构的水印信息 W_{A^*}. 图像块 B^* 中所有可翻转像素点的像素值对

应图像块 A^* 提取的水印信息 W'_{A^*}.

$$W_{A^*} = \{W_{A^*,k} | k = 1, 2, \cdots, L_{B^*}\} \tag{7}$$

$$W'_{A^*} = \{W'_{A^*,k} | k = 1, 2, \cdots, L_{B^*}\} \tag{8}$$

其中, L_{B^*} 为图像块 B^* 中可翻转像素点的个数.

3) 结合邻域篡改特征进行篡改检测定位:

步骤 1. 判断图像块 A^* 重构水印与提取水印的一致性,

$$D_{A^*} = \begin{cases} 1, & W_{A^*} \neq W'_{A^*} \\ 0, & W_{A^*} = W'_{A^*} \end{cases} \tag{9}$$

其中, $D_{A^*} = 1$ 表示图像块 A^* 重构水印与提取水印不相同, $D_{A^*} = 0$ 表示重构水印与提取水印相同.

步骤 2. 继续判断图像块 A^* 的真实性, $d8_{A^*}$、$d8_{B^*}$ 分别表示 D_{A^*} 和 D_{B^*} 八邻域中 1 的个数, 则

$$C_{A^*} = \begin{cases} 1, & D_{A^*} = 1, d8_{A^*} \geq d8_{B^*} \\ 0, & D_{A^*} = 1, d8_{A^*} < d8_{B^*} \end{cases} \tag{10}$$

步骤 3. H_{A^*} 是对块 A^* 真实性的最终判定, 若 H_{A^*}=1, 则块 A^* 真实; 若 $H_{A^*} = 0$, 则块 A^* 被篡改.

$$H_{A^*} = \begin{cases} 1, & C_{A^*} = 1, d8_{A^*} > T_1 \\ 0, & C_{A^*} = 1, d8_{A^*} \leq T_1 \\ 1, & C_{A^*} = 0, d8_{A^*} > T_2 \\ 0, & C_{A^*} = 0, d8_{A^*} \leq T_2 \end{cases} \tag{11}$$

依照如上步骤按光栅扫描顺序对所有图像块的真实性进行判断, 最终得到篡改定位图像.

2　仿真结果分析

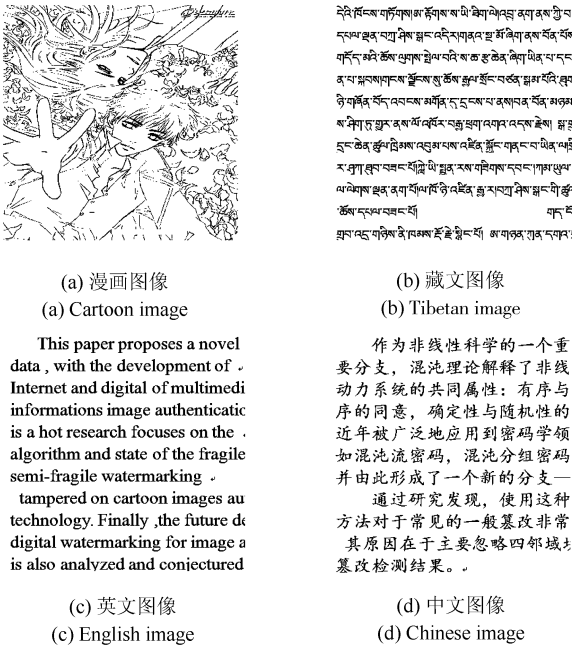
本文利用一定篡改比例 ρ_T 下的漏检概率 P_{fa} 和虚检概率 P_{fr} 定量恒量算法的篡改检测性能[20],

$$\rho_T = \frac{N_T}{N} \tag{12}$$

$$P_{fa} = \frac{N_T - N_{TD}}{N_T} \tag{13}$$

$$P_{fr} = \frac{N_{VD}}{N - N_T} \tag{14}$$

其中, N 为图像块个数, N_T 为篡改图像块个数, N_{TD} 为检测出的篡改图像块个数, N_{VD} 为被误判为篡改的真实图像块个数. 为体现本文算法对二值图像的通用性, 分别选取如图 4 所示的漫画、藏文、英文和中文 4 类二值图像为测试图像.



(a) 漫画图像
(a) Cartoon image

(b) 藏文图像
(b) Tibetan image

(c) 英文图像
(c) English image

(d) 中文图像
(d) Chinese image

图 4　原始二值图像
Fig. 4　Original binary images

2.1　阈值选取

本文结合图像块的邻域篡改特征进行图像块真实性的判断, 从式 (11) 可以看出, 阈值 T_1、T_2 的选取对篡改检测有至关重要的影响, 本文采用实验统计法对阈值 T_1、T_2 进行探讨.

实验对以上 4 类各 25 幅大小为 384×384 二值图像在不同篡改比例下, 采用不同的阈值进行篡改检测. 图 5 分别给出 $\rho_T = 0.0469$, $\rho_T = 0.1172$, $\rho_T = 0.2188$ 时, 不同阈值下算法的漏/虚检概率 P_{fa}/P_{fr} 曲线. 其中横坐标为阈值 T_2, 纵坐标为漏/虚检概率. 可以看出, ρ_T 不同阈值相同时检测结果会有所不同. 当一个阈值不变, 另一阈值增大时, 漏检率上升, 虚检率下降. 考虑到攻击者不会大肆修改图像内容, 故结合图像块的邻域篡改特征, 兼顾虚检概率和漏检概率, 实验中采用阈值 $T_1 = 1$, $T_2 = 3$.

2.2　篡改检测性能

为验证算法在一般篡改、拼贴攻击和混合篡改等不同伪造攻击下的检测性能, 以图 4 中的测试图像为对象, 对比给出本文、Kwon[13] 算法和 Li[16] 算法在各种篡改下的检测结果. 其中, 本文算法分块大小为 24×24, Kwon 算法最小分块为 24×24, Li 算法将图像分为 5 个等级, 最低级子块大小为 24×24. 为公平比较, 各检测率均以 24×24 的图像块为单位计算. 实验中取本文的密钥 $k_1 = 0.4$, $k_2 = 0.5$, $k_3 = 0.6$.

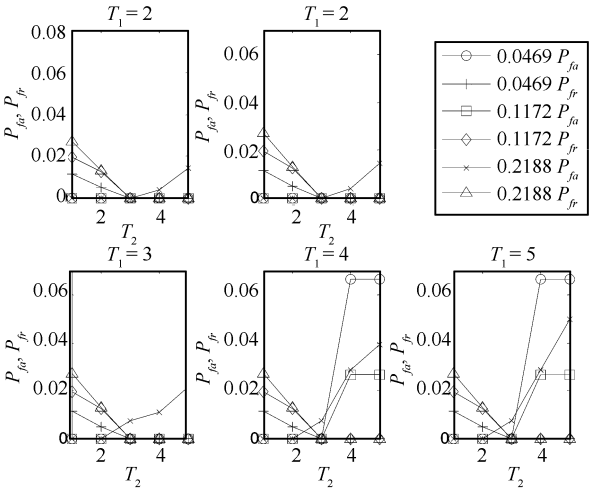


图 5 不同 ρ_T 的漏检率和虚检率曲线图

Fig. 5 P_{fa} and P_{fr} graph in different ρ_T

2.2.1 一般篡改

以图 4 (a) 漫画图像为测试图像, 利用本文算法生成的含水印图像见图 6 (a), 水印容量 2 436 bits, SSIM = 0.9595. 用 “这个世界我主宰” 矩形框替换含水印图像的右下部分如图 6 (b) 所示, 篡改比例为 9.38 %.

图 6 (c) ∼ (f) 分别是本文、Li 算法第 5 级、第 4 级和 Kwon 算法的认证结果, 相应的虚、漏检概率见表 1 的 “一般篡改” 行. Li 算法第 5 级有较多篡改块没有检测到, 漏检率为 12.50 %. 第 4 级依然存在漏检, 并且由于图像块的增大使虚检率增大到 2.59 %. Kwon 算法的虚检率达到 6.90 %. 而本文结合块邻域篡改特征进行检测漏检率为 4.17 %, 没有虚检出现, 表明本文算法对一般篡改拥有较好的认证能力.

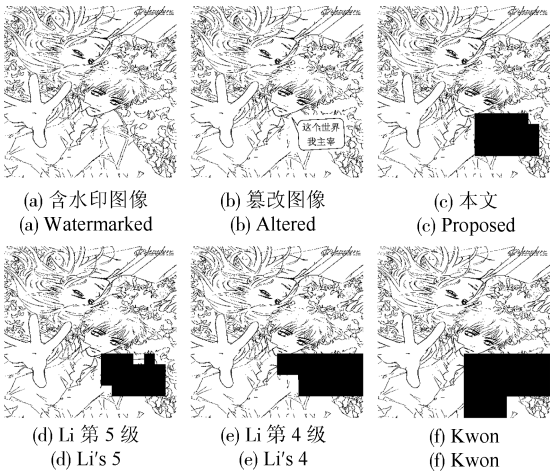


图 6 一般篡改认证结果

Fig. 6 Authentication results in the general tampering

2.2.2 拼贴攻击

为测试算法抵抗拼贴攻击的能力, 分别以藏文和英文二值图像为测试图像, 测试在单区域和多区域拼贴攻击时算法的检测性能.

图 7 (a) 是利用本文算法生成的含水印图像, 水印容量为 3 598 bits, SSIM = 0.9503. 利用相同密钥生成图 4 (b) 的含水印图像, 将其 (100:200, 100:200) 的区域拼贴到图 7 (a) 的相应区域, 得到单区域拼贴图像如图 7 (b) 所示, 拼贴比例为 9.77 %.

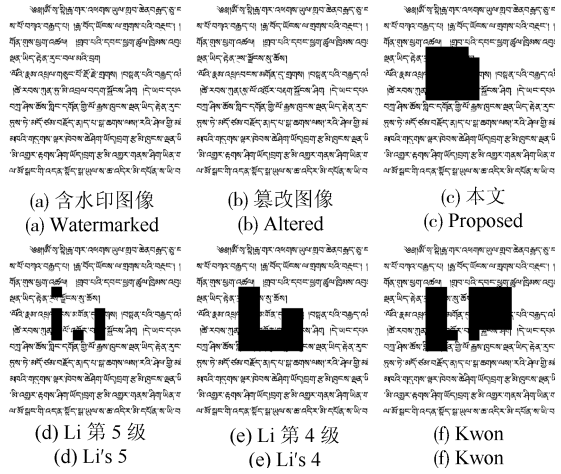


图 7 单区域拼贴认证结果

Fig. 7 Authentication results in single area collage attack

图 7 (c) ∼ (f) 分别为三个算法的认证定位结果. 相应的虚、漏检概率见表 1 的 “单区拼贴” 行. 可以看到, Li 算法第 5 级认证能力较差, 漏检率达到 68 %, 第 4 级漏检情况有所改善, 但带来了 3.9 % 的虚检. Kwon 算法同样没有检测到完全拼贴块, 分块的不固定使虚检率达到 6.06 %. 而本文只漏检了一个篡改块, 没有虚检, 表明本文算法能较好地抵抗单区域拼贴攻击.

图 8 (a) 是利用本文算法生成的含水印图像, 水印容量为 3 050 bits, SSIM = 0.9573. 利用相同密钥生成图 4 (c) 的含水印图像, 用其对图 8 (a) 进行多区域拼贴篡改, 包括: 1) 第 4 行开始处的 “documents” 拼贴为 “informations”; 2) 倒数第 4 行的 “binary” 拼贴为 “cartoon”. 得到多区域拼贴图像如图 8 (b) 所示, 拼贴比例为 8.59 %.

图 8 展示了三个算法的定位结果. 相应的虚、漏检概率见表 1 的 “多区拼贴” 行. 可以看到, 当拼贴区域较大时, Li 算法较低等级子块的认证能力较差, 第 5 级的漏检高达 90.91 %, 第 4 级漏检率只下降了少许, 没有得到明显改善. 由于篡改图像和原始图像的分块情况相同, 故 Kwon 算法也没有检测到拼贴篡改, 漏检了 77.27 % 的篡改块. 从图 8 (c) 及表 1 可以清楚看到本文算法对多区域拼贴篡改认证的有

效性.

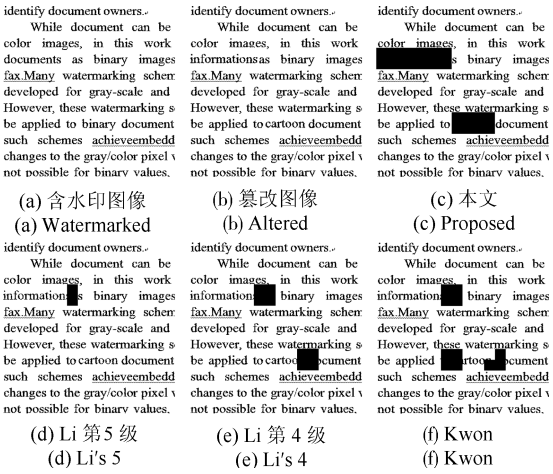


图 8　多区域拼贴认证结果

Fig. 8　Authentication results in multiple areas collage attacks

从以上两组拼贴篡改实验可以看出, Li 算法第 5 级只能检测到一些边缘篡改, 漏检率很高, 而第 4 级图像块的增大使漏检情况得到改善, 但若篡改区域大于该等级子块时, 依然无法抵抗拼贴攻击. Kwon 算法对于分块情况未改变的图像无法检测到拼贴篡改, 且存在较大的虚检. 而本文通过在图像块之间引入相关性对单区域和多区域拼贴都有较好地检测性能.

2.2.3　混合攻击

当添加、删除、替换和拼贴篡改同时存在于一幅图像中时, 本文算法依然具有优越的篡改定位能力. 图 9 (a) 是含水印图像, 水印容量为 3 180 bits, SSIM = 0.9535. 同时, 利用相同密钥生成图 4 (d) 的含水印图像以实施拼贴攻击. 对图 9 (a) 进行以下篡改: 1) 在第 1 行开始处添加 “如”; 2) 在第 4 行 “独立” 篡改为 “相关”; 3) 将倒数第 4 行的 “虚检” 删除; 4) 用图 4 (d) 的含水印图像右下角包含 “一般篡改”、“忽略四” 的区域替换图 9 (a) 相应区域的 “拼贴攻击”、“考虑八”. 得到包含 4 处篡改的混合攻击图像如图 9 (b) 所示, 篡改比例为 12.50 %.

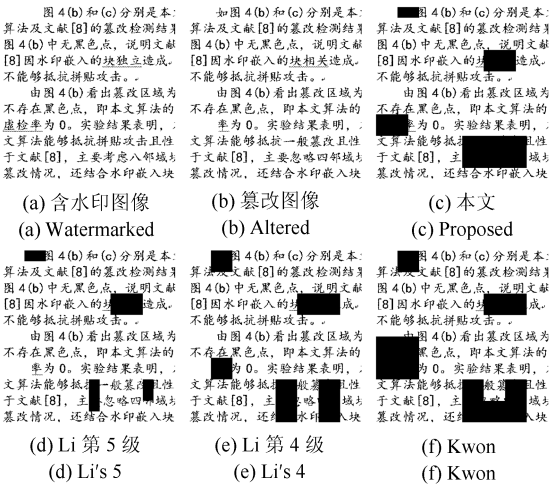


图 9　混合攻击认证结果

Fig. 9　Authentication results in blended attacks

三个算法认证结果分别见图 9 (c) ∼ (f), 相应的虚、漏检概率见表 1 “混合攻击” 行. 可以看出: Li 算法第 5 级由于水印容量分配不足与 “块独立” 导致漏检 59.38 % 的篡改块, 第 4 级漏检率下降到 31.25 %, 同时虚检了 4.46 % 的真实块. Kwon 算法也没有检测到图像块的拼贴篡改. 相比而言, 本文算法的检测结果没有虚漏检情况出现, 即能够较好地检测到不同篡改, 具有优越的认证能力.

综合以上实验可看出: 对于一般替换、添加、删除篡改以及拼贴攻击, 与现有算法相比, 本文算法能够在保证定位精度的前提下有效地对篡改进行认证定位.

3　结语

本文重点针对现有二值图像认证算法不能抵抗拼贴攻击的缺陷, 提出了一种抗拼贴攻击的二值图像认证算法. 该算法将原始图像分块, 对于每个图像块, “不可翻转” 像素点 Hash 加密生成水印信息. 为抵抗拼贴攻击, 引入图像块之间的相关性, 将水印信息置乱嵌入到其他图像块中. 篡改认证时通过结合邻域篡改特征提高算法的篡改检测能力. 实验结果表明, 本文算法在保证对替换、添加、删除攻击等具备良好的篡改检测性能的前提下, 能够有效抵抗拼贴攻击. 如何结合实际应用扩大水印容量, 在算法不

表 1　现有算法认证能力比较

Table 1　Comparison of authentication performance with existing algorithms

篡改方式	P_{fa} (%)				P_{fr} (%)			
	Proposed	Li's 5	Li's 4	Kwon	Proposed	Li's 5	Li's 4	Kwon
一般篡改	4.17	12.50	8.33	0.00	0.00	0.00	2.59	6.90
单区拼贴	4.00	68.00	40.00	36.00	0.00	0.00	3.90	6.06
多区拼贴	0.00	90.91	81.82	77.27	0.00	0.00	1.71	2.56
混合攻击	0.00	59.38	31.25	12.50	0.00	0.00	4.46	8.93

可见性与篡改认证能力间达到更好的平衡是下一步研究重点.

References

1 Zhou Lin-Na. Study of Digital Forensics Based on Image Content [Ph. D. dissertation], Beijing University of Posts and Telecommunications, China, 2007
(周琳娜. 数字图像盲取证技术研究 [博士学位论文], 北京邮电大学, 中国, 2007)

2 Huang F J, Huang J W, Shi Y Q. Detecting double JPEG compression with the same quantization matrix. *IEEE Transactions on Information Forensics and Security*, 2010, **5**(4): 848−856

3 Peng F, Nie Y Y, Long M. A complete passive blind image copy-move forensics scheme based on compound statistics features. *Forensic Science International*, 2011, **212**(1−3): e21−e25

4 Mahalakshmi S D, Vijayalakshmi K, Priyadharsini S. Digital image forgery detection and estimation by exploring basic image manipulations. *Digital Investigation*, 2012, **8**(3−4): 215−225

5 Stamm M C, Liu K J R. Anti-forensics of digital image compression. *IEEE Transactions on Information Forensics and Security*, 2011, **6**(3): 1050−1065

6 Haouzia A, Noumeir R. Methods for image authentication: a survey. *Multimedia Tools and Applications*, 2008, **39**(1): 1−46

7 Qi Ke, Xie Dong-Qing. Watermarking scheme against geometrical attacks based on second generation Bandelet. *Acta Automatica Sinica*, 2012, **38**(10): 1646−1653
(綦科, 谢冬青. 基于第二代 Bandelet 变换的抗几何攻击图像水印. 自动化学报, 2012, **38**(10): 1646−1653)

8 Aldeeb N H H, Abuhaiba I S I. An improved watermarking scheme for tiny tamper detection of color images. *International Journal of Image, Graphics and Signal Processing*, 2013, **5**(5): 1−14

9 He H J, Chen F, Tai H M, Kalker T, Zhang J S. Performance analysis of a block-neighborhood-based self-recovery fragile watermarking scheme. *IEEE Transactions on Information Forensics and Security*, 2012, **7**(1): 185−196

10 Tzeng C H, Tsai W H. A new approach to authentication of binary images for multimedia communication with distortion reduction and security enhancement. *IEEE Communications Letters*, 2003, **7**(9): 443−445

11 Yang H J, Alex C K. Binary image authentication with tampering localization by embedding cryptographic signature and block identifier. *IEEE Signal Processing Letters*, 2006, **13**(12): 741−744

12 Xi Y H, Zhang M R. A document watermarking algorithm based on partitioned character image. In: Proceedings of the 2009 International Conference on E-Business and Information System Security. Wuhan, China: IEEE, 2009. 1−4

13 Kwon O J, Kim W G, Kim Y. Binary image watermarking with tampering localization in binary wavelet transform domain. *International Journal of Multimedia and Ubiquitous Engineering*, 2012, **7**(2): 341−346

14 Wu M, Liu B D. Data hiding in binary image for authentication and annotation. *IEEE Transactions on Multimedia*, 2004, **6**(4): 528−538

15 Zhu Cong-Xu, Chen Zhi-Gang. Sensitive chaotic fragile watermarking technique for binary images verification. *Mini-Micro Systems*, 2006, **27**(1): 151−154
(朱从旭, 陈志刚. 一种灵敏的文本图像认证混沌脆弱水印技术. 小型微型计算机系统, 2006, **27**(1): 151−154)

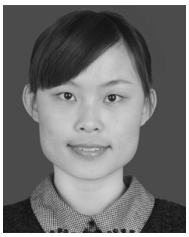
16 Li Zhao-Hong, Hou Jian-Jun, Song Wei. Binary document image authentication watermarking technique based on hierarchical structure. *Acta Automatica Sinica*, 2008, **8**(8): 841−848
(李赵红, 侯建军, 宋伟. 基于等级结构的二值文本图像认证水印算法. 自动化学报, 2008, **8**(8): 841−848)

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

18 Wang Z, Bovik A C, Sheikh H R, Simoncelli E P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, **13**(4): 600−612

19 He Hong-Jie, Zhang Jia-Shu. Chaos-based scramble self-embedding watermarking algorithm. *Journal on Communications*, 2006, **27**(7): 80−86, 93
(和红杰, 张家树. 基于混沌置乱的分块自嵌入水印算法. 通信学报, 2006, **27**(7): 80−86, 93)

20 He Hong-Jie. Digital Image Secure Authentication Watermarking Algorithms and Their Performance Analysis of Statistical Detection [Ph. D. dissertation]. Southwest Jiaotong University, China, 2009
(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, 中国, 2009)



秦 瑶 西南交通大学信息科学与技术学院硕士研究生. 主要研究方向为二值图像数字水印技术.
E-mail: yaoqin525@yeah.net
(**QIN Yao** Master student at the School of Information Science and Technology, Southwest Jiaotong University. Her main research interest is binary image digital watermarking.)

和红杰 西南交通大学信息科学与技术学院教授. 主要研究方向为数字图像处理, 信息隐藏.
E-mail: hjhe@swjtu.cn
(**HE Hong-Jie** Professor at the School of Information Science and Technology, Southwest Jiaotong University. Her research interest covers digital image processing and information hiding.)

陈 帆 西南交通大学信息科学与技术学院副教授. 主要研究方向为数字图像取证, 信息隐藏. 本文通信作者.
E-mail: mrchenfan@126.com
(**CHEN Fan** Associate professor at the School of Information Science and Technology, Southwest Jiaotong University. His research interest covers digital image forensics and information hiding. Corresponding author of this paper.)