Computer Science and Application 计算机科学与应用, 2017, 7(1), 67-73
Published Online January 2017 in Hans. http://www.hanspub.org/journal/csa
http://dx.doi.org/10.12677/csa.2017.71009

Hans汉斯

Fragile Watermarking Algorithm for the Text Image Based on Hash Function

Xiaoxue Ma, Ying Wang

Qingdao University, Qingdao Shandong
Email: 1182683153@qq.com

Received: Jan. 6th, 2017; accepted: Jan. 21st, 2017; published: Jan. 24th, 2017

Abstract

Aiming at the actual demand of authenticity and integrity authentication of text image, a fragile watermarking algorithm for text image based on Hash function is proposed. The idea of the algorithm is to first block the text image, and then calculate the reversible and non-reversible pixel points in each block based on smoothness and connectivity. The watermark information is generated by MD5 algorithm in Hash function. Finally, the watermark information is embedded into the reversible pixel points. The watermark information extracted in the received text image is compared with the previously generated watermark information, and the authenticity and integrity of the text image can be authenticated. The algorithm has a good ability of tampering positioning.

Keywords

Text Image, Hash Function, Fragile Watermarking, Tampering Positioning

基于Hash函数的文本图像脆弱水印算法

马小雪，王　英

青岛大学，山东 青岛
Email: 1182683153@qq.com

收稿日期：2017年1月6日；录用日期：2017年1月21日；发布日期：2017年1月24日

摘　要

针对文本图像内容真实性和完整性认证的实际需求，提出一种基于Hash函数的文本图像脆弱水印算法。算法思想是首先对文本图像进行分块，然后根据平滑度和连通性来计算每个分块中的可翻转像素点和不

文章引用: 马小雪, 王英. 基于 Hash 函数的文本图像脆弱水印算法[J]. 计算机科学与应用, 2017, 7(1): 67-73.
http://dx.doi.org/10.12677/csa.2017.71009

可翻转像素点，然后利用Hash函数中的MD5算法加密产生水印信息，最后将这些水印信息嵌入到可翻转像素点中。将在接收到的文本图像中提取的水印信息与之前产生的水印信息相比较，就可以实现文本图像的真实性及完整性认证。该算法有良好的篡改定位能力。

关键词

文本图像，Hash函数，脆弱水印，篡改定位

1. 引言

随着互联网的发展，电子商务和电子政务随之兴起，重要文件资料、经济合同、电子发票等更多文本图像需要网络传送。这些文件一旦出现恶意篡改而无法证明真伪，就会造成严重后果。因此，研究文本图像认证方法，对电子政务和电子商务具有重要作用[1] [2]。文本图像特点：前后景区域对比鲜明，隐藏噪声易于显现，隐藏冗余非常有限，所以在文本图像中隐藏信息的难度较大[3] [4]。脆弱水印技术，由于其极容易被破坏的特性，非常适合用于图像内容的认证，可以有效检测篡改和篡改定位。文献[5]提出利用混沌算法将水印嵌入的最不重要像素块中，视觉质量比较好，但是定位能力比较差；文献[6]将水印嵌入文本空白处，定位效果比较好，但是视觉效果不理想。本文提出一种基于 Hash 函数的脆弱水印算法，利用 Hash 函数对输入值的高敏感性，来实现脆弱水印在文本图像的嵌入及篡改定位。

2. Hash 函数

Hash 函数在现代密码学中起着重要作用。其有主要性质：1) 易计算：对于任意给定的消息，容易计算其哈希值；2) 压缩性：任意长度的数据，算出的摘要长度固定；3) 敏感性：对数据进行任何改动，甚至只修改 1 个字节，所得到的摘要会有很大区别；4) 抗碰撞性：对于给定的消息 M_1，要发现另一个消息 M_2，满足 $H(M_1) = H(M_2)$在计算上是不可行的；找一对不同的消息 M_1，M_2，使 $H(M_1) = H(M_2)$在计算上是不可行的。

Hash 函数主要有 MDx 系列和 SHA 系列。MDx 系列包括 MD5、HAVAL、RIPEMD 一 128 等；SHA 系列包括 SHA0、SHA1、SHA256 等。MD5 和 SHAl 应用最广，MD5 产生的摘要信息为 128 位，SHA 产生的摘要信息为 160 位。

单向散列算法，即信息-摘要算法(Message-Digest Algorithm 5, MD5)，在 20 世纪 90 年代初由 MIT Laboratory for Computer Science 和 RSA Data Security Inc 的 Rbnald L. Rivest 开发，经 MD2、MD3 和 MD4 发展而来。MD5 以 512 位分组处理输入文本，每个分组又划分为 16 个 32 位子分组。算法的输出由 4 个 32 位分组组成，级联成一个 128 位散列值[4]。虽然 2004 年王小云教授在国际密码学会议公布了 MD 系列算法的破解，但在一般的应用场合，可以认为 MD5 算法是足够安全的。MD5 有良好的压缩性和抗碰撞性等，通过对一段信息产生信息摘要以保证信息的安全性。

MD5 摘要生成流程如图 1。

第一步填充：若输入信息长度(bit)对 512 求余的结果不等于 448，则需要填充。填充一个 1 和 n 个 0，使得求余结果等于 448。填充完后，信息长度为 $N \times 512 + 448$；

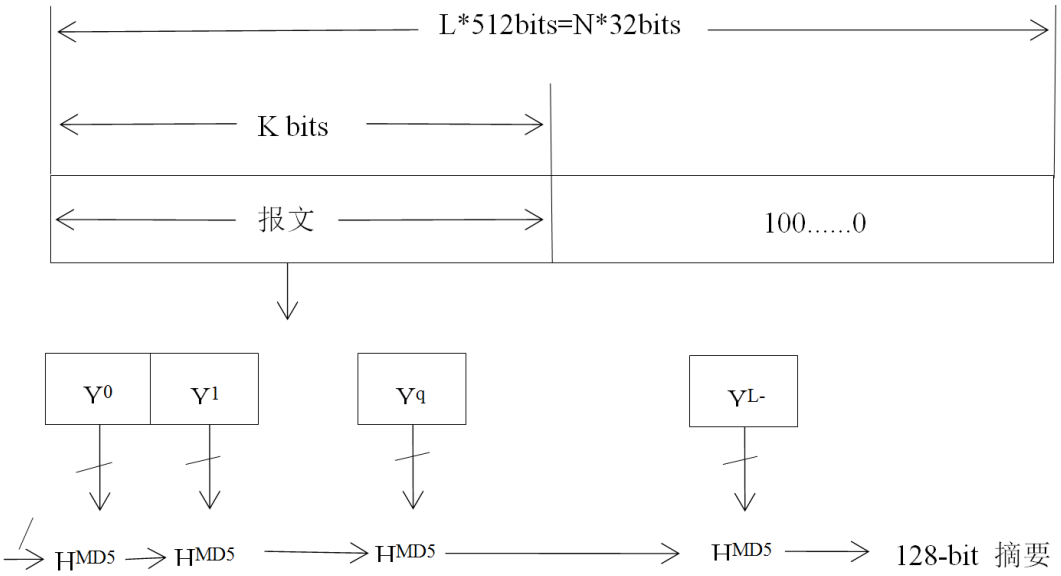


Figure 1. MD5 calculation flow
图 1. MD5 计算流程

第二步记录：用 64 位来存储填充前信息长度。这样信息长度就变为 $N \times 512 + 448 + 64 = (N + 1) \times 512$ 位。

第三步装入：装入标准幻数，物理顺序是 A = (01234567)16，B = (89ABCDEF)16，C = (FEDCBA98)16，D = (76543210)16。

第四步循环：循环运算，循环次数是分组的个数(N + 1) [7] [8] [9]。

3. 水印算法的设计及实现

3.1. 水印的生成与嵌入

3.1.1. 水印生成

1) 原始文本图像分块：分成 3 m × 3 n 大小的块，每块标记为 Ai，即每块中含有 $m \times n$ 个 3 × 3 的小块；

2) 可翻转像素点确定：根据 MinWu 提出的可翻转像素准则[10]，统计每块中可翻转像素点以及不可翻转像素点；

3) Hash 摘要生成：将每块中不可翻转像素作为 MD5 算法的输入，生成长度为 L 的二值序列，作为水印 Wi；

4) 将水印 Wi 加密保存：用 logistic 混沌生成混沌序列 $W1$，Wi 与 $W1$ 异或生成加密水印。存储 Hash 摘要要比存储原始文本图像更有效。

以分块 $A1$ 为例，MD5 摘要生成如表 1 所示。

3.1.2. 水印的嵌入

1) 将原始文本图像分成 3 m × 3 n 大小的分块，每块标记为 Ai；

2) 用 logistic 混沌映射生成一个混沌序列 $W2$，用 $W2$ 来选定与 Ai 相对应的块 Bi；

3) 将水印信息 Wi 嵌入到 Ai 中对应的可翻转像素中，即得到嵌入水印的文本图像。

用混沌序列指引水印嵌入块，保密水印的嵌入位置，提高水印的抗攻击能力[11]。对可翻转像素进行翻转，对图像的视觉影响很小，所以将水印嵌入到可翻转像素中的视觉隐蔽性很高[10]。在保证图像视觉质量的前提下，将水印均匀的嵌入到每个分块中，实现水印的脆弱性[12]。这样含水印图像如果遭到恶意篡改，

就会导致水印发生相应的变化，根据水印的变化可以确定篡改位置。

水印嵌入流程如图 2 所示。

3.2. 水印提取与篡改检测

3.2.1. 水印提取

1) 把待检测文本图像分成 3 m × 3 n 大小的分块；

2) 用 logistic 混沌映射确定分块 Bi'；

3) 从映射块 Bi'中可翻转像素点提取 Wi'。

3.2.2. 水印篡改检测

1) 篡改检测：可以通过计算提取水印和原始水印的归一化系数(NC)判断图像是否被篡改，NC 越接近于 1，说明提取的水印与原始水印越接近。理论上，当 NC 为 1 时，图像未被篡改，当 NC 不为 1 时，说明图像被篡改。

$$NC = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} W(i,j)W'(i,j)}{\sum_{i=1}^{m}\sum_{j=1}^{n} W(i,j)^2}。$$

2) 篡改定位：如果 $Wi \neq Wi'$则 $A\_flag(x, y) = 1$，可以判定图像被篡改；如果 $Wi = Wi'$，则 $A\_flag(x, y) = 0$，则图像未被篡改，从而实现篡改区域的认证和定位。

水印提取与检测流程如图 3 所示。

4. 算法仿真

建立图像库，现存 50 幅文本图像。在 MATLAB 环境下，对该算法进行仿真实验，验证算法的有效性。

4.1. 不可见性

图 4 为原始文本图像示例。图 5 为对应嵌入水印的文本图像示例。对 50 幅文本图像进行仿真，视觉

Table 1. Encryption watermarking
表 1. 加密水印

图像类型	Hash摘要	混沌序列	加密水印
文字文本	cabdd132b2e92026 704072e0c657dace	d7b6aa18b02ff3a1 383744f09f7a6b12	1d0b7b2a02c6d387 48773610592db1dc
医学单据	be2ae3d82199298a 04d186ee8d6e15c8	d7b6aa18b02ff3a1 383744f09f7a6b12	699c49c091b6da2b 3ce6c21e12147eda
电子发票	7d6671213c0bf3db 017d41b3e94a5c7e	d7b6aa18b02ff3a1 383744f09f7a6b12	aad0db398c24007a 394a05437630376c

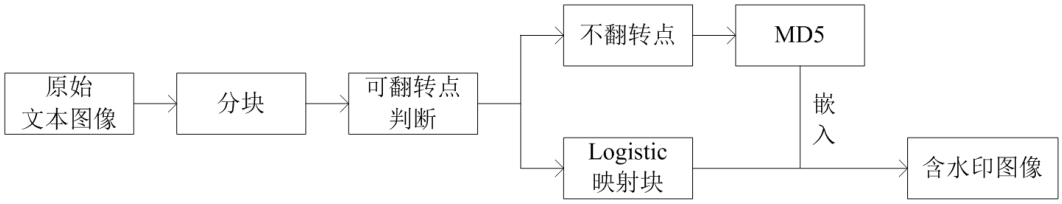
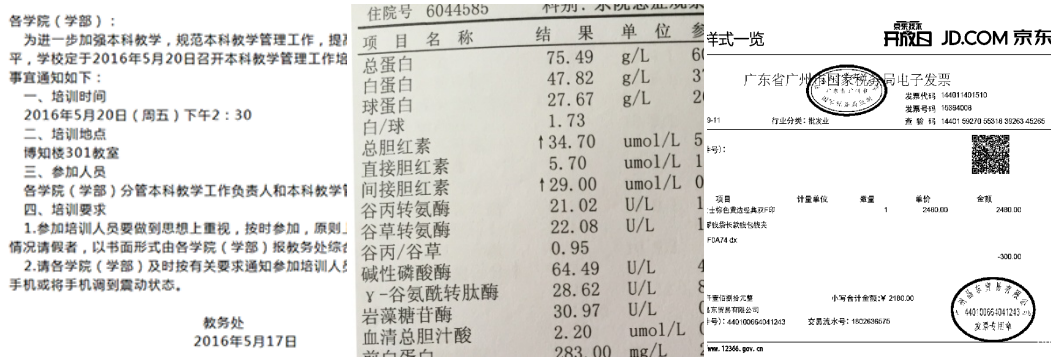


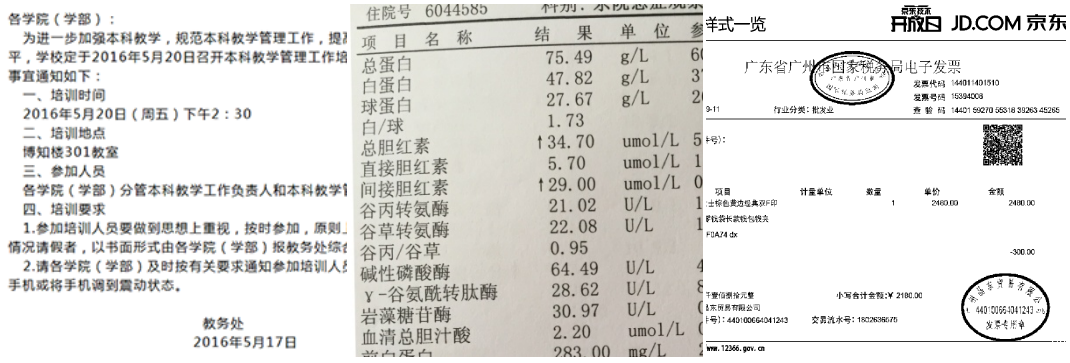
Figure 2. Watermark embedding flow diagram
图 2. 水印嵌入流程框图

Figure 3. Watermark extraction flow diagram
图 3. 水印提取流程框图



(a) (b) (c)

Figure 4. Watermark extraction flow diagram: (a) Text; (b) Medical documents; (c) Electronic invoice
图 4. 水印提取流程框图：(a)文字文本；(b) 医学单据；(c) 电子发票



(a) (b) (c)

Figure 5. Examples of watermarked images: (a) Watermarked text; (b) Watermarked medical documents; (c) Watermarked electronic invoice
图 5. 含水印图像示例：(a) 含水印文字文本；(b) 含水印医学单据；(c) 含水印电子发票

不可见性良好。客观评定用峰值信噪比 *PSNR*，其值越高，不可见性越好。它的临界值一般取为 30 dB，峰值信噪比超过 30 dB，就说明含水印的图像的透明性较好。50 幅文本图像的 *PSNR* 都在 45 dB 以上，从而证明该算法有良好的不可见性。在含水印图像中提取水印信息并计算对应的归一化系数，*NC* 及 *PSNR* 如表 2 所示。

4.2. 篡改定位

对含水印文本图像进行局部的篡改。图 5(a)第 6 行中下午篡改为上午，第 15 行中静音改为震动；图 5(b)第 3 行数据 27.67 篡改为 45.36，第 10 行数据 0.95 改为 3.08；图 5(c)将单价 2480 改为 4560，将合计金额 2180 篡改为 4260。篡改定位仿真示例如图 6 所示。计算每种篡改图像的归一化系数、虚检率及漏检率，如表 3 所示。

仿真实验结果表明，该算法具有很好的视觉不可见性，同时对水印图像的篡改具有很高的敏感性，可以鉴别文本图像的正确性和完整性。

Table 2. Invisible features
表 2. 不可见性

图像类型	*NC* 系数	*PSNR*
文字文本	1	46.182 dB
医学单据	1	54.236 dB
电子发票	1	50.513 dB

Table 3. P_{fa} and P_{fr} of tampering and positioning
表 3. 篡改定位虚检率及漏检率

图像类型	*NC* 系数	虚检率(%)	漏检率(%)
文字图像	0.89	1.67	0.34
医学检验图像	0.91	1.93	0.58
电子发票	0.86	2.07	0.65

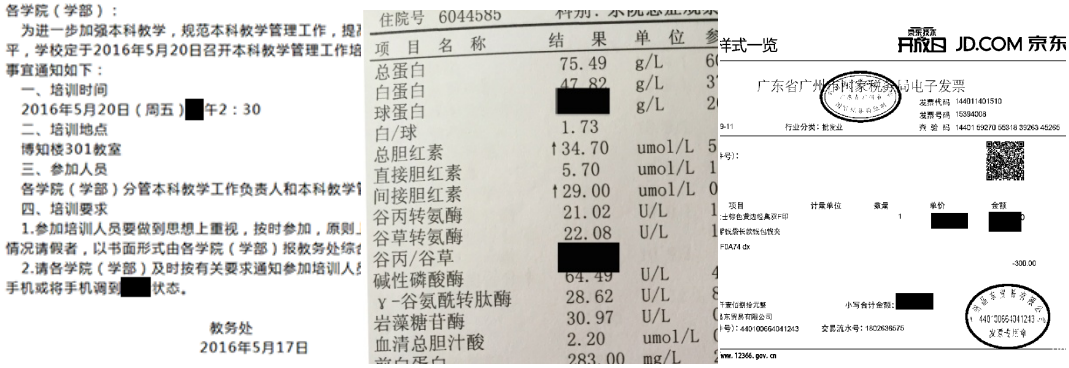


Figure 6. Examples of tampering and positioning
图 6. 篡改定位示例

5. 结束语

本文提出了一种针对于文本图像认证的脆弱水印算法。通过利用 Hash 函数对输入值的高敏感性，来实现脆弱水印在文本图像的嵌入及篡改定位。通过将文本图像分块，然后计算出每个块中不可翻转像素和可翻转像素，然后将不可翻转像素利用 Hash 函数的 MD5 算法加密生成含有图像内容的水印信息，然后将其嵌入到不可翻转像素中就得到了嵌入水印的文本图像，此算法通过实验证明有很好的水印不可见性和篡改定位能力。

致　谢

在此感谢导师的指导和帮助，在我有问题不能解答的时候对我细心的指导，也感谢实验室同学和朋友的帮助，在我遇到困难的时候帮助我解决问题并给予关心，并对给予转载和引用权的资料、图片、文献、研究思想和设想的所有者，表示感谢。

参考文献 (References)

[1] 谭利娜. 文本图像鲁棒认证技术研究[D]: [博士学位论文]. 长沙: 湖南大学, 2012.

[2] 孙圣和, 陆哲明, 牛夏牧, 等, 著. 数字水印技术及其应用[M]. 北京: 科学出版社, 2004: 536.

[3] 秦瑶. 内容真实性认证的文本图像脆弱水印算法[D]: [硕士学位论文]. 成都: 西南交通大学, 2014.

[4] 余振山. 自然语言文本中数字水印的设计与研究[D]: [博士学位论文]. 合肥: 中国科学技术大学, 2009.

[5] 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[J]. 小型微型计算机系统, 2006, 27(1): 151-154.

[6] 胡精易, 毛建旭, 赵希. 一种用于认证二值文本图像窜改定位的数字水印算法[J]. 计算机应用研究, 2012, 29(12): 4631-4634.

[7] 林克正, 李东勤, 李绍华. 基于 Hash 函数的脆弱图像水印算法[J]. 哈尔滨工程大学学报, 2008, 29(1): 61-64.

[8] 毛熠, 陈娜. MD5 算法的研究与改进[J]. 计算机工程, 2012, 38(24): 111-114.

[9] Bruce, Schneier, 祝世雄. 应用密码学——协议、算法和 C 源程序[J]. 信息安全与通信保密, 1994(3): 312-313.

[10] 白树锋, 邓立新. 脆弱水印在电子病历中的应用[J]. 软件导刊, 2012, 11(6): 149-150.

[11] 陈善学, 彭娟, 李方伟. 基于二维 Logistic 混沌映射的 DWT 数字水印算法[J]. 重庆邮电大学学报(自然科学版), 2012, 24(4): 495-500.

[12] Wu, M., Tang, E. and Lin, B. (2000) Data Hiding in Digital Binary Image. *IEEE International Conference on Multimedia and Expo*, **1**, 393-396. https://doi.org/10.1109/ICME.2000.869623