

基于点阵式字符编码的文本信息隐藏算法^①

杨玉军 杨夷梅 (怀化学院 计算机科学与技术系 湖南 怀化 418008)

摘要: 针对电子文档内容的版权保护问题和现有文本信息隐藏算法鲁棒性不强的特点,提出了一种基于点阵式字符编码的文本信息隐藏算法。该算法先对要隐藏的信息进行点阵式编码处理,使字符具有图像的特性,然后根据 Logistic 混沌系统对该点阵式编码信息进行二值化处理加密,结合 HVS 和加密后信息对文本的颜色进行微小修改,从而实现信息的隐藏。实验结果显示该信息隐藏算法对文本内容的编辑具有很好的鲁棒性和实用性。

关键词: HVS; 点阵式字符编码; 信息隐藏; 文本信息

Text Information Hiding Algorithm Based on Dot-Matrix Character Code

YANG Yu-Jun, YANG Yi-Mei

(Department of Computer Science and Technology, Huaihua University, Huaihua 418008, China)

Abstract: To address the problem of copyright protection for electronic document content and the poor robustness of watermarking algorithm based on text information, a text information hiding algorithm based on dot-matrix character code is proposed. The hiding information is firstly coded by dot-matrix, and the character has the image characteristic property. Then the coded information is encrypted by a two-value chaotic sequence generated by Logistic system. To hide the information, a slight modification of color is made based on HVS and the encrypted information. Experiments have shown that the algorithm has a better watermarking robustness and better practicability for editor attacking of text content.

Keywords: HVS; dot-matrix character code; information hiding; text information

1 引言

信息隐藏方法^[1]在进行秘密信息传递时,相对于单纯加密等其它方法而言具有更高的安全性和可靠性,它可将秘密信息隐藏在合法的文件中,使得攻击者不易发现和分析,由于信息隐藏方法的这种优势,它的使用日益广泛。在计算机网络技术飞速发展的今天,人们进行信息交流更加方便和快捷,但是给人们带来了越来越严重的安全问题,如媒体作品的版权问题。

目前的研究主要是利用利用图像、声音等载体来进行信息隐藏。这主要是因为图像、声音等数字载体自身具有一定的数据冗余度,结合人的视觉特性,可

以在这些载体对象里隐藏信息而不容易察觉。而在文本信息隐藏方面,由于文本自身结构的一些特殊性,限制了它的发展速度,但也涌现了一些信息隐藏的方法。例如 Brassil 和 Maxem chunk 等人^[2]提出了在 Postscript 格式中嵌入水印的三种方法,即行移处理、字移处理和特征处理。Purdue 大学的 Atallah 等人^[3]提出了一种鲁棒性较好的文本隐藏算法。但目前大多数的研究论文都很少涉及完整的、失效性强的文本信息隐藏算法。

本文提出一种基于点阵式字符编码的文本信息隐藏算法,该算法先对要隐藏的信息进行点阵式编码处理,使字符具有图像的特性,然后根据 Logistic 混沌

① 基金项目:湖南省怀化学院科研项目(HHUY2008-17)

收稿时间:2009-12-08;收到修改稿时间:2010-01-09

系统对该点阵式编码信息进行二值化处理加密,结合HVS和加密后信息对文本的颜色进行微小修改,从而实现信息的隐藏。实验结果显示该信息隐藏算法对文本内容的编辑具有很好的鲁棒性和实用性。

2 点阵式字符编码和人眼视觉系统

2.1 点阵式字符编码

点阵式字符编码是以字符的点阵字库为基础的字符编码技术,点阵字库是一个保存了所有文字的点阵数据的数据文件。使用过“文曲星”之类的电子辞典和LED的人们会发现上面显示的汉子就能够明显地看出“点阵”的痕迹,而计算机显示器的显示分辨率很高,肉眼一般无法区分,因此“点阵”的痕迹不是那么明显。

点阵、矩阵、位图这三个概念在本质上是有关联的,从某种程度上讲,这三个就是同义词。点阵从本质上讲就是单色位图,他使用一个比特来表示一个点,如果这个比特为0,表示某个位置没有点,如果为1表示某个位置有点。矩阵和位图有着密不可分的联系,矩阵其实是位图的数学抽象,是一个二维的阵列。位图就是这种二维的阵列,这个阵列中的(x,y)位置上的数据代表的就是对原始图形进行采样量化后的颜色值。以UCDOS中文宋体字库为例,根据国标汉字库中的每一个字均由256点阵来表示的标准,每一个字由16行16列的点阵组成,每一个点可以理解为一个像素,因此每一个字的字形可以理解为一幅图像。事实上任何一个字符不仅可以显示为字符,也可以显示为256像素的图像。

2.2 人眼视觉系统

人眼视觉系统(Human Visual System,HVS)是由眼球、神经系统及大脑的视觉中枢构成。一般来说,人眼对亮度区域所加噪声的敏感性较小,即图像的背景亮度越高,它所能嵌入的信息就越多。对于频率特性来说,频率越低人眼的分辨能力就越高,即人眼对低频内容比较敏感。从图像类型特性来说,图像可以分为平滑区域和纹理区域,人眼的视觉系统对平滑区域的敏感性要远高于纹理密集区域。从颜色感知特性来说,HVS对不同的颜色有不同的感知。对实体颜色的像素值 $I = RGB(V_R, V_G, V_B)$,由Weber定律^[4]其可见性检测门限为:

$$\Delta I = 0.02 \times I = RGB(0.02V_R, 0.02V_G, 0.02V_B) \quad (1)$$

这表明人眼是无法对实体颜色在 $RGB(\pm 0.02 \times$

$V_R, \pm 0.02 \times V_G, \pm 0.02 \times V_B)$ 范围内进行区分的。因此,从人体视觉系统特性来看,对文本颜色做微小的改变来隐藏信息会具有很好的隐藏效果。与此同时,文本颜色的改变不会改变文本的内容和阅读,隐藏信息后的文本与没有隐藏信息的文本实际上没有任何差别的。

3 基于点阵式字符编码的文本信息隐藏算法

算法涉及到的参数定义: T表示载体文本; W表示要隐藏的信息; K表示加密算法的密钥; $R \times C$ 表示字符点阵阵列。 T_w 表示经过隐藏 W 信息后得到的文本。

3.1 信息隐藏算法

基于点阵式字符编码的文本信息隐藏算法描述如下:

(1)将要隐藏的信息 W 转换成为点阵式字符编码序列 $M = \{m_1, m_2, \dots, m_i, \dots, m_n\}$, 其中 m_i 是要隐藏信息的某个字符的点阵式编码;

(2)将点阵式字符编码序列 M 转换成二进制形式 $B = \{b_1, b_2, \dots, b_j, \dots, b_{r \times c \times n}\}$, 其中 b_j 是要隐藏信息字符的点阵式编码的一位二进制位;

(3)在密钥 K 的作用下, 由 Logistic 混沌系统 A:

$$X_{n+1} = 4 * X_n * (1 - X_n) \quad (2)$$

生成混沌序列, 经过二值化处理, 得到二值化的混沌序列 $S = \{s_1, s_2, \dots, s_k, \dots, s_n\}$, 其中 $s_k \in \{0, 1\}$, 密钥 $K \in (0, 1)$, 且密钥 K 是一个具有多位小数的有理数, 具体数值由用户确定;

(4)根据序列 S 的值, 当 s_k 值为“1”时, 如果所对应的信息 $b_j = 1 (j < k)$, 则将对对应文字颜色的三基色值(即 Red, Green, Blue 基色值, 简称 RGB 值)减 1, 如果 $b_j = 0$, 则将对对应文字颜色的 RGB 值保持不变; 当 s_k 值为“0”时, 则不在该文字对象中隐藏信息;

(5)待所有信息嵌入后, 即获得含有隐藏信息的文本 T_w 。

3.2 隐藏信息提取或检测算法

基于点阵式字符编码的文本隐藏信息提取或检测算法描述如下:

(1)在密钥 K 的作用下, 由混沌系统 A 生成混沌序列, 经过二值化处理, 得到二值化的混沌序列 $S = \{s_1, s_2, \dots, s_k, \dots, s_n\}$, 其中 $s_k \in \{0, 1\}$;

(2)根据序列 S 的值, 当 s_k 值为“1”时, 如果所对

应文字颜色的 RGB 值偏差为 1, 则对应的信息 $b_j = 1 (j < k)$, 否则 $b_j = 0$; 当 s_k 值为“0”时, 则不做任何处理, 继续后面的操作;

(3) 循环执行(2), 直到所有的信息都被检测处理, 得到二进制集合 $B = \{b_1, b_2, \dots, b_j, \dots, b_{r \times c \times n}\}$;

(4) 对集合 B 进行为点阵式字符编码序列 $M = \{m_1, m_2, \dots, m_i, \dots, m_n\}$;

(5) 对 m_i 进行点阵式字符编码校正和识别处理, 即隐藏的信息 W。

4 算法的性能分析

算法测试在 P4 2.0GHz×2, DDRAM 1024M, Windows XP Professional 5.1 以及 VBA 开发环境中进行, 图 1 和图 2 分别是原始的部分文本和含有隐藏信息“HuaiHua University”的部分文本。对比图 1 和图 2, 信息隐藏在实际应用中是不可察觉的, 同时经过测试, 可以正确提取隐藏信息“HuaiHua University”。

西班牙, 是位于欧洲西南部伊比利亚半岛上的一个国家, 该国的舞蹈种类繁多, 主要的有博雷罗舞、佛拉门哥舞、几丹文舞、巴斯克舞等等。这些舞蹈热烈鲜艳如烈火雄狮, 奔放有力如海浪滔滔, 有着激动人心的节奏和旋律, 有着端庄自傲和刚柔并蓄的美妙风姿……受到了世界各国人民的喜爱和欢迎。

著名古典芭蕾舞剧《天鹅湖》中的西班牙舞, 是各国芭蕾舞剧热门演出的第一个版本, 它最初是在王子齐格弗里德挑选未婚妻的盛典上所出现的娱乐即兴节目之一, 由两个徒手男青年和两个手持折扇的女青年表演。舞者以大滑步和伸展的舞姿, 以及多变的舞臂动作, 开始对舞。舞者大幅度的后仰前踢, 前仰后踢,

用足掌足跟击地轻巧的打点, 上身微微晃动; 男舞者用倾身弯腰小跳作陪衬。女舞者大幅度的踢腿翻身动作, 增添和发展了欢乐活泼的气氛。婉转柔美的手臂动作、显示了高超的技艺。

风格独特而斑斓多姿的西班牙舞, 集中地体现了西班牙人民乐观豪迈、傲然自得的民族个性和俊美矫健、神采飞扬的精神面貌。其舞蹈风格的形成, 除了本民族的内在因素, 还有外在的影响。影响最大的是 14 世纪时吉卜赛人个性强烈、作风泼辣, 兼之喜爱音乐, 能歌善舞, 他们在与西班牙人民相处的过程中, 很自然地将其奔放无羁的风度和富有感染力的表演特色融入了西班牙舞之中, 使之异常美妙, 而令人难忘。

图 1 原始的部分

西班牙, 是位于欧洲西南部伊比利亚半岛上的一个国家, 该国的舞蹈种类繁多, 主要的有博雷罗舞、佛拉门哥舞、几丹文舞、巴斯克舞等等。这些舞蹈热烈鲜艳如烈火雄狮, 奔放有力如海浪滔滔, 有着激动人心的节奏和旋律, 有着端庄自傲和刚柔并蓄的美妙风姿……受到了世界各国人民的喜爱和欢迎。

著名古典芭蕾舞剧《天鹅湖》中的西班牙舞, 是各国芭蕾舞剧热门演出的第一个版本, 它最初是在王子齐格弗里德挑选未婚妻的盛典上所出现的娱乐即兴节目之一, 由两个徒手男青年和两个手持折扇的女青年表演。舞者以大滑步和伸展的舞姿, 以及多变的舞臂动作, 开始对舞。舞者大幅度的后仰前踢, 前仰后踢,

用足掌足跟击地轻巧的打点, 上身微微晃动; 男舞者用倾身弯腰小跳作陪衬。女舞者大幅度的踢腿翻身动作, 增添和发展了欢乐活泼的气氛。婉转柔美的手臂动作、显示了高超的技艺。

风格独特而斑斓多姿的西班牙舞, 集中地体现了西班牙人民乐观豪迈、傲然自得的民族个性和俊美矫健、神采飞扬的精神面貌。其舞蹈风格的形成, 除了本民族的内在因素, 还有外在的影响。影响最大的是 14 世纪时吉卜赛人个性强烈、作风泼辣, 兼之喜爱音乐, 能歌善舞, 他们在与西班牙人民相处的过程中, 很自然地将其奔放无羁的风度和富有感染力的表演特色融入了西班牙舞之中, 使之异常美妙, 而令人难忘。

图 2 含有隐藏信息的部分文本

4.1 隐藏信息容量分析

由于该算法是通过修改文字的颜色来隐藏信息,

且每一个文字只能隐藏 1bit 的信息, 且算法采用混沌系统来产生二值化的伪随机序列来选择文字进行信息隐藏, 根据混沌系统的特性, 其“0”和“1”的分布是均匀的, 且接近 0.5, 因此对于一个包含 N 个文字的文本文件来说, 其理论可隐藏信息容量 Capa 为 $N/2$ bit, 即:

$$Capa = N \times \frac{1}{2} = \frac{N}{2} \text{bit} \quad (2)$$

4.2 抗攻击能力分析

在实际应用中, 文本内容需要进行部分的修改和移动, 甚至删除等编辑操作, 因此, 需要分析该算法是否具有很好的抗编辑等攻击能力。

考虑文档的正常编辑, 以及文档破坏、信息隐藏攻击等行为通常针对文本连续的若干词、句、段落, 这样对隐藏信息连续嵌入的算法是不利的, 而对于非连续的编辑和攻击, 算法具有很好的效果。信息隐藏检测和提取算法的预取效果和实验效果是吻合的, 如表 1 所示。

表 1 算法针对各种编辑和攻击时信息提取的情况

攻击方式	检测和提取情况 (编辑量百分比为 P)
随机删除或修改	$P \leq 30\%$, 仍可以完整地检测和提出隐藏信息
攻击	$30\% < P \leq 40\%$, 利用点阵式字符校正技术可以大部分地检测和提出隐藏信息
攻击	$40\% < P \leq 60\%$, 利用点阵式字符校正技术可以小部分地检测和提出隐藏信息
攻击	$P > 60\%$, 基本上不能检测和提出隐藏信息
随机插入或移动	$P \leq 100\%$, 利用点阵式字符校正技术仍可以完整地检测和提出隐藏信息
攻击	$P \leq 10\%$, 利用点阵式字符校正技术可以完整或大部分地检测和提出隐藏信息
连续删除或修改	$P > 10\%$, 利用点阵式字符校正技术可以小部分地检测和提出隐藏信息
攻击	$P \leq 100\%$, 利用点阵式字符校正技术仍可以完整地检测和提出隐藏信息
连续插入或移动	$P \leq 100\%$, 利用点阵式字符校正技术仍可以完整地检测和提出隐藏信息
攻击	$P \leq 100\%$, 利用点阵式字符校正技术仍可以完整地检测和提出隐藏信息
随机或连续修改	$P \leq 100\%$, 利用点阵式字符校正技术仍可以完整地检测和提出隐藏信息
标点攻击	$P \leq 100\%$, 利用点阵式字符校正技术仍可以完整地检测和提出隐藏信息

5 结论

针对电子文档内容的版权保护问题和现有文本信息隐藏算法鲁棒性不强的特点, 提出了一种基于点阵式字符编码的文本信息隐藏算法。该算法先对要隐藏的信息进行点阵式编码处理, 使字符具有图像的特性, 然后根据 Logistic 混沌系统对该点阵式编码信息进行二值化处理加密, 结合 HVS 和加密后信息对文本的颜色 (下转第 207 页)

色进行微小修改,从而实现信息的隐藏。实验结果显示该信息隐藏算法对文本内容的编辑具有很好的鲁棒性和实用性,但对连续攻击效果比较差,这也是以后努力解决的问题。

参考文献

- 1 Cox IJ, Kalker T, Pakura G, et al. Information transmission and steganography. Proc of the 4th Int Workshop on Digital Watermarking. LNCS 3710. Berlin: Springer, 2005: 15 – 28.
- 2 Brassl JT, Low S, Maxemchukn F. Copyright protection for the electronic distribution of text documents. Proceedings of the IEEE, 1999, 87(7): 1181 – 1195.
- 3 Atallahm J, Rashn V, Hemplemann C F, et al. Natural language watermarking antamper-proofing proceedings of the 5th Information Hiding Workshop. S1, 2005: 117 – 122.
- 4 Ganzalez RC, Wintz P. Digital image processing(2nd edition). New York: Addison-Wesley Publishing Co, 1987: 38 – 42.