

# 彩色图像中基于LSB的文字Unicode加密隐藏

徐畅凯, 徐文华

(重庆工商大学派斯学院, 重庆 401520)

**摘要:** 针对公众在网上交流时文字信息被窃取的问题, 提出一种文字加密并隐藏于彩色图像的方法。将文字信息的Unicode码编码为20位二进制, 并以7位以下的数字作为密码, 以此为参数构造可逆运算置乱信息二进制编码, 实现加密, 利用LSB算法将加密二进制数据嵌入蓝光通道, 实现信息隐藏。经实验仿真和数据分析表明, 该方法易于实现, 采用本方法能在彩色图像中隐藏大量文字信息而不显著影响图像的质量, 且密图有一定的抗破解能力。

**关键词:** 彩色图像; Unicode码; 隐藏; 加密

**中图分类号:** TP391.41 **文献标识码:** A

## Encryption Hidden of Text Unicode in Color Image Based on LSB

XU Changkai, XU Wenhua

(Dept. of Basic Courses, Pass College of Chongqing Technology and Business University, Chongqing 401520, China)

**Abstract:** Aiming at the problem of information theft when the public communicates online, a method of information encryption and transmission hidden in color images is proposed. The Unicode of text information is extended to 20-bit binary coding, then enter the number less than 7 bits as the key, and the key's binary value is used to construct an invertible function to scramble the information binary coding for the encryption. Then the encrypted binary data is added to the blue-light channel based on LSB algorithm to realize information hiding. With the experiment and data analysis, this paper reveals that this method is easy to implement and the scheme can hide lots of sound information in color images without significantly affecting the quality of the imagery, and the secret image has certain anti-cracking ability.

**Keywords:** color image; Unicode; hiding; encryption

### 1 引言(Introduction)

随着信息技术特别是网络技术的发展, 各种信息的交换与传输变得越来越便捷, 尤其随着智能手机的普及, 人们的各种需要都可以通过手机终端的APP得到帮助, 但在信息传递的过程中, 传输的信息随时面临被窃取, 篡改的危险, 因此, 如何保证网络中信息传输的安全已经成为人们关注的焦点。信息隐藏和加密技术是解决信息安全传输的一种有效途径。将信息隐藏在图像中也是近20年来最常见的技术之一, 在版权保护、认证等方面有着广泛的应用<sup>[1-3]</sup>。

目前, 国内外的信息隐藏术发展为可逆数据隐藏方法和非可逆数据隐藏方法<sup>[4,5]</sup>。根据载体是否可恢复, 在信息隐藏的过程中, 若秘密信息通过载体发送给接受者以后, 可以还原载体的过程称为可逆数据隐藏。对于可逆数据隐藏目前主要有四种方法, 即基于量化的方法、基于直方图修正的方法、基于压缩的方法、双图像的方法。双图像可逆数据隐藏是最近很多学者提出的方法<sup>[6]</sup>, 即在嵌入隐藏数据的过程中生

成两个相似的密图, 该方法和之前的方法相比具有较高的数据嵌入能力和较低的图像失真率。

LSB图像隐藏算法是一种量化的方法。由于图像一般由像素构成的, 每个像素有8位, 通常最后几位的变化, 通过肉眼是无法察觉到。LSB算法就是利用视觉的这一特征, 通过对图像二进制低位进行量化达到嵌入数据的目的。文献[7]的论证表明该方法易于实现, 不可感知性好, 且隐藏容量较大。由于LSB算法对载体的不可逆性, 且易检测、攻击和破解等问题, 很多学者提出了改进的方法, 如文献[8]—文献[15]。其中文献[8]提出基于相邻灰度值对互补嵌入的LSB匹配隐写改进算法。文献[9]提出了利用差值扩展和直方图平移的思想给出了一种可逆数据隐藏的方法, 文献[10]利用离散余弦变换和JPEG图像编码特征给出一种在JPEG图像中的可逆数据隐藏方法。文献[11]—文献[15]针对嵌入效率和嵌入容量提出了改进的可逆数据隐藏方法。

由于现有的数据隐藏方法都是已经公开发表的, 因此,

若这些方法被窃密者了解和掌握其思想或算法，那么，这些数据隐藏方法都是不安全的，针对这个问题，学者们提出了结合密码学进行加密隐藏的方法。这样即使窃密者能够检测出该图像是载密图像，若不能够给出正确的密钥也将无法提取正确的信息。文献[16]—文献[19]给出了利用密钥加密图像的方法。其中文献[18]利用密钥和混沌模型产生一种混沌嵌入模式，使隐藏的数据有较高的安全性，能够抵抗大多数常规攻击，文献[19]中，Ke等人提出利用公钥加密过程中产生的冗余嵌入数据的方法，通过LWE算法加密后产生的信息冗余设计了一种多层加密隐藏方案，可以在载体图像中嵌入多重隐藏信息并且带有多重数据隐藏密钥，实现在特定层次的密钥只能解开特定层次的隐藏信息。

尽管已经有了很多的数据隐藏技术，但是将这些技术具体应用在保护文字安全的不多，大多数数据隐藏技术都应用在数字水印、数字认证和版权保护等方面。人们在利用即时通讯工具或网络上发送私密文字消息时候总是担心信息被第三方偷窥和利用。比如家庭住址、身份证号、银行账号、账户号和密码等等。针对这个问题，本文给出了一种利用LSB算法结合密码将文字信息加密隐藏于彩色图像中并通过密码解密隐藏信息的方法。其中密码是用户任意设定的6位的数字密码，文字是Unicode库中对应的文字，其中包括汉字及外国文字，标点符号，数字和英文字母等。利用该方法，当用户在传输私密信息时，任选一种彩色图像作为伪装，然后设定一组密码将该信息隐藏在图像中，将该图像传输给另一方以后，对方用该密码便可解密其中的信息，文本给出的算法具体内容如下文所述。

## 2 算法理论基础(Theoretical basis of algorithm)

### 2.1 彩色图像

由于彩色是由多种光谱合成的，因此彩色图像也称为多光谱图像。人的视觉系统中存在着杆状细胞和锥状细胞两种感光细胞。杆状细胞为暗视器官，主要功能是辨识高亮度信息；锥状细胞是明视器官，主要功能是在一定的亮度下分辨颜色。因此锥状细胞是负责彩色视觉的传感器，其可分为三个主要的类别。第一类是大约65%的对红光敏感的锥状细胞，第二类是大约33%的对绿光敏感的锥状细胞，第三类是大约2%对蓝光敏感的锥状细胞。由于人眼对光线的这些吸收特性，所以我们常看到的彩色就被认为是红色(Red)、绿色(Green)和蓝色(Blue)的各种组合，即三基色。根据不同的应用，彩色图像常用的颜色模型有RGB、CMY、HSV、HIS、YUV、YIQ等。本文以RGB颜色模型为研究基础。RGB模型采用三基色构成表色系统，也就用红绿蓝三色混成自然界的任一颜色。颜色传感器把数字图像上的一个像素编码成(R, G, B)，每个分量量化分为256级，因此RGB模型可表示 $256 \times 256 \times 256$ 约1670万种颜色。假设有一副像素 $n \times m \times 3$ 的彩色图像，可将该彩色图像表示为二维空间变量和光谱变量的函数 $f = f(x, y, \lambda)$ ，其中 $(x, y)$ 为像素位置， $\lambda$ 为光谱分量。 $\lambda = 1, 2, 3$ 分别代表像素红、绿、蓝分量。

### 2.2 文字编码

目前，我们记录信息的文字主要是由26个英文字母、十个数字、标点符号和约7万个汉字组成。这些文字在国际上有统一的Unicode编码，也是一种国际标准编码，每一个文字都用一个16进制数进行编码。编码采用的是UCS-2，即用

两个字节来编码一个字符，两个字节就是16位二进制，2的16次方等于65536，所以UCS-2最多能编码65536个字符，它的前128个字符和ASCII码一致，中、日、韩的三种文字占用了Unicode中0x3000到0x9FFF的部分，目前绝大多数常用汉字已经有了对应Unicode编码。假设输入文字信息的长度为N，则其对应的Unicode编码可以看作空域变量的函数 $g = g(x, y)$ ，其中 $x$ 代表文字顺序， $x \in [1, N]$ ， $y$ 代表第 $x$ 个字符编码的位置， $y \in [1, 16]$ ， $g$ 为编码值，其值为0或1。

## 3 文字加密隐藏算法(Text encryption and hiding algorithm)

为了实现文字信息的加密和隐藏，其基本思路是先将文字转化为对应的Unicode码，然后将Unicode码转为20位的二进制编码，利用密码对编码加密，最后将加密的编码嵌入到图像。解密则是先读取像素中特定编码，然后利用密码将编码解码，并将解出的编码转化为Unicode码，最后将Unicode码转换为对应的文字，其过程如图1所示，其关键技术如下所述。

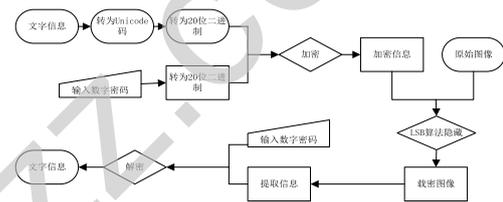


图1 算法流程图

Fig.1 Algorithm flowchart

### 3.1 文字加密

文字加密是对其Unicode编码加密，密码是其加密参数。设定加密信息所用密钥为1—6位的自然数，则其数值范围为0—999999，共100万个组合，其对应数字的二进制值长度最大为20位。如999999的二进制数为11110100001000111111。由于输入字符信息和国际通用的16位Unicode编码是一一对应的，因此加密文字信息可以通过密码与Unicode编码建立一一映射，以此给出加密文字的编码，从而达到加密的目的。文字加密算法如下：

- (1)输入 $N$ 个文字。
- (2)获取 $N$ 个文字的Unicode码。
- (3)将(2)Unicode码编码为20位二进制，可将其二进制编码看作空间变量的函数 $g = g(x, y)$ ，其中， $x$ 代表文字序号， $x \in [1, N]$ ， $y$ 代表第 $x$ 个字符编码位， $y \in [1, 20]$ ， $g$ 为编码值，其值为0或1。
- (4)输入一个6位及以下的数字 $k$ ，作为密码。
- (5)将 $k$ 编码为20位2进制，记 $k$ 的编码为函数 $h = h(x)$ ，其中 $x$ 为编码位， $h$ 为编码值0或1。
- (6)将(3)中得到的每个文字的二进制码 $g(x, y)$ 分别和密码编码 $h(x)$ 进行逻辑排斥或运算，就得到加密后的文字信息，记加密后的编码为 $g'(x, y)$ ，则：

$$g'(x, y) = g(x, y) \oplus h(y) \tag{1}$$

其中，符号 $\oplus$ 为排斥或运算符。

因为 $g'(x, y)$ 是文字信息加密后二进制码，而16位二进制代码和文字是一一映射，所以此时直接显示 $g'(x, y)$ 对应的文字信息就是一个密文，也就是加密后的文字，要得到原文字信息，需要利用密钥 $h(x)$ 和 $g'(x, y)$ 反解 $g(x, y)$ 。

3.2 文字隐藏

文字信息由式(1)加密以后，得到加密信息的二进制码  $g'(x, y)$ 。信息隐藏就是将加密信息  $g'(x, y)$  隐藏于彩色图像  $f = f(x, y, \lambda)$  中。其算法如下：

(1)计算加密编码  $g'(x, y)$  在图像中占用的行数  $c_1$  和列数  $w_1$ 。

其计算方法如下：

$w_1 = 20 \times \lfloor m / 20 \rfloor, d = \lfloor m / 20 \rfloor, c_1 = \lceil N / d \rceil$  (2)

其中， $m$  为图像的列数， $N$  为文字中总数。

(2)对图像蓝光通道  $c_1 \times w_1$  部分像素进行归一化处理，将该部分数值全部变为偶数，得到新的图像  $f'(x, y, \lambda)$ 。 $f'(x, y, \lambda)$  计算方法如下：

$f'(x, y, \lambda) = \begin{cases} f(x, y, \lambda) - 1 & \lambda = 3 \wedge f \equiv 1(\text{mod } 2) \\ f(x, y, \lambda) & \text{其他} \end{cases}$  (3)

(3)将加密编码  $g'(x, y)$  依次加入到图像  $f'(x, y, \lambda)$  蓝色  $c_1 \times w_1$  的空间中得到新的彩色图像  $h(x, y, \lambda)$ ，其方法如下：

$h(x, y, \lambda) = \begin{cases} f'(x, y, \lambda) + g'(x, y) & \lambda = 3 \\ f'(x, y, \lambda) & \text{其他} \end{cases}$  (4)

此时图像  $h(x, y, \lambda)$  的  $c_1 \times w_1$  蓝色灰度的奇偶性即对应加密信息  $g'(x, y)$  的1和0值。

(4)当(3)过程完成以后，继续将其余蓝色灰度归一化为偶数直到  $c_1 + 1$  行停止。

根据以上算法， $n \times m$  的图像  $f'(x, y, \lambda)$  可以隐藏的文字个数  $N$  为

$N = \lfloor m / 20 \rfloor \times n$  (5)

3.3 文字解密

在得到隐藏信息的彩色图像  $h(x, y, \lambda)$  后，解密信息的过程就是先读取隐藏在图像  $h(x, y, \lambda)$  蓝色灰度的二进制加密信息  $g'(x, y)$ ，然后通过之前的密码编码  $h(x)$  还原原文的二进制编码  $g(x, y)$ ，再转换为Unicode码，进而解密文字。由真值表或等值演算易证：

若  $g'(x, y) = g(x, y) \oplus h(y)$  则  $g(x, y) = g'(x, y) \oplus h(y)$  (6)

其证明的真值表如表1所示。

表1 证明式(4)的真值表

Tab.1 Truth table of proof formula(6)

Table with 4 columns: g(x,y), h(x), g'(x,y), g'(x,y)oplush(x). Rows show binary combinations (0,0), (0,1), (1,0), (1,1) and their corresponding g'(x,y) and g'(x,y)oplush(x) values.

显然(6)式成立，由此提取隐藏文字的算法如下：

(1)读取隐藏信息的载体图像  $h(x, y, \lambda)$ 。

(2)依次从左到右，从上到下读取蓝光波段数据，若  $h(x, y, 3) = 1(\text{mod } 2)$ ，则加密文字的二进制编码为1；若  $h(x, y, 3) = 0(\text{mod } 2)$ ，则加密文字的二进制编码为0。每获取20个编码，存为一个文字编码。

(3)记录连续编码为0的个数  $k$ ，若  $k > 20$  则停止存储文字。

(4)假设(2)中提取的编码为  $\bar{g}(x, y)$ ，则编码就是(1)式中中对信息加密后的编码  $g'(x, y)$ ，即

$\bar{g}(x, y) = g'(x, y)$  (7)

(5)利用密钥  $h(x)$  和  $g'(x, y)$ ，反解  $g(x, y)$ ，即

$g(x, y) = \bar{g}(x, y) \oplus h(y)$  (8)

(6)提取编码  $g(x, y)$  右边16位二进制编码，将其转化为16进制数，即得到文字的Unicode码。

(7)根据(6)中的Unicode码，转化为对应的文字，即实现了文字的解密。

4 实验结果及分析(Experimental results and analysis)

在仿真实验中，本文选择大小为  $512 \times 512 \times 3$  的图像 Lana.bmp 为隐藏文字的载体，利用本文引言部分文字作为密文，在64位win7系统下利用MATLAB 2010b进行仿真实验，通过Matlab的API和GUI实现了语音的加密和隐藏算法。软件界面及运行截图如图2所示。



(a)文字加密程序界面



(b)文字解密程序界面

图2 算法运行程序界面

Fig.2 Algorithm running program interface

4.1 实验结果

按照本文算法，由式(5)可知，该照片可以隐藏12800个文字。选择本文引言部分含空格520个字符复制4遍，共计2080个字符作为隐藏信息嵌入载体图3(a)中，得到载密图3(b)如图3所示。



(a)载体图



(b)载密图

图3 载体和载密图

Fig.3 The carrier image and secretcarried image



$ER = 3bpp$ 。通常,人们日常使用手机相机的像素都是千万像素的,对一张手机拍摄的 $4000 \times 3000 \times 3$ 彩色照片来说,其单通道隐藏的文字数为60万字,就是说一张照片可以隐藏一篇180万字的长篇小说。实验仿真也进一步验证了该算法在图片中的确可以隐藏大量文字,隐藏容量远远满足日常的需要。

考虑到方便和实用,算法没有设计复杂的密码,采用跟银行卡密码一样纯6位数字的密码作为密钥加密信息,其密钥空间为 $10^6$ ,与文献[20]加密算法相比,密钥空间很小,不能抵御穷举攻击,对于日常应用,该密码空间可能足够,但若需要保密级别更高的话,需要拓展更多的密钥空间。从表3可以看出密钥对信息的影响是敏感的,密钥相差一个值,解密信息差之千里。因此,对于在网络上传输敏感信息或记录个人隐私信息时,通过此算法将信息隐藏于彩色图像中,就不用担心信息泄露。因为彩色图像本身就是一种伪装,我们无法通过视觉来辨识一张图像有没有隐藏信息,即使知道某张图像隐藏了信息,要解密只有试探出100万个密钥中的唯一正确密码才能解密出真实信息。

## 5 结论(Conclusion)

本文针对网络信息传输的安全问题,提出了一种将文字信息加密隐藏于彩色图像的方法。该方法利用6位自然数的密钥,构造了一种可逆运算,对文字的Unicode编码进行加密,然后将加密信息隐藏于彩色图像中。解密信息则是通过特定算法提取照片中加密的Unicode编码,然后利用加密运算的逆运算解码出原文的Unicode编码,从而实现解密信息。通过仿真及数据分析可以得出,本文方法可以实现在彩色图像中隐藏大量文字而不影响原图的质量,具有很强的隐蔽性,抗直方图检测和统计监测,而且解密的可能性是百万分之一。因此,该算法可以有效解决文字信息通过QQ、微信和邮件等方式传输泄密的风险,不足之处是密图不能抗攻击、抗噪,以及任何形式的图像更改都能破坏加密信息,从而无法解密,后续可以继续研究密文的抗攻击能力。

## 参考文献(References)

[1] Bender W, Gruhl D, Morimoto N, et al. Techniques for data hiding[J]. IBM System Journal, 1996, 35(3,4): 313-336.

[2] Subhedar M S, Mankar V H. Current status and key issues in image steganography: A survey[M]. Elsevier Science Publishers B.V. 2014.

[3] Ni Z C, Shi Y Q, Ansari N, et al. Reversible data hiding[J]. IEEE Transactions on Circuits Systems Video Technology, 2006, 16(3): 354-362.

[4] Arooj Nissar, A. H. Mir. Classification of steganalysis techniques: A study[J]. Digital Signal Processing, 2010, 20(6): 1758-1770.

[5] Zhang Tao, Zhang Hao, Wang Ran, Wu Yun da. A new JPEG image steganalysis technique combining rich model features and convolutional neural networks[J]. Mathematical biosciences and engineering: MBE, 2019, 16(5): 4069-4081.

[6] Ki-Hyun Jung. A Survey of Reversible Data Hiding Methods in Dual Images[J]. IETE Technical Review, 2016, 33(4): 1-12.

[7] Chang C C, Lin M H, Hu Y C. A FAST AND SECURE IMAGE HIDING SCHEME BASED ON LSB SUBSTITUTION[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2002, 16(04): 399-416.

[8] 奚玲, 平西建, 张涛. 基于相邻灰度值对互补嵌入的LSB匹配隐写改进算法[J]. 计算机科学, 2010, 37(09): 101-104.

[9] 王继军. 利用差值扩展和直方图平移的可逆数字水印算法[J]. 小型微型计算机系统, 2014, 35(05): 1192-1195.

[10] Huang F, Qu X, Kim H J, et al. Reversible Data Hiding in JPEG Images[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(9): 1610-1621.

[11] Wedaj F T, Kim S, Kim H J, et al. Improved reversible data hiding in JPEG images based on new coefficient selection strategy[J]. EURASIP Journal on Image and Video Processing, 2017, 2017(1): 63.

[12] Weng S W, Zhang G H, Jeng-Shyang Pan et al. Optimal PPVO-based reversible data hiding[J]. Journal of Visual Communication and Image Representation, 2017(48): 317-328.

[13] Li X L, Li J, Li B, et al. High-fidelity reversible data hiding scheme based on pixel-value-ordering and prediction-error expansion[J]. Signal Processing, 2013, 93(1): 198-205.

[14] 李桂芸, 邓桂英, 赵逢禹. 一种基于LSB图像信息隐藏的改进算法[J]. 计算机系统应用, 2012, 21(04): 156-160.

[15] 任克强, 肖璐瑶. 融合CFT和LSB的高容量可逆数据隐藏[J]. 液晶与显示, 2019, 34(04): 410-416.

[16] Hamidreza Rashidy Kanan, Bahram Nazari. A novel image steganography scheme with high embedding capacity and tunable visual image quality based on a genetic algorithm[J]. Expert Systems With Applications, 2014, 41(14): 6123-6130.

[17] Gyan Singh Yadav, Aparajita Ojha. Secure data hiding scheme using shape generation algorithm: a key based approach[J]. Multimedia Tools and Applications, 2018, 77(13): 16319-16345.

[18] Gyan Singh Yadav, Aparajita Ojha. Chaotic system-based secure data hiding scheme with high embedding capacity[J]. Computers and Electrical Engineering, 2018(69): 447-460.

[19] Ke Y, Zhang M Q, Liu J, et al. A multilevel reversible data hiding scheme in encrypted domain based on LWL[J]. Journal of Visual Communication and Image Representation, 2018(54): 133-144.

[20] 陈善学, 唐义嫻. 基于混沌系统的RGB彩色图像三重置乱算法[J]. 重庆邮电大学学报(自然科学版), 2018, 30(06): 812-818.

## 作者简介:

徐畅凯(1983-), 男, 硕士, 讲师. 研究领域: 计算机图形图像处理.

徐文华(1984-), 女, 硕士, 讲师. 研究领域: 应用数值代数.