

基于线性预测和位操作的信息隐藏算法

朴红吉, 郑品, 田雄, 冯林

(大连理工大学 大学生创新院, 辽宁 大连 116023)

摘要: 鉴于信息的安全性问题, 信息隐藏技术也已成为信息安全领域的研究热点。利用图像中当前像素点邻近的相关像素的像素值, 对该点的值进行线性预测。再根据预测的准确度, 确定相应的阈值, 并结合位操作, 提出了一种新的简单易于实现的信息隐藏算法。并在 Lena 图像上进行了实验, 得到了隐藏率为 0.160bits/Byte, 信噪比为 42.276dB 的结果。经初步测试, 证明了在保证较高隐藏率的前提下, 能够较好地提高信噪比。

关键词: 信息隐藏; 线性预测; 位操作

中图分类号: TP309.2

文献标识码: A

文章编号: 1673-629X(2008)01-0185-03

Data Hiding Algorithm Based on Linear - Prediction and Bit - Operation

PIAO Hong-ji, ZHENG Pin, TIAN Xiong, FENG Lin

(Institute of University Students' Innovation, Dalian University of Technology, Dalian 116023, China)

Abstract: People are facing more and more problems on information security, and information hiding has been a research hotspot in this field. First, linear - predict the value of a pixel in an image based on the values of the surrounding pixels. According to the veracity of the linear - prediction, and based on bit - operation, propose a new watermarking algorithm for digital images, which is easy to realize. A test on Lena image has been done, the hiding rate is 0.160bits/Byte, and the SNR (Signal - to - Noise) is 42.276dB. The new algorithm has been proved that people can use it to embed a large payload while keeping the distortion low.

Key words: data hiding; linear - prediction; bit - operation

0 引言

在数字信息飞速发展的今天, 信息的安全性问题日益突出, 对信息隐藏技术的要求也越来越高。从 LSB 隐藏算法^[1], 到 Fridrich 等人提出的基于图像空域和变换域的无损信息隐藏方法^[2,3], 再到 Thodi 提出的一种基于预测误差扩展的无损数据隐藏方法^[4], 以及谢于明等人对线性预测的进一步研究^[5], 国内外专家学者提出了许多信息隐藏的方法和应用, 使信息隐藏技术得以迅速发展。

文中提出的隐藏算法是: 首先根据某个像素点 X 邻近相关像素点的像素值, 对 X 的值进行线性预测, 根据预测值的准确程度, 确定相应的阈值。再根据 X 的原始值和预测值, 利用位操作的方法确定 X 点处的隐藏值, 从而达到信息隐藏的目的。提取算法则是隐藏算法的逆过程。该算法经过初步测试, 在嵌入数据量和

信噪比两方面都得到了较好的隐藏效果。

1 预测算法

1.1 预测模版

将一幅数字图像 P 如图 1 简化表示。设像素点 X 的原始值为 originalValue, 预测值为 predictionValue, 隐藏值为 hideValue。用 X 的邻近相关像素 A、B、C、D、E、F 的值 a、b、c、d、e、f 来预测 originalValue, 预测算法为:

$$\text{predictionValue} = \lfloor (a + b + c + d + e + f) / 6.0 + 0.5 \rfloor \quad (1)$$

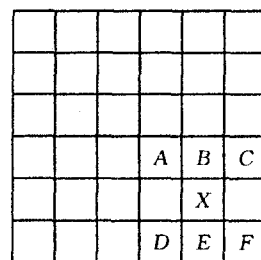


图1 简化后的数字图像

1.2 可嵌入区域

设数字图像 P 的像素矩阵有 m 行 n 列。为保证提

收稿日期: 2007-03-19

基金项目: 国家自然科学基金资助项目(50575031)

作者简介: 朴红吉(1985-), 男(朝鲜族), 吉林人, 主研方向为信息隐藏; 冯林, 博士, 教授, 研究方向为图像压缩、配标及融合、演化算法。

取和嵌入时 X 的预测值相同且在尽可能多的像素点嵌入数据,应采用隔行扫描方式对图像进行扫描。则可嵌入区域 $S(x, y)$ 的取值范围为(逆光栅顺序):

$$x = (m - 1) - (2k + 1), 0 \leq k \leq (m - 3)/2, k \text{ 为整数} \quad (2)$$

$$1 \leq y \leq m - 1$$

2 阈值的确定

2.1 阈值 NEGLECT

为增大嵌入数据量,在保证不太大地影响信噪比的前提下,改变的图像数据可以与原图像数据不相同。用阈值 NEGLECT 来衡量改变的数据与原数据相差的程度。NEGLECT 定义如下:

$$\text{NEGLECT} = \lfloor \log_2 |\text{originalValue} - \text{hideValue}| \rfloor \quad (3)$$

2.2 阈值 RESERVATION

为最大程度地增大嵌入数据量,且使隐藏信息后的数字图像有较高的信噪比,应适当选择参与运算的像素值的位数,即不考虑像素值的高位,而只对像素值的低位进行相应位移运算。把这个参与运算的位数阈值称为 RESERVATION。而 RESERVATION 的选取应取决于预测算法的预测准确度。

以 131×131 的 24 位 Lena 灰度图像为例,按 1.1 节中的预测算法,采用隔行扫描方式,其预测准确度如表 1 所示。

表 1 1.1 节预测算法在 Lena 图像上的统计

pos	7	6	5	4	
p	0.0367	0.0919	0.1358	0.1402	
pos	3	2	1	0	-1
p	0.1602	0.1440	0.1068	0.0781	0.1059

表 1 中数据表示的意义是:pos 表示首次出错位, p 表示概率。originalValue 与 predictionValue 从高到低逐位比较,第一次不同出现在第 3 位的概率是 0.1602;originalValue 与 predictionValue 完全相同的概率是 0.1059。由统计数字可知,概率较大的首次出错位为 5、4、3、2。参考 NEGLECT 的选取(RESERVATION 的值应大于 NEGLECT + 1),可确定 RESERVATION 应取为 3,即只有低 3 位参与位移运算。

利用原始图像数据,在 RESERVATION = 3 时再次对 131×131 的 24 位 Lena 灰度图像进行统计,得到表 2。

表 2 RESERVATION = 3 时 1.1 节预测算法在 Lena 图像上的统计

首次出错位	2	1	0	-1
概率	0.1467	0.1324	0.2562	0.4646

3 隐藏算法

Step 1: 获取寄主图像的像素矩阵。

Step 2: 准备预隐藏信息的 bit 流。

Step 3: 嵌入信息。

设原始图像数据的像素值为 originalValueR(该值的位数为 RESERVATION,即原值 originalValue 的低 RESERVATION 位);预测值为 predictionValueR(predictionValue 的低 RESERVATION 位);隐藏值为 hideValueR(hideValue 的低 RESERVATION 位);预隐藏信息的 bit 流为: $b_0 b_1 b_2 b_3 b_4 \dots$ 。

(1) 将 originalValueR 和 predictionValueR 从高位到低位逐位比较,找到第一个不同的位的索引值,记为 predictionIndex(predictionIndex 代表着预测的精度),则 predictionIndex 的取值范围为: RESERVATION - 1 ~ -1(注:predictionIndex 等于 -1 表示 originalValueR 和 predictionValueR 的各个位完全相同)。若 predictionIndex 的值不在 [NEGLECT, -1] 范围内,则跳转到(2),否则跳转到(3)。

(2) 将 originalValueR 左移 RESERVATION - 1 - predictionIndex 位(溢出部分全部舍去不考虑),从预隐藏信息的 bit 流中提取前 RESERVATION - 1 - predictionIndex 位,并用提取的 RESERVATION - 1 - predictionIndex 位取代 originalValueR 的低 RESERVATION - 1 - predictionIndex 位,所得到的新的值就是隐藏值的低 RESERVATION 位,即 hideValueR。比较 hideValueR 与 originalValueR 的值,若高 RESERVATION - 1 - NEGLECT 位都相同,说明 hideValueR 相对 originalValueR 的变化在预定阈值限定范围内,可以在这个像素中隐藏信息;否则,说明 hideValueR 相对 originalValueR 的变化不在预定阈值限定范围内,不能在这个像素中隐藏信息,这时将 predictionValue 赋给 hideValue,即不在这个像素中隐藏信息,这时要将从预隐藏信息流中提取的 RESERVATION - 1 - predictionIndex 位补回预隐藏信息流中。

(3) 这种情况下与(2)相似,只是须添加一些附加信息。因为允许隐藏值与原始像素值的后 NEGLECT + 1 位不同,为了在信息提取时能够准确地提取信息,必须把预测精度,即 predictionIndex 的值以附加信息流的形式保存起来,以供提取信息时使用。因为 predictionIndex 只能取到 NEGLECT + 2 个值(从 NEGLECT 到 -1),所以每条附加信息应用 $\lfloor \log_2 (\text{NEGLECT} + 2) \rfloor + 1$ 个 bit 表示。附加信息流用 addition 表示。

(4) 用 hideValueR 的取代 originalValue 的低 RESERVATION 位,所得结果即为 hideValue 的值。

若 hideValue 和 originalValue 的值相同,则不在此像素点隐藏信息,即把 predictionValue 的值赋给 hideValue。

(5)用 hideValue 取代原始图像矩阵中的 originalValue,实现信息隐藏。

用上述方法,从最后一个数据起,按逆光栅顺序隔行扫描图像矩阵,直到将所有信息都隐藏完为止,记录此时的像素的行列数,记为 (i, j) 。需要注意的是欲隐藏信息的长度要在图像矩阵的可容纳范围内。

addition 的处理方法:将 $i + j + \text{addition} + \text{EOF}$ (addition 结束标志)用 LSB 算法隐藏到图像矩阵中,从第一个数据起,按光栅顺序逐点扫描图像数据,直到 $i + j + \text{addition} + \text{EOF}$ 全部隐藏完,记录此时的像素的行列数,记为 (i_1, j_1) 。要保证 $i_1 * n + j_1 < i * n + j$ 。

4 提取算法

Step 1:获取带有隐藏信息的图像数据。

Step 2:提取 $i + j + \text{addition}$ 。从第一个数据起,按光栅顺序逐点扫描图像数据。

Step 3:提取隐藏信息,其 bit 流记为 secretBits

(1)若 hideValue 和 predictionValue 的值相同,则说明该像素内没有隐藏信息,跳过这个像素。若 hideValueR 和 predictionValueR 的值不同,跳转到(2)。

(2)将 hideValueR 和 predictionValueR 从高位到低位逐位比较,找到它们第一个不同的位的索引值,记为 predictionIndex(predictionIndex 的意义和取值范围与嵌入时相同)。若 predictionIndex 的值在 $[\text{NEGLECT}, -1]$ 内,则说明预测精度不能从 hideValueR 和 predictionValueR 的比较中得出,应从 addition 中提取出 $\lfloor \log_2 (\text{NEGLECT} + 2) \rfloor + 1$ 个 bit 值确定 predictionIndex 的值。

(3)以 predictionValueR 的高 $\text{RESERVATION} - 1 - \text{predictionIndex}$ 为高位,以 hideValueR 的高 $\text{predictionIndex} + 1$ 位为低位构成 originalValueR。再用 originalValueR 取代 hideValue 的低 RESERVATION 位,即得到恢复的像素值 originalValue。而 hideValueR 的低 $\text{RESERVATION} - 1 - \text{predictionIndex}$ 为隐藏信息的一部分,添加到 secretBits 中。

隔行扫描整个图像矩阵,从最后一个数据起,按逆光栅顺序隔行扫描图像数据,直到第 (i, j) 个数据为止。

(4)保存 secretBits。

5 嵌入量分析

嵌入量的大小取决于预测算法的优劣,而相同的

预测算法对于不同的图像也会产生不同的结果。嵌入量的大小还取决于阈值 RESERVATION 和 NEGLECT 的选取。因此要对具体的算法和图像做统计后才能计算出可能的嵌入量。

对 1.1 节中的预测算法和 131×131 的 24 位 Lena 灰度图像,当阈值 $\text{RESERVATION} = 3, \text{NEGLECT} = 1$ 时的理论嵌入量计算如下:

$$\frac{3}{2} \times 0.1467 + \frac{2}{2+3} \times 0.1324 + \frac{1}{2} \times 0.2562 + \frac{0}{2} \times 0.4646 = 0.3069$$

即对于每个可嵌入区域的图像数据,都能隐藏 0.3069 bit 的信息。

6 实验及结果

6.1 文中算法在 Lena 图像上的实验结果

对一幅 131×131 像素的 24 位 Lena 灰度图像的试验结果如下(见图 2):

当设定阈值 $\text{RESERVATION} = 3, \text{NEGLECT} = 1$, 嵌入信息量为 2744 bits 时, $\text{PSNR} = 42.276\text{dB}$, 提取后图像相对原图像的信噪比 $\text{PSNR} = 42.603\text{dB}$, 可以无损地提取隐藏信息。



(a)原始图像



(b)嵌入信息的图像



(c)提取信息后的图像

图2 文中算法在 Lena 图像上的实验结果

6.2 文中算法与其它算法的比较

文中算法与其它算法比较见表 3。

文中算法与 Thodi 算法比较,隐藏率略低,但信噪比略高;文中算法与 JPEG-LS 预测算法比较,虽隐藏率较低,但信噪比却较高。文中算法更适用于对隐藏率和信噪比都有较高要求的应用。

(下转第 191 页)

```

/* 计算 TCP 校验和 */
libnet_do_checksum(buf, IPPROTO_TCP, TCP_H);
/* 发送包 */
libnet_write_ip(libnetsock, buf, TCP_H + IP_H)

```

4 结束语

使用协议分析技术,对 SMTP 协议进行解析,进而对邮件底层特性和邮件内容特性检测并发现邮件入侵,使邮件防护系统提高垃圾邮件的正确阻截率,解除邮件用户的烦恼,不受到有害邮件的攻击,对于维护网络安全具有重要意义。就目前而言,对 SMTP 协议的分析的研究还比较少,从捕获到的邮件内容信息中找到敏感的信息,如果单纯采用简单的字符匹配,误判率高,所以对数据进行预处理、智能分词、快速索引等

技术手段,将是以后研究的方向。

参考文献:

(上接第 184 页)

- [3] Hoglund G. A Real NT Rootkit, Patching the NT Kernel[J]. Phrack Magazine, 1999, 9(55): 55 - 65.
- [4] Warrender C, Forrest S, Pearlmuter B. Detecting intrusions using system calls alternative data models[C]//IEEE Symposium on Security and Privacy. Oakland: IEEE Computer Society, 1999: 133 - 145.
- [5] 杨风召,朱扬勇,施伯乐. IncLOF: 动态环境下局部异常的增量挖掘算法[J]. 计算机研究与发展, 2004, 41(3): 477 - 484.
- [6] Samhain Labs. The Basics - Subverting the Kernel[EB/OL]. 2003 - 01. <http://la-samha.de/library/rootkits/basics.html>.
- [7] Samhain Labs. Detecting Kernel Rootkits[EB/OL]. 2003 - 01. <http://la-samha.de/library/rootkits/detect.html>.

- [1] 谢希仁. 计算机网络[M]. 大连: 大连理工大学出版社, 2004.
- [2] Comer D E, Stevens D L. 用 TCP/IP 进行网际互连[M]. 林瑞, 蒋慧轩译. 北京: 电子工业出版社, 1998.
- [3] 周 婕. 协议分析在入侵监测系统中的应用[J]. 计算机与网络, 2003(6): 137 - 140.
- [4] 蔡伟鸿, 汤立浩. 电子邮件分析系统的设计[J]. 汕头大学学报, 2004(2): 75 - 78.
- [5] Jonathan. Simple mail transfer protocol[S]. RFC821. 1982.
- [6] 胡吉明, 刘少君. 状态协议分析技术在 TCP 中的应用[J]. 计算机技术与发展, 2006, 16(3): 212 - 216.

(上接第 187 页)

表 3 文中算法与其它算法的比较

	隐藏率(bits/Byte) (统计平均值)	信噪比(dB) (统计平均值)
文中算法	0.160	42.276
Thodi 算法	0.182	40.754
JPEG-LS 预测算法	0.364	31.270

7 结束语

文中结合线性预测和位操作,提出了一种新的简单易实现的信息隐藏算法,具有较高的嵌入量和信噪比,能够无损提取隐藏信息。但图像不能无损提取,在传输的安全性和抗干扰,以及提取预测算法的预测准确度等方面是今后的研究重点。

参考文献:

- [1] 付 兵. 一种增加 LSB 信息隐藏量的方法[J]. 长江大学学报: 自然版, 2006, 3(4): 73 - 75.
- [2] Fridrich J, Goldjan M, Du R. Invertible authentication[C]//Proceedings of SPIE. San Jos, California: [s. n.], 2001: 197 - 208.
- [3] Goldjan M, Fridrich J, Du R. Distortion-free data embedding[C]//Proceedings of the 4th Information Hiding Workshop. Pittsburg, PA: [s. n.], 2001: 27 - 41.
- [4] Thodi D M, Rodriguez J J. Reversible watermarking by prediction-error expansion[C]//Proceedings of the 6th IEEE Southwest Symposium on Image Analysis and Interpretation. Lake Tahoe, Nevada: [s. n.], 2004: 21 - 25.
- [5] 谢于明, 程义民, 王以孝, 等. 基于线性预测的图像无损信息隐藏方法[J]. 计算机辅助设计与图形学学报, 2006, 18(4): 585 - 591.