一种基于相邻像素差值的隐写分析方法①

张 淞, 范明钰

(电子科技大学 计算机科学与工程学院, 成都 611731)

摘 要: 首先介绍 BMP 文件结构及特点, 基于相邻像素相关的特性, 提出了一种新的针对 LSB 替换的隐写分析算法. 分别从横向和纵向扫描图像像素并统计相邻像素之间差值, 然后根据差值奇偶和的比值对图像是否含有隐秘信息进行判断. 实验编程环境为 VS2008, 实验结果表明此检测算法统计量小、实现简单, 在隐秘信息嵌入量较多时具有较高检测率.

关键词: 隐写分析; BMP 文件; 相邻像素

Steganalysis Method Based on the Difference of Adjacent Pixels

ZHANG Song, FAN Ming-Yu

(Dept. of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: This dissertation first introduces the structure and characteristic of BMP file. Based on the features of adjacent pixels, a steganalysis algorithm for the replacement of LSB is proposed. Scanning and counting the difference of adjacent pixels horizontally and vertically are the first step, and then this thesis judges whether the image contains secret information according to the ratio of odd and even. The development environment is VS2008, and the experiment shows that the algorithm can be simply applied with a small statistics. In addition, the new algorithm has a high detection capability with a large amount of secret information.

Key words: steganalysis; BMP file; adjacent pixels

20 世纪 90 年代以来, 随着计算机网络的广泛普及, 计算机技术、信息处理技术和网络通信技术的迅猛发展, 信息隐藏(Information Hiding)[1,2]技术越来越受到人们的重视, 信息隐藏的最大优点是: 除通信双方以外的第三方不知道隐藏消息存在这个事实. 其中以数字图像作为载体的信息隐藏技术中, 通过替换图像空域的最低有效比特位(LSB)来实现隐藏秘密信息的方法是出现较早, 也是研究较为深入的. 目前网络上的许多信息隐藏工具都是基于 LSB 嵌入方法的, 比如 S-Tools、Hide4PGP 等.

信息隐藏技术虽然有利于确保国家机密在网络上安全传递, 但也会被非法分子所利用, 于是隐写分析技术就应运而生了. 隐写分析技术是对信息隐藏技术的攻击技术, 是对抗非法信息传播的一个重要方法.

目前国内外已提出的隐写分析方法有很多, 比如 Westfeld[3]提出的 $\chi 2$ 检验, 对于顺序嵌入的情况可以估计出嵌入信息长度和位置, 但当载体的颜色频率成分较多、嵌入信息不均匀时检测准确率将大大下降. J. Fridrich 等人提出的 RS 检测方法[4]对随机嵌入秘密信息具有较高的检测率, 但是仅适用于灰度和彩色图像; 随后又提出了 RQP 检测方法[5], 此方法仅针对 24 位真彩色图像, 且对于颜色数目较多的图像检测效果不理想. Dumitrescu 等人提出了 SPA[6](Sample Pair Analysis)算法, 这个方法在 RS 的基础上对 LSB 替换做了更深入的分析, 并从理论上证明了 RS 算法从实质上来说是 SPA 算法的一个特例. 张涛等人[7]提出的差分直方图转移系数法, 在隐藏信息量高于 40%时优于 RS 的检测性能. 张新鹏[8,9]提出的 GPC(gray-level plane crossing)

① 基金项目:国家高技术研究发展计划(863)(2009AA01Z403)

收稿时间:2012-03-29;收到修改稿时间:2012-05-06

隐写分析方法计算简单，但检测性能还有待提高.

本文提出了一种新的利用相邻像素相关性进行隐写分析的算法，该算法统计量小、实现简单，在隐秘信息嵌入量较多时可以快速有效地检测出隐秘信息的存在.

1　BMP文件结构及特点[10,11]

BMP 是 Window 操作系统中的标准图像文件格式，典型的 BMP 图像文件由文件头、位图信息头、颜色信息和图形数据四部分组成.

文件头结构和信息头结构分别为 14 字节和 40 字节的数据类型，这两个类型共占 54 字节，因此在读取图像数据的时候，是从第 55 个字节开始读取的. 需要说明的是，在 BMP 位图信息头数据用于说明位图的尺寸等信息，其中 19-22 字节存储图像的宽度(像素)，表示为 biWidth. 23-26 字节存储图像的高度(像素)，表示为 biHeight. 在提取宽度和高度信息时，从高位字节开始计算，比如 19-22 字节数字(16 进制)分别为: 35、17、02、00，则表示图像宽度为 0*16*16*16*16*16*16+2*16*16*16*16+1*16*16*16+7*16*16+3*16+5=131072+4096+1792=137013(像素).

使用 WinHex 软件打开如图 1(a)所示的原始图，可以看到其数字化表示如图 1(b)，其数据是以 16 进制数表示，以 16 个字节为一行，左边表示行数，中间是图像的数据值，右边表示每个图像的数据值作为 ASCII 码对应的字符.

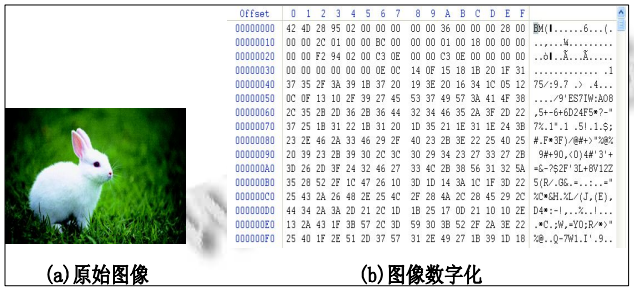

图 1　图像数字化

由图 1(b)可以看出，第一个字节和第二个字节分别为 42 和 4D，对应的字符分别为 B 和 M，这个是位图文件的类型，为固定值; 19-22 字节为 2C、01、00、00 表示图像宽度为 300(像素), 23-26 字节为 BC、00、00、00，则表示图像高度为 188(像素).

从第 55 个字节开始，是该 BMP 图像文件的像素数据部分，对于 24 位的 BMP 图像，每连续 3 个字节便描述图像一个像素点的颜色信息，这 3 个字节分别代表 B(蓝)、G(绿)、R(红)三基色在此像素中的亮度，若某连续 3 个字节为: FFH，00H，00H，则表示该像素的颜色为纯蓝色.

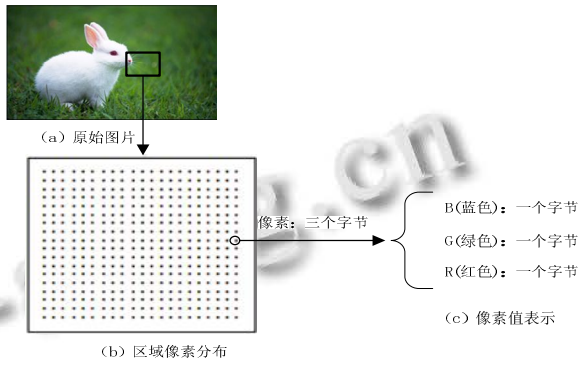

图 2　图像结构表示

在图 2 中，截取原始图像的某个区域，放大后可以看到图像是由很多像素点所组成(如图 2(b))，而每个像素点分别由 B、G、R 三个字节表示.

有一点需要说明，每一个像素值记录顺序在扫描行内是从左到右，扫描行之间是从下到上. 即是从图像的左下角开始，横向是往右扫描，纵向往上扫描，最后到达右上角.

2　算法改进设计

本算法的理论基础是，对于自然图像的像素，在一定的邻域内有一定的相关性，在嵌入隐秘信息后会破坏这种相关性. 本算法对自然图像的像素分别进行横向和纵向差值运算，两两像素差值范围从 0 到 255，嵌入隐秘信息后差值为 0 的个数会减少，差值为 1 的个数会增加，由此假设嵌入信息前，像素差值的奇偶和基本相等，即两者比值约为 1，而潜入后奇数和会增加，偶数和会减少，则奇偶和比会大于 1，由此来检测图像中是否含有隐秘信息. 在此称此种检测算法为 S 检测. 首先进行的是训练，算法具体实现为:

假设图像的宽和高分别为 M、N 像素，即 biWidth=M，biHeight=N，数组 sub_heng[256]存储横向扫描相邻像素差值从 0 到 255 的个数，sub_zong[256]存储横向扫描相邻像素差值从 0 到 255 的个数. E_odd 表示 sub_heng 和 sub_zong 中差值为奇数的个数与差值的乘积的和，E_even 表示 sub_heng 和 sub_zong 中差值

为偶数的个数与差值的乘积的和，P 表示 E_odd 与 E_even 的比值，也即使用 P 来判断图像嵌入信息与否. 算法伪代码如下:

```
Open image a.bmp //打开位图
get biWidth,biHeight   //位图的宽度和高度
Position = 54; //定位于第 55 个字节，因为是从 0 字节开始
get first pixel to cont[3]//顺序取一个像素，即三个字节
for i= 0 to biHeight-1   //横向差值统计
  Position = 54 + i * biWidth; //定位
  get a pixel to con [3] //顺序取一个像素
for j= 0 to biWidth-1
  Position = 54 + i * biWidth + j * 3   //定位
  sub_heng[cont[3]-con [3]]++ //统计差值
  cont[3]= con [3]
for i= 0 to biWidth -1   //纵向差值统计
  Position = 54 + i * 3; //定位
  get a pixel to con [3] //顺序取一个像素
  for j= 0 to biHeight-1
    Position = 54 + j *biWidth + i*3; //定位
    sub_zong[cont[3]-con [3]]++; //统计差值
    cont[3]= con [3]
for i= 0 to 255//统计奇偶和
  if(i%2!=0)
   E_odd+=i*sub_heng[i]+i*sub_zong[i]//奇数和
  Else
   E_even+=i*sub_heng[i]+i*sub_zong[i]//偶数和
P=E_odd/E_even //计算 P 值
```

需要说明的是，横向扫描时，每次取出的是一个像素，在进行差值运算时，是分别将 B、G、R 与前一个像素的 B、G、R 进行差值运算，而不是相邻字节之间取差值，这是由于相邻像素之间相同颜色的相关性更强，从而减少统计误差.

由训练结果可以得到 P 的取值范围，由此来选择合适的阈值 μ 进行检测. 检测过程既是计算未知图片的 P 值，记为 P'，再通过比较 P'是否在阈值范围内来确定图像是否含密.

3 仿真实验与分析

按照设计的检测算法可以知道，P 的取值范围为 [1-μ ,1+μ](0< μ <1)，因此实验的关键是通过对图像的训练找出 μ 的取值. 训练过程为首先统计原始载体图像(未载入隐秘信息)的 P 值，然后分别再在这些图像中嵌入信息，再统计 P 值.

在 CBIR 图像库[12]和 USC-SIPI 图像库[13]中任意选取 1500 幅图像，这些图像都是 24 位 BMP 图像，其中 500 幅用于训练, 1000 幅用于检测，训练图像有 500 幅是原始图像，与之对应的数据的此载体载密之后的数据，隐藏使用 S-Tools 工具.

首先进行训练，根据算法得到如图 3 所示的图像，其中蓝色为载体图像得到的 P 值，红色为载密图像得到的 P 值.

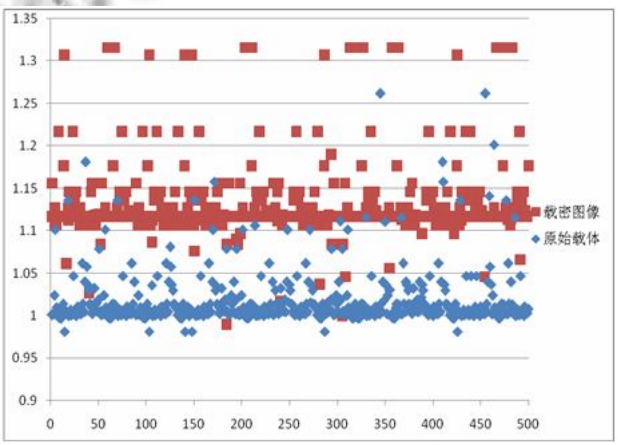

图 3 载体图像与隐秘图像的 P 值分布

从图 3 可以看到，载体图像的值绝大多数都在[1-0.05, 1+0.05]以内，载密图像绝大多数值在 1.1 以上，由此可以将阈值 μ 设为 0.05.

为了检验阈值设定是否准确，使用隐藏工具 S-Tools 和 Hide4GP 分别在载体图像中按 5%、20%、50%、80%、100%的比例嵌入秘密消息，然后对嵌入不同比例秘密消息的图像用本文所提出的 S 检测方法进行检测，检验结果还包括误报率，因为漏报率=1-检测率，表中就不再列出这个结果，具体检测结果图表 1 所示.

由表 1 可以看出，对于不同的隐藏算法，检测得到的数据基本相当，由此进行下一个实验.

将用于检测的 1000 幅 24 位 BMP 图像做以下操作: 其中 800 幅使用 S-Tools 工具嵌入了秘密信息，200 幅为原始载体，即没有嵌入任何信息. 表 2 是本文提出的 S 检测算法与 RS、RQP 以及 SPA 检测算法的比较，

表中的数字表示检测的准确率.

表 1 算法检测结果(%)

	S-Tools	Hide4GP	误报率	
100%	100	100	0	0
80%	96.91	94.81	1.21	1.34
50%	90.22	89.16	2.17	2.01
20%	85.25	83.12	5.34	6. 14
5%	75.41	70.68	8.23	8.35

表 2 对比检测率(%)

	S 检测	RS	RQP	SPA
100%	100	100	100	100
80%	96.91	96.67	96.72	96.85
50%	90.22	95.62	95.81	95.98
20%	85.25	86.57	87.54	87.74
5%	75.41	78.54	77.21	79.51

由表中的信息可以看出，在嵌入比率较大的情况下，本文提出的检测算法具有较高的效率.

4 结语

本文提出的一种利用相邻像素相关性的隐写分析算法，通过实验得到了预期的结果，从而验证了算法的正确性与可行性. 理论分析和实践表明，该方法只需统计横向和纵向像素的差值情况，统计量小、方法实现简单，在嵌入率比较高的情况下能够达到非常好的检测效果，嵌入量较低的情况下检测效率就不如其他几种方法. 但是该算法只对 LSB 隐写有效，一旦图像采用其他的隐写方法，检测算法就会失效. 因此下一步研究方向为根据自然图像固有的自然特性，找一个通用且有效的隐写分析方法.

参考文献

1 Li B, Fang YM, Huang JW. Steganalysis of Multiple-Base Notational System SteganograPhy. IEEE Signal Processing Letters. 2008, 15: 493–496.

2 Yang CF, Liu FL, Luo XY, Liu B. Steganalysis Frameworks of Embedding in Multiple Least-Signifieant Bits. IEEE Trans. on Information Forensies and Secunty. 2008, 3(4):662–672.

3 Westfeld A, Pfitzmann A. Attacks on steganographic systems. Proceedings of Information Hiding, Third International Workshop. Berlin: Springer-Verlag, 2000:61–67.

4 Fridrich J, Goljanm DR. Detecting LSB stegano graphy in color and grayscale images. IEEE Multmiedia, 2001, 8(4): 22–28.

5 Fridrich J, Du R, Meng L. Steganalysis of LSB encoding in color images. Proc. of IEEE International Conference on Multimedia and Expo. New York, 2000: 1279–1282.

6 Dumitreseu S, Wu XL, Wang Z. Deteetion of LSB SteganograPhy via Sample Pair Analysis. IEEE Trans. on Signal Proeessing, 2003,51(7): 1995–2007.

7 张涛,平西建.基于差分直方图实现 LSB 信息伪装的可靠检测.软件学报,2004,15(1):151 –158.

8 Zhang XP, Wang SZ, Zhang KW. Steganography with leasthistogram abnormality. Computer Network Security, Lecture Notes in Computer Science. Springer-Verlag, 2003: 395–406.

9 张新鹏,王朔中,张开文.数字密写和密写分析.北京:清华大学出版社,2005.

10 贾玉珍,靳冰,刘琮,等.BMP 文件结构的信息隐藏方法与实现.江西理工大学学报,2009,30(1):42–44.

11 周文锦,范明钰,王光卫.一种针对 BMP 格式图像的 LSB 数字隐藏方法.信息安全与通信保密,2005:253–255.

12 CBIR Image Database. University of Washington. [2012-03-22]. http://www.cs.washington.edu/ research/ imagedata-base/groundtruth/

13 USC-SIPI Image Database.[2012-03-22]. http://sipi.usc.edu/ services/database/Database.html

(上接第 155 页)

Data Mining and Knowledge Discovery, 2003, 7.

2 Huang ZX. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. Data Mi-ning and Knowledge Discovery, 1998,2: 283–304.

3 雷小锋,谢昆青,林帆,夏征义.一种基于 K-Means 局部最优性的高效聚类算法.软件学报,2008,7.

4 赵伟,张姝,李文辉.改进 K-means 的空间聚类算法.计算机应用研究.2008,7.

5 Tzortzis G, Likas A. The Global K-Means Clustering Algorithm. Proc. of the International Joint Conference on Neural Networks,2008.

6 夏宁霞,苏一丹,覃希.一种高效的 K-medoids 聚类算法.计算机应用研究,2010,12.