几类典型隐写术分析算法的分析与评价

郎荣玲[1] 夏 煜[2] 郅 艳[3] 戴冠中[2]

[1](西北工业大学应用数学系,西安 710072)

[2](西北工业大学自动控制系,西安 710072) [3](西北工业大学机电学院,西安 710072)

摘 要 隐写术分析是检测、提取、破坏隐蔽载体中秘密信息的技术。检测方法可以分为对比检测和盲检测两类,由于通常得不到用于秘密通信的原始载体,因此一般使用盲检测方法。为使人们对基于图像的隐写术分析的研究现状和现有各种算法的优缺点有个概括了解,首先简要介绍了现有的多种图像隐写术分析方法,然后集中对分辨函数、χ^2 检测和 *RS* 检测 3 种盲检测方法,从算法原理和实验结果两个方面进行了深入的分析,并且从算法的适用范围、检测效果等几个方面对这几种算法进行了比较。在保证误报率合理而尽量降低漏报率的前提下,给出了 *RS* 检测方法的置信区间。

关键词 隐写术 隐写术分析 χ^2 检验 *RS* 检测方法

中图法分类号: TP391 **文献标识码**: A **文章编号**: 1006-8961(2004)02-0249-08

Analysis and Evaluation of Several Typical Steganalysis Algorithms

LANG Rong-ling[1], XIA Yu[2], ZHI Yan[3], DAI Guan-zhong[2]

[1](*Department of Applied Mathematics, Northwestern Polytechnic University, Xi'an* 710072)

[2](*Department of Automatic Control, Nothwestern Polytechnic University, Xi'an* 710072)

[3](*Colloege of Mechanical and Electrical Engineering, Northwestern Polytechnic University, Xi'an* 710072)

Abstract Steganalysis is a technique of finding, distilling and destroying the hiding information in stego-cover. The methods of finding the hiding information can be divided into two kinds. One finds the secret information by contrasting the original cover and stego-cover, and the other ,named blind detection does not have the original cover for contrasting. The original cover for secret communication usually can't be obtained, so the blind detection is widely adopted. In order to give an outline of the research work of steganalysis and compare the advantages and disadvantages of various methods available now, some analysis methods are introduced. The discrimination function test, χ^2-test and *RS*-test are analyzed from both principle and experiment results, and comparisons of the three methods above are also given in certain aspects, such as the applicability and effectiveness of tests. The thresholds of *RS*-test are selected at the condition of minimizing the ratio of missed detection while keeping the number of false accusations reasonable.

Keywords steganography, steganalysis, χ^2-test, *RS*-test

1 引 言

许多年来,各国政府和信息产业部门都很重视网络信息安全技术的研究和应用,近年来国际信息技术研究领域出现了一个新的研究方向——信息隐藏技术,通常也称为隐写术。该技术是将秘密信息隐藏在其他载体中,通过载体的传输,实现信息的传递。隐写术与密码技术不同,隐写术将信息嵌在原始对象中,要求这种嵌入不易被人或计算机检测到,从而使得第三方不知道有秘密消息在传递。密码技术是把明文加密为密文,如果没有正确的密钥将无法读解信息,但是密文也预示了重要信息的存在,必会引起第三方的关注。

隐写术在对安全领域做出贡献的同时,也同样被不法份子所利用。有报道称,恐怖分子利用信息隐

基金项目:国家 863 计划项目资助(2001AA142170)

收稿日期:2003-04-07;改回日期:2003-07-11

藏通过 Internet 传递秘密信息、组织恐怖袭击等。针对这种情况，各国安全机构开展了对隐写术的分析和攻击技术的研究，也就是隐写术分析。对隐藏信息的分析和攻击有检测、提取、破坏等几种形式，目前隐写术分析主要集中在检测和破坏两种方式。检测方法可以分为对比检测和盲检测两种形式。对比检测技术需要隐秘载体和原始载体对比，这种方法相对简单。盲检测技术是指在没有原始载体对比的情况下，仅通过隐秘载体检测隐藏信息。但在通常情况下得不到用于秘密通信的原始载体，因此只能采用盲检测，这无疑增加了检测的难度。

基于图像的隐写术分析在近几年得到了广泛研究，例如，文献[1]提出了分辨函数检验方法，此种方法是通过比较各个位平面特别是最低位平面的随机程度，来判断是否有信息嵌入；文献[2]提出了一种统计检验方法——χ^2 检验，该方法是利用 χ^2 检验比较隐秘载体中像素值的理论频率和实际从样本得到的频率，从而检验是否有信息嵌入；文献[3]提出了在彩色图像中检测隐藏信息的方法，当嵌入 LSB 信息后，新的调色板会产生很多颜色相近的颜色对，该方法是以相近的颜色对与所有颜色对的比值 R 作为衡量是否有隐藏信息的判据。该方法不仅仅适用于 GIF、PNG 等调色板图像，而且还适用于真彩色图像，可以分析出嵌入信息的长度，但此方法不适用于独立颜色数较多的图像以及灰度图像，并且算法复杂度也很高；文献[4]还提出了 RS 检测方法，原始图像具有一定的规律，但是嵌入信息后有些规律被破坏，RS 检验方法就是利用了其中一种规律来检验；文献[5]提出了一种高阶统计量的检测模型，该方法通过建立原始图像的高阶统计量模型，检测该模型的偏差从而确定是否存在隐藏信息，构造的统计量有两种：一种是在多个方向和规格上的子频带的数字特征，包括均值、方差、斜率、峰度等；另一种是基于系数大小的最佳线性预报的错误率，这些统计量共同构成了特征向量。重复的对每一个子频带进行计算，在每一点估计线性预报，采用模式识别的 2 类 FLD 方法，可以确定模式，用来判断是否存在隐藏信息。以上介绍的这几种方法均为盲检测方法。

2 算法原理

2.1 分辨函数检验方法

灰度图像的 8 个位平面有以下两个特点：一是，在较重要的位平面中若存在随机性，在其相邻的次要位平面中的相应位置也存在随机性；二是，从位 7^{th} 平面到 0 位平面，随机性逐渐增加。如果第 k 个位平面嵌入信息，第 k 个位平面到第 $k-1$ 个位平面的随机度的变化是很突然的，则改变这个特征。利用这两个特点，可以判断是否在一个图像中嵌入了信息。

该方法定义了一个位于 2 值图像中的矩形滑窗 W，作用在该滑窗上的转移密度定义为 TD。对于图像中的每一个点 P，以 P 为中心选择 W 滑窗，计算 TD 值，然后计算 TD 的平均值。

设图像 W 是一个 $w\times w$ 的 2 值图像，定义

$$TD=\sum_{(i,j)\in W}|x(i,j)-x(i,j+1)|+\sum_{(i,j)\in W}|x(i,j)-x(i+1,j)|$$

其中，$x(i,j)$是(i,j)的像素值。

对于给定的 w，TD 的取值范围为 $TD\in[0,2w(w-1)]$。若图像中的两种颜色分布是随机的，则 TD 值大约为 $w(w-1)$。当某个位平面的 TD 值与其相邻的位平面的值相差很大，或其 TD 值接近于 $w(w-1)$，就说明在此位平面中有可能隐藏了均匀分布的信息。

2.2 χ^2 检验

χ^2 检验是一种统计攻击的方法。统计攻击的思想就是把隐秘图像中像素的理论频率n_i'（即隐藏信息后索引值 i 应该出现的频率）和从载体中观测到的样本频率 n_i 进行比较，从而找出差异，检测是否有信息嵌入。因为进行的是盲检测，没有原始载体作为比较，因此统计攻击的关键是如何获得理论频率n_i'。

在最低位平面嵌入信息，有可能改变的只能是最不重要位。因此当嵌入均匀分布信息后，索引值为奇数和偶数出现的频率在理论上应该是一样的。因此索引值为偶数的像素点出现的理论频率n_{2i}' 应该是 $n_{2i}'=\frac{n_{2i}+n_{2i+1}}{2}$。$\chi^2$ 检验通过构造统计量函数 $X=\sum_{i=0}^{k}\frac{(n_{2i}-n_{2i}')^2}{n_{2i}'}$，利用样本计算 X 的值 x。从而计算衡量是否有信息嵌入的概率值 p

$$p=1-\frac{1}{2^k\Gamma\left(\frac{k}{2}\right)}\int_0^x e^{-\frac{x}{2}}x^{\frac{k}{2}-1}dx$$

检测图像中是否嵌入了信息，并且可以检测出信息的长度以及信息嵌入的位置。若有均匀分布的信息嵌入，n_{2i}'与 n_{2i}非常接近，因此 x 非常小，则概率 p 接近于 1。反之若概率 p 接近于 1，说明 x 值非常小，即

n'_{2i}与 n_{2i}非常接近，从而检验出有信息嵌入。若 p 的值非常小，甚至接近于 0，则得出相反的结论。

2.3 *RS* 检验方法

利用 LSB 方法嵌入信息后，会增加噪声。构造衡量噪声的函数 $f(x_1,\cdots,x_n)=\sum_{i=1}^{n-1}|x_{i+1}-x_i|$，满足 $f(x_1,\cdots,x_n)\in\mathbf{R}$，若图像为 8-bit 灰度图，像素的索引值 $x_i\in\boldsymbol{P}$，其中 $\boldsymbol{P}=\{0,1,2,\cdots,255\}$，嵌入信息后函数 $f(x_1,\cdots,x_n)$的值会增加。RS 检测方法定义了一个二轮置换函数，即 $F^2(x)=F(F(x))=x,x\in\boldsymbol{P}$。这里定义：$F_1:0\leftrightarrow1,2\leftrightarrow3,\cdots,254\leftrightarrow255$；$F_{-1}:-1\leftrightarrow0,1\leftrightarrow2,\cdots,255\leftrightarrow256$；$F_0$ 为单位置换，即 $F_0(x)=x$。利用函数 f 和置换 F 定义 $\boldsymbol{R}$，$\boldsymbol{S}$ 和 $\boldsymbol{U}$ 3 类像素组：

$$\boldsymbol{R}:G\in\boldsymbol{R}\Leftrightarrow f(F(G))>f(G);$$
$$\boldsymbol{S}:G\Leftrightarrow\in\boldsymbol{S}\Leftrightarrow f(F(G))<f(G);$$
$$\boldsymbol{U}:G\in\boldsymbol{U}\Leftrightarrow f(F(G))=f(G)$$

为了使 G 中不同的像素随机的应用不同的置换，引入了掩码，掩码也是 n 元数组，每个元素可以为 -1，0，1 中的任意一个。在引入掩码的情形下定义置换 $F(G)$为$(F_{M(1)}(x_1),\cdots,F_{M(n)}(x_n))$。在这里把掩码为 M 的 $\boldsymbol{R}$ 的个数定义为 R_m，掩码为 M 的 $\boldsymbol{S}$ 的个数定义为 S_m，对于掩码 $-M$，相应的定义为 R_{-m}和 S_{-m}。

由置换 F_1 与 F_{-1}的定义可以得出对于任意的 x，都有 $F_{-1}(x)=F_1(x+1)-1$，所以有 $R_m\cong R_{-m}$和 $S_m\cong S_{-m}$成立。可是嵌入信息后，这两个方程将不再成立。方程不成立的原因至今也没有给出理论上的证明，只是由大量的实验得到。*RS* 检测方法就是利用上述方程来检测图像中是否有信息嵌入。

3 实验分析

实验所用的图像如图 1 所示为 Girl. bmp (128×128)和 Lena. bmp(128×128)。针对 Ezstego 和 Hide4PGP 两种算法进行了实验，并且信息都是嵌在 0 位平面。

(a) Girl. bmp(128×128)　　(b) Lena. bmp(128×128)

图 1　实验图像

3.1 利用 Ezstego 方法嵌入

3.1.1 分辨函数检验方法

在实验中选取 $w=5$。嵌入信息前 Girl. bmp 的 TD 平均值为(从 $7^{th}\sim0$ 位平面)：1.334 1，2.897 5，4.970 4，7.111 5，8.421 8，10.858 0，11.272 4，13.202 4。Lena. bmp 的平均值为(从 $7^{th}\sim0$ 位平面)：4.029 6，7.455 4，9.936 4，14.051 0，17.766 5，19.680 1，19.911 6，20.096 3。分析这两组数据可以得到从 $7^{th}\sim0$ 位平面的 TD 值的变化是缓慢的，并且逐渐增加。Lena 图的 0 和 1^{st}位平面几乎是随机的，因此隐藏在 Lena 图的 0 和 1^{st}位平面中的均匀分布信息很难被检测出来。

把长度为 8K 的均匀分布信息分别嵌到上面图像的 0 位平面中。嵌入信息后Girl. bmp的 TD 平均值为(从 $7^{th}\sim0$ 位平面){1.334 1，2.897 5，4.970 4，7.111 5，8.421 8，10.858 0，11.272 4，17.546 4}。Lena. bmp 图的 TD 平均值为(从 $7^{th}\sim0$ 位平面){4.029 6，7.455 4，9.936 4，14.051 0，17.766 5，19.680 1，19.911 6，20.137 7}。嵌入信息后Girl. bmp 的 0^{th}位平面的 TD 值与 1^{st}位平面的 TD 值之间相差很大，因此必然引起检测者的怀疑。但用分辨函数方法检测 Lena 图却会引起误报。

在 0 位平面嵌入 0K、2K、4K、6K、8K、10K、12K、14K、16K 的均匀分布信息后，0 位平面的 TD 平均值为{13.202 4，14.221 1，15.765 2，16.675 1，17.546 4，18.275 0，19.539 1，19.877 0，19.936 5}。这说明随着信息嵌入量的增加，0 位平面的 TD 平均值逐渐趋向 20(如图 2 所示)，表明最低位平面的随机度随嵌入信息量的增加而逐渐增加。

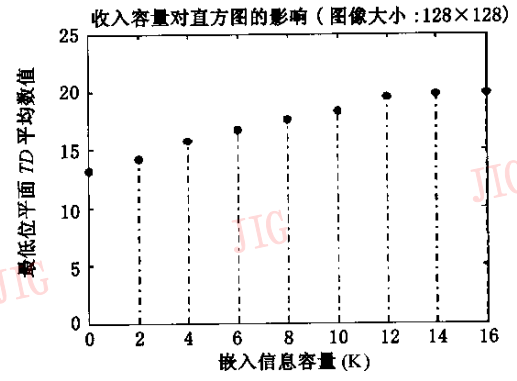
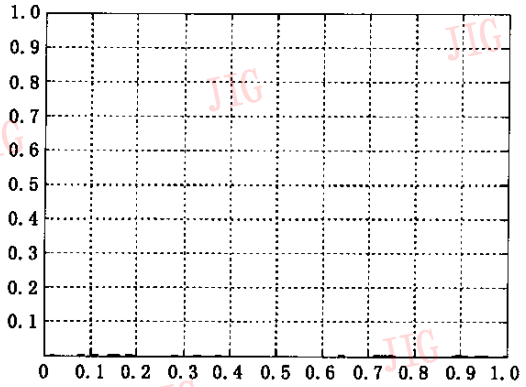


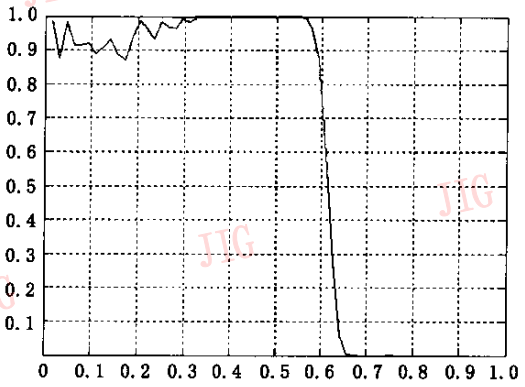
图 2　TD 值随嵌入信息量的变化曲线

3.1.2 χ^2 检验

利用 χ^2 检验得到 χ^2 曲线(图 3)，从图 3(a)的曲线可以看出图像 Girl. bmp 在嵌入信息前，p 几乎为



(a) 嵌入信息前



(b) 嵌入信息后

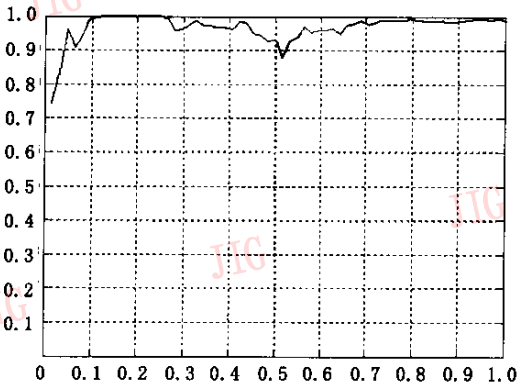
图 3 嵌入信息前后的检测结果(Girl. bmp)

(横坐标表示嵌入信息量与载体长度的百分比,纵坐标表示概率值 p)

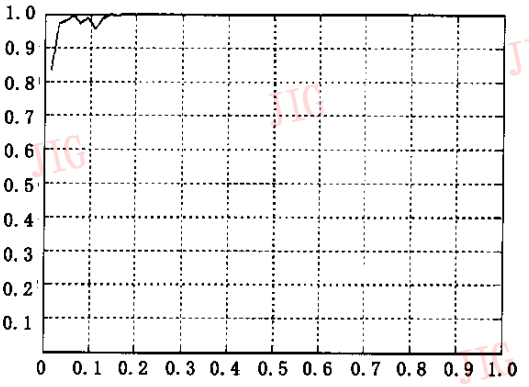
零。在图像前半部分嵌入 8K 均匀分布信息后,这时所得的 χ^2 检验曲线的前半部分的 p 接近于 1,并且大约在 50%的位置曲线突然下降,以后 p 几乎为零(如图 3(b))。说明图像中可能嵌入了信息,并且可以看出信息大约嵌在图像的前半部分。

但是 χ^2 检验并不是对于任何图像作为载体的情况都适用,例如当用 Lena 图像作为载体时,利用 χ^2 检验就会发生误报。Lena 图像在嵌入信息前、后的 χ^2 检验曲线如图 4 所示,嵌入信息前、后的 χ^2 检验曲线中的 p 值都接近于 1。虽然嵌入信息前、后的 χ^2 检验曲线存在区别,但进行的是盲检测,没有原始图像作为比较,因此遇到这种情况用 χ^2 方法检验往往就会发生误判,把没有隐藏信息的图像判断为隐藏了信息。

χ^2 检验的检测结果与利用分辨函数方法得到的结果一致。Girl. bmp 的 0^{th} 位平面的 TD 值为 13.202 4,即 0 位平面的随即性很小,因此索引值为奇数、偶数的像素点出现的频率相差很大,所以利用样本计算的 x 值较大,从而使得 p 值较小。当嵌入均匀分布的信息后,情况刚好相反。Lena 图像嵌入信息前后其位平面的 TD 值都接近于 20,因此嵌入信息前后的 x 值都较小,p 值都接近于 1。



(a) 嵌入信息前



(b) 嵌入信息后

图 4 嵌入信息前后的检测结果(Lena. bmp)

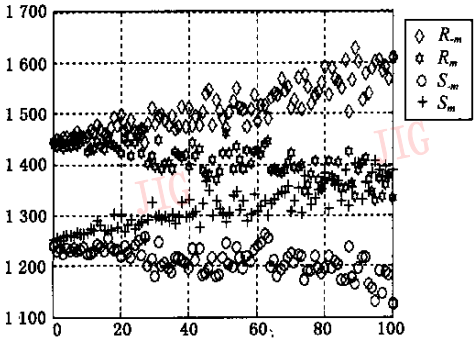
(横坐标表示嵌入信息量与载体长度的百分比,纵坐标表示概率值 p)

3.1.3 *RS* 检验方法

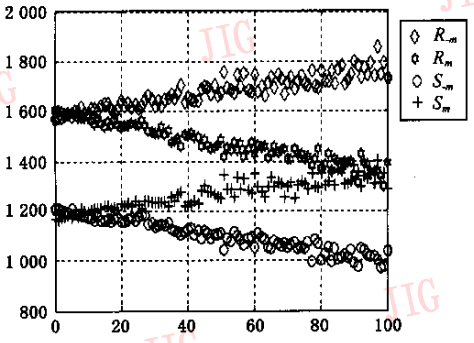
为了清楚描述出 R_m、R_{-m}、S_m 与 S_{-m} 的变化情况,图 5 给出了它们随嵌入信息量的改变而变化的曲线。没有嵌入信息时,Girl. bmp 图像的 $R_m=1\,445$,$R_{-m}=1\,444$,$S_m=1\,252$,$S_{-m}=1\,239$,满足 $R_m \cong R_{-m}$,$S_m \cong S_{-m}$,与图 5(a)中横坐标为 0 的点相对应,因此判断没有嵌入信息。在图像的前半部分嵌入信息后,得到 $R_m=1\,371$,$R_{-m}=1\,511$,$S_m=1\,318$,$S_{-m}=1\,172$,几乎与图 5(a)中的横坐标为 50 的点相对应。对于 Lena 图像就存在着误差,嵌入信息前,

R_m 与 R_{-m},S_m 与 S_{-m}交点的横坐标不是零,但由图 5(b)可以看出误差并不大,当信息长度占载体图像的 5%以上时就可以检测到有信息嵌入。

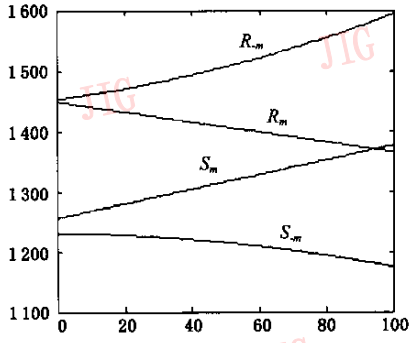
经过大量的试验证明,R_{-m}和 S_{-m}的曲线可以用直线很好地拟合,二次多项式也可以很合理地拟合 R_m 和 S_m 的曲线。由拟合曲线可以更加清晰地看到随着嵌入信息量的增加,R_{-m}与 S_m 单调增,R_m 与 S_{-m}单调减,并且 R_m 与 S_m 必能相交,利用这几点还可以计算出嵌入信息的长度。



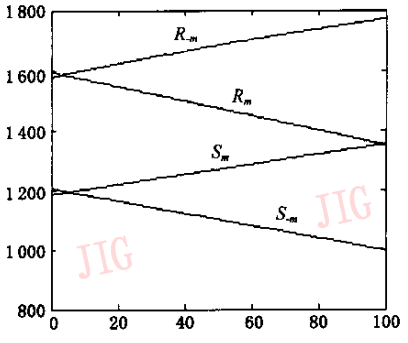
(a) *RS* 曲线(Girl. bmp)



(b) *RS* 曲线(Lena. bmp)



(c) *RS* 曲线的拟合曲线(Girl. bmp)



(d) *RS* 曲线的拟合曲线(Lena. bmp)

图 5　*RS* 曲线及其拟合曲线

(横坐标表示嵌入信息量占载体长度的百分数,纵坐标表示 R_m、R_{-m}、S_m 与 S_{-m}的值)

3.2　利用 Hide4PGP 方法嵌入

3.2.1　分辨函数检验方法

嵌入 8K 信息后 Girl. bmp 所对应的 *TD* 平均值为{1.334 1,2.897 5,4.970 4,7.111 5,8.421 8,10.858 0,11.272 4,14.122 4}。这时 *TD* 值缓慢增加,并且 0 位平面的 *TD* 值与 20 相差很大,因此检测不出图像中隐藏了信息。这是因为随机度不集中,因此值的增加不明显,所以对于利用离散方法嵌入的信息,分辨函数方法失效。

3.2.2　χ^2 检验

χ^2 检测方法检测时利用连续嵌入算法嵌入的信息效果很好,但对于离散嵌入算法,漏报率就会提高。例如对于图 Girl. bmp(128×128),用 Hide4PGP 分别嵌入信息长度为 4K、8K、12K、15K 的均匀分布信息。由图 6 中的 χ^2 检验曲线可以看出,只有当信息长度占图像的 75%以上时,才有可能检测到有信息嵌入。

3.2.3　*RS* 检验方法

(1) 计算 R,S 值

嵌入信息前 Gril. bmp 图的各个值为

$$R_m=1\,445,R_{-m}=1\,444,S_m=1\,252,S_{-m}=1\,239,$$

$$R=\frac{R_m}{R_{-m}}=1.000\,7,S=\frac{S_m}{S_{-m}}=1.010\,5$$

利用 Hide4PGP 在 Girl. bmp 图的前半部分嵌入信息后,得到的值为

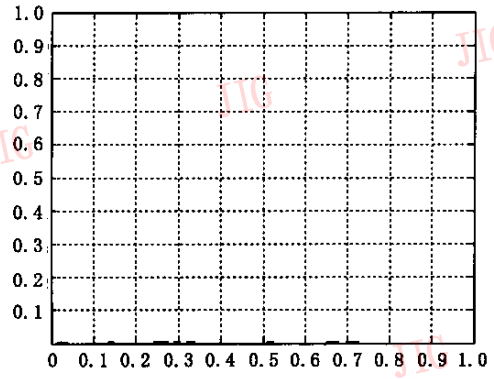
$$R_m=1\,438,R_{-m}=1\,603,S_m=1\,517,S_{-m}=1\,357,$$

$$R=\frac{R_m}{R_{-m}}=0.897\,1,S=\frac{S_m}{S_{-m}}=1.117\,1$$

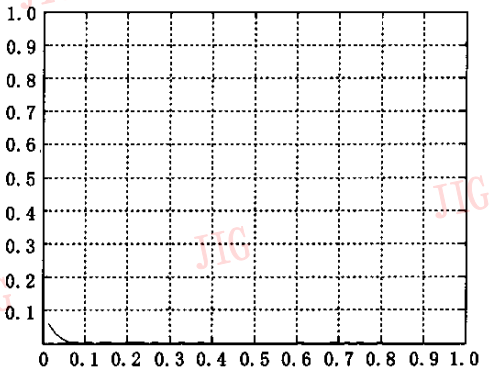
嵌入信息前 Lena 图的各个值为

$$R_m=1\,607,R_{-m}=1\,570,S_m=1\,166,S_{-m}=1\,211,$$

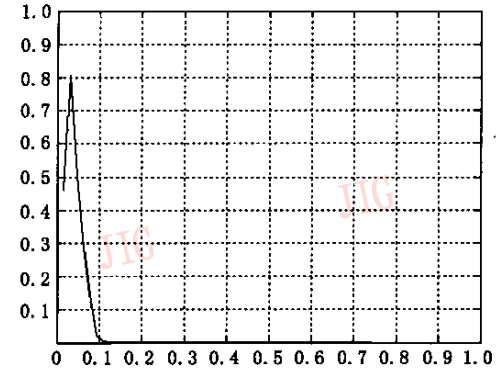
$$R=\frac{R_m}{R_{-m}}=1.023\,6,S=\frac{S_m}{S_{-m}}=0.962\,8$$



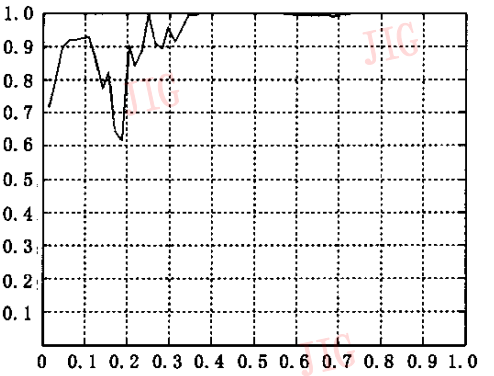
(a) 嵌入 4k 信息



(b) 嵌入 8k 信息



(c) 嵌入 12k 信息



(d) 嵌入 15k 信息

图 6 利用 Hide4PGP 嵌入信息后的曲线

(横坐标表示嵌入信息量与载体长度的百分比,纵坐标表示概率值 p)

利用 Hide4PGP 在 Lena 图的前半部分嵌入信息后,得到的值为

$$R_m=1\,497, R_{-m}=1\,685, S_m=1\,250, S_{-m}=1\,125,$$

$$R=\frac{R_m}{R_{-m}}=0.914\,3, S=\frac{S_m}{S_{-m}}=1.091\,1$$

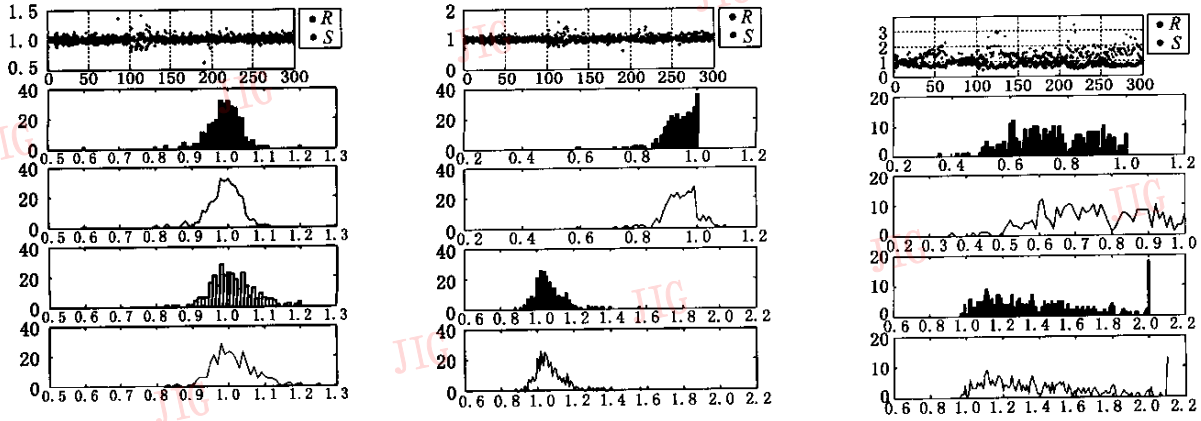
对于 Girl.bmp 和 Lena.bmp 嵌入信息前几乎都满足 $R\cong 1, S\cong 1$,只是 Lena 图的误差大一些。嵌入信息后 R,S 与 1 的距离增大,因此可以检测出嵌入了信息。当嵌入信息量较小时,效果就不会这样显著,因此需要选择置信区间。

(2) 置信区间的选取

利用 300 幅 128×128 的灰度图像做了实验,分别对这 300 幅图利用 Hide4PGP 嵌入 1K,2K,4K,8K 信息。并且绘制出嵌入信息前、嵌入 1K,2K,4K,8K 信息后 R、S 的直方图和直方图的拟合曲线如图 7 所示。图 7(a)显示的是没有嵌入信息时的情况,其中,R、S 的值近似服从正态分布,其正态分布的峰值很接近于 1。计算其均值 $\overline{R}=0.991\,5, \overline{S}=1.013\,5$,与 1 很接近。嵌入 1K,2K 信息后,这时 R,S 的值仍可以认为近似服从正态分布,但其峰值与 1 的距离增大,嵌入 1K 时 $\overline{R}=0.941\,4, \overline{S}=1.064\,7$;嵌入 2K 时 $\overline{R}=0.901\,5, \overline{S}=1.114\,8$。当嵌入 4K,8K 时,$R,S$ 不再近似服从正态分布,嵌入 4K 时,$\overline{R}=0.836\,2$, $\overline{S}=1.214\,0$;嵌入 8K 时,$\overline{R}=0.749\,9, \overline{S}=1.418\,3$。

随着嵌入信息长度的增加,R 的值减小,并且向小于 1 的一侧移动;而 S 的值增大,并且向大于 1 的一侧移动。这与用 Ezstego 嵌入方法的结果一致,因为随着嵌入信息长度的增加 R_{-m} 与 S_m 单调增,R_m 与 S_{-m} 单调减,所以 R 的值减小,而 S 的值增大。

由图 7(a)可以看出嵌入信息前 R,S 值的直方图几乎关于 $R=1$ 和 $S=1$ 对称。因此可以取一个以均值为中心的对称区间作为置信区间。如果图像的 R,S 值都落在该区间内,则认为其没有信息嵌入,否则就认为嵌入了信息。经过分析计算选 R 的置信区间为 (0.951 5, 1.031 5),S 的置信区间为 (0.963 5, 1.063 5)。没有嵌入信息时,误报率为 28.33%。这时误报率虽然比较高,但是漏报率相对较小,当嵌入信息长度为 1K 时,漏报率为 21.34%,



(a) 嵌入信息前　　(b) 嵌入 1K 信息　　(c) 嵌入 8K 信息

图 7　RS 的直方图及曲线

当嵌入信息长度为 4K 时,漏报率为 7.27%,当嵌入信息长度为 8K 时,漏报率仅为 3.67%。如果还需降低漏报率,可以减小置信区间的长度,但这时必然会增大误报率。

还可以利用概率统计中求置信区间的方法求置信区间。嵌入信息前,$\overline{R}$,$\overline{S}$ 近似服从正态分布,并且方差未知。因此构造统计量 T

$$T=\frac{\overline{X}-u}{S}\sqrt{n}\sim t(n-1)$$

即统计量 T 服从自由度为 $n-1$ 的 t 分布。其中,$\overline{X}$ 为样本均值,n 为样本点个数,S 为样本方差。选择置信度为 α,经过计算可得置信区间的上、下限分别为

$$\hat{\theta}_1=\overline{X}-t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}$$

$$\hat{\theta}_2=\overline{X}+t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}$$

其中,$t_{\alpha/2}(n-1)$是自由度为 $n-1$ 的 t 分布的上 $\alpha/2$ 分位数。

如果取 $\alpha=0.05$,R 的置信区间为(0.973 1, 1.014 3),S 的置信区间为(0.980 3,1.034 7)。这时当嵌入信息长度为 1K 时,漏报率为 5.23%,当嵌入信息长度为 4K 时,漏报率为 4.17%,当嵌入信息长度为 8K 时,漏报率为 1.33%。虽然漏报率减小了,但误报率却增加了很多。漏报率与误报率本身就是一对矛盾,这就要根据自己的需要来选择。一般情况下希望在误报率合理的情况下,尽量降低漏报率。

4　算法比较与分析

4.1　适用范围

(1) 嵌入信息

分辨函数检验方法及 χ^2 检测方法检测均匀分布的信息效果较好,若嵌入的信息不服从均匀分布,就会影响检测效果,甚至检测不出隐藏了的信息。RS 检验方法对信息分布没有特殊的要求,因为只要有信息嵌入就增加了噪声,破坏了方程 $R_m\cong R_{-m}$ 和 $S_m\cong S_{-m}$。分辨函数检验方法还要求嵌入信息的容量比较大。

(2) 嵌入算法

分辨函数检验方法及 χ^2 检测方法适应于检验利用连续嵌入方法嵌入的信息。因为利用连续嵌入方法嵌入的信息,可以使随机性比较集中,从而使 TD 值增加比较明显,利用分辨函数检验方法才比较有效。对于 χ^2 检测方法,从理论上来讲,嵌入均匀分布信息后索引值为偶数的像素点出现的频率是 $n'_{2i}=\frac{n_{2i}+n_{2i+1}}{2}$,但离散嵌入算法是把信息随机的分散在整个载体,因此若嵌入的信息长度比较短,则使上述特点不再集中,则统计量 $X=\sum_{i=0}^{k}\frac{(n_{2i}-n'_{2i})^2}{n'_{2i}}$ 的值就会增大,p 值就会减小,因此漏报率就会增加。当信息长度与载体的长度接近时,用 χ^2 检验方法检测就可以检测出隐藏了信息。RS 检验方法对嵌入算法没有严格要求。

(3) 载体

若把 0 位平面随机性很强的图像作为载体,则 0 位平面 TD 的值接近于 20,利用分辨函数检验方法必会发生误报。在这种情况下,统计量 $X=\sum_{i=0}^{k}\frac{(n_{2i}-n'_{2i})^2}{n'_{2i}}$ 的值很小,p 值接近于 1,因此利用 χ^2 检

测方法就会误认为嵌入了信息。*RS* 检验对载体没有特别的要求，但对于随机性很强的载体会存在一定的误差。

4.2 实验效果

分辨函数方法不能明显地检测出图像是否隐藏了信息，但其可以用于预检测，若发现图像某一个位平面的 TD 值接近于 20，或两个位平面的 TD 值相差很大，就把其列入可疑范围用其他分析方法再进一步分析。χ^2 检验是一种很有效的统计方法，利用此种方法可以在没有原始载体的情况下检测出图像中是否隐藏了信息，并且可以判断出信息的长度以及嵌入的位置。但是此方法也有一定的局限性。它能否检测出隐藏了信息，与所用的原始载体和所用的嵌入算法密切相关。目前 *RS* 检测法是检测范围比较广的一种检测方法，它可以有效地检测出应用连续嵌入方法或离散算法嵌入的信息，并且对于离散算法的检测效果更佳。

现有的检测方法中各有优点与局限性，因此只用一种方法检测不能保证检测结果一定正确。但可以对一幅图像综合应用各种检测方法，对所有的检测结果进行分析，最终给出结果。这样无疑会提高检测的准确率。

5 结 论

隐写术是信息安全领域的一个重要方面，隐写术分析作为隐写术的对立面，已成为信息安全领域研究的焦点。图像作为信息隐藏的良好载体，已经在很多方面得到了应用。目前虽然已经提出了多个基于图像的信息隐藏检测的算法和系统，但每种算法或系统都有着各自的优势和局限性，至今仍然还没有形成一套完整的理论与系统。因此，基于图像的隐写术分析技术还有待进一步发展和完善。

参 考 文 献

1 Lee Yeuan-Kuen, Chen Ling-Hwei. An adaptive image steganographic model based on minimum-error LSB replacement [A]. In: Proceedings of the Ninth National Conference on Information Security[C], Taiwan, China, 1999:8～15.

2 Westfeld A, Pfitzmann A. Attacks on steganographic systems [A]. In: Proceedings of Third international Workshop, IH'99 [C], Dresden, Germany, Springer Verlag, 1999:61～75.

3 Fridrich J, Du R, Meng L. Steganalysis of LSB encoding in color images [A]. In: 2000 IEEE international conference on multimedia and Expo[C], New York City, 2000, **3**:1279～1282.

4 Fridrich Jessica, Goljan Miroslav, Du Rui. Detecting LSB steganography in color and gray-scale images [J]. IEEE Tranactions on Multimedia, 2001, **8**(4):22～28.

5 Hary Farid. Detecting hidden messages using higher-order statistical models [A]. In: Proceedings IEEE international conference on image processing[C], Rochester, NY, 2002:23～27.

6 Provos N, Honeyman P. Detecting steganographic content on the internet [R]. Technical Report CITI 01-1a, University of Michigan, U. S., 2001.

郎荣玲 1975 年生，2001 年获西北工业大学应用数学专业硕士学位，现在西北工业大学自动控制系攻读博士学位。主要研究方向为信息安全、计算机密码学、运筹学。

夏 煜 1975 年生，2000 年获西北工业大学自动控制专业硕士学位，现在西北工业大学自动控制系攻读博士学位。主要研究方向为计算机控制与智能控制、信息安全。

郅 艳 1975 年生，2000 年获西北工业大学自动控制专业硕士学位，现在西北工业大学自动控制系攻读博士学位。主要研究方向为信息安全。

戴冠中 1937 年生，博士生导师，教授。主要研究方向为自动控制、信息安全。