空域信息隐藏算法隐写分析

喻鹏 401030918057

June 7, 2019

摘要

文章研究了 RS 隐写分析算法、卡方隐写分析算法、信息量估计法、GPC 分析法和基于相邻像素统计特性的隐写算法。通过进行大量的隐写测试实验表明，上述五种算法均能对 LSB 算法进行有效检测。

关键词：信息隐藏；隐写分析算法；LSB 算法

1 引言

信息隐藏技术是信息安全的重要组成部分，这种技术通过某种嵌入方式将秘密信息嵌入载体而不被发现，接收者可以根据事先约定的规则提取秘密信息。相对应地，隐写分析技术，是攻击者尝试分析出数字图像载体中有没有嵌入含有秘密信、嵌入率是多少以及嵌入规则如何的一种技术。遗憾的是，往往只能以概率的形式去估计嵌入率的多少，而对于提取秘密信息则无能为力，难度很大。

按嵌入的角度来说，隐写分析技术可以分为基于空域的和频域的技术。空域的方法主要用于检测基于时空域的信息隐藏算法，而频域的方法主要是用以分析基于傅里叶变换、小波变换和余弦变换等等频域变换的信息隐藏算法。本文主要介绍空域信息隐藏隐写算法。

针对 LSB(least significant bit) 算法的隐写分析也进行得比较早。Westfeld 等人在文献 [1] 中提出卡方隐写分析算法，根据像素对服从卡方分布以此建立卡方统计量，进而来分析秘密信息是否存在并计算嵌入率；Fridrich 等人在文献 [2-3] 提出了 RS(Regular Singular) 算法，利用空间块之间存在的相似性进而进行隐写分析，该算法适用性更广，能分析秘密信息的有无和嵌入率的多少；XinpengZhang 在文献 [4] 中提出 GPC(Gray-Level Plane Crossing) 算法，通过利用相邻像素之间的相关性进行隐写分析，遗憾的是不能进行嵌入率的分析;XIa 等人在文献 [5] 中提出信息量估计法，通过计算负载率 α 来计算嵌入率；同样基于相邻像素的统计特性，秦姣华等人在文献 [6] 提出一种新的估计嵌入率的方法, 基于相邻像素统计特性的隐写算法 (Oddity Or Evenness Of the Gray-Scale Values Between Adjacent Pixels,OEP)，取得了比较好的结果；张涛等人在文献 [7] 中提出 DIH(Difference Image Histogram) 算法，在一定的条件下效果比 RS 更好；Sorina Dumitrescu 等人在文献 [8] 提出的 SPA(Sample Pair Analysis) 分析方法，通过利用信号的空间相关性进行隐写分析，取得了比较高的精度。

2 五种隐写分析算法原理

时空域下 LSB 算法是最简单、最经典的隐写分析算法，对它的研究也非常透彻。很多商用软件就是采用该算法来嵌入秘密信息的，相应地，同时针对该算法的隐写分析技术也得到飞速的发展，很多隐写算法就是在这种背景下诞生。下面介绍这五种算法。

2.1 RS 隐写分析算法

Fridrich 在他的文献 [2-3] 中指出，经过统计分析，嵌入秘密信息的载体图像，最低有效位的值 (0 和 1) 服从概率值为 0.5 的均匀分布，而没有嵌入秘密信息的图像则不会存在上述规律。

现在假设载体图像是 8 位的灰度图像，它的大小是 M*N，用 P 表示图像的像素值，$P = \{0, 1, 2, \cdots, 255\}$。首先用 ZigZag 方式对图像分块,生成一个含有 n 元素的向量组 $G = \{x_1, x_2, \cdots, x_n\}, x_i \in P$。嵌入信息后的图像就相当于添加了一定的噪声，这时可以构造一个衡量噪声的函数：

$$f(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n-1} |x_{i+1} - x_i|$$

对于这个定义在图像块 G 上的噪声函数 f，有 $f(x_1, x_2, \cdots, x_n) \in R$，这个噪声函数 f 的性质是，相邻像素之间的差距越大，起伏越大，那么 f 的值也越大, 说明图像块像素之间的相关性也越小，相应地，如果图像相邻像素之间比较平稳，f 就比较小，图像块空间相关性就比较大。

这时定义三个翻转函数：F_0、F_1 和 F_{-1}，即不变函数、交换函数和偏移函数，其中有：

$$F_0(x) = x, x \in P$$

$$F_1(x) : 2n \leftrightarrow 2n+1 \Leftrightarrow 0 \leftrightarrow 1, 2 \leftrightarrow 3, \cdots, 254 \leftrightarrow 255$$

$$F_{-1}(x) : 2n \leftrightarrow 2n-1 \Leftrightarrow -1 \leftrightarrow 0, 1 \leftrightarrow 2, \cdots, 255 \leftrightarrow 256$$

对于其中的 F_0 和 F_1，称之为非负翻转，对 F_0 和 F_{-1} 则称之为非正翻转。对像素块 G 应用 F 翻转，就会得到 3 种不同的像素组 R、S 和 U:

$$\begin{aligned} Regular \quad & Group: G \in R \quad if \quad f(F(G)) > f(G) \\ Singular \quad & Group: G \in S \quad if \quad f(F(G)) < f(G) \\ Unusable \quad & Group: G \in U \quad if \quad f(F(G)) = f(G) \end{aligned}$$

进行翻转后，若图像块 G 相关函数 f 值增大，说明空间相关性减弱，像素起伏程度增加，那么称该图像块是正常的 (Regular)，反之，称该图像块是异常的 (Singular)，如果没有变化，则称之为无用的 (Unusable)。RS 隐写分析算法也就是这么来的。

有了上述函数 f 和 F 翻转操作之后，就可以针对 LSB 算法来进行分析了。LSB 算法实质就是对于原始图像进行 2 种翻转，秘密信息和最低有效位相同就应用 F_0 翻转，如果不同，即最低有效位为 0 而秘密信息为 1 或者最低有效位为 1 而秘密信息为 0 就进行 F_1 翻转，需要注意的是不进行 F_{-1} 翻转。

2.1.1 嵌入有无的判定

上面是 RS 隐写分析需要用到的一些基础理论。RS 隐写分析的思想是，引入掩码 M，$M \in \{-1, 0, 1\}$ 对应于 3 个翻转函数, 这里主要用到-M 和 +M；接下来对 G 进行-M 和 +M 置换, 对于 +M 置换之后得到 R 组和 S 组概率 R(p/2) 和 S(P/2)，-M 置换会得到 R(1-p/2) 和 S(1-p/2)。那么，对图像块 G 进行翻转的结果就为 $F(G) = \{F_{M(x_1)}, F_{M(x_2)}, \cdots, F_{M(x_n)}\}$，$M(x_i), i \in \{1, 2, \cdots, n\}$ 是-1,0,1 中的任何一个。例如，令 M 为 $\{1, 0, 1, 0\}$，G 为 $\{23, 40, 88, 79\}$，操作之后的结果就是 $\{22, 40, 89, 79\}$。

对每个图像块应用非负翻转，计算像素起伏程度增加的图像块的比例，记为 R_M, 类似地记为 S_M；同理应用非正翻转，则得到 R_{-M}, S_{-M}。

一般来说，对图像应用 3 种置换后，会有

$$R_M + S_M \leq 1,$$

$$R_{-M} + S_{-M} \leq 1.$$

对于自然图像，从统计上说，非负翻转或非正翻转会同等程度增加图像块的混乱程度, 此时会有

$$R_M \approx R_{-M},$$

$$S_M \approx S_{-M}.$$

而含有秘密信息的图像，翻转会破坏图像块的空间相关性，一般情况下有

$$R_M > S_M,$$

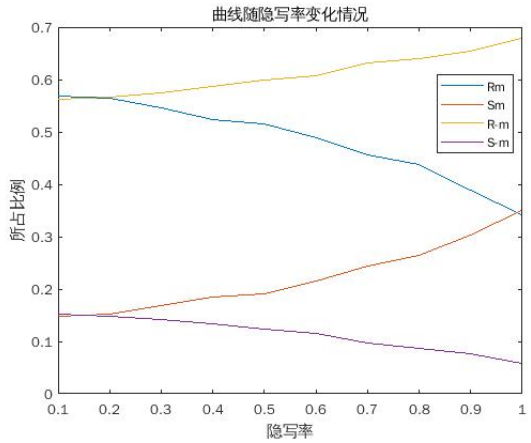
$$R_{-M} > S_{-M}.$$

所以，对待检测图像，进行非负翻转和非正翻转，计算 R_M、S_M 、R_{-M} 和 S_{-M}，如果 $R_{-M} - S_{-M}$ 显著大于 $R_M - S_M$，则认为图像经过隐写。自然图像所得结果就如 1(a)中嵌入率为 0 时所示。

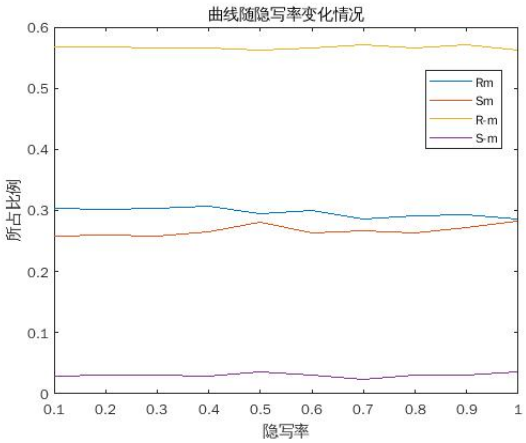
2.1.2 嵌入率的计算

对于 LSB 隐写图像，由于嵌入只采用非负翻转，进行非正翻转之后图像块翻转的变化会比较大，所以采用非负翻转和非正翻转的结果有明显不同。设原图隐写率为 α，根据大量的统计数据，图像中有 $\alpha/2$ 的像素应用了 F_1 翻转，也有 $\alpha/2$ 比例的像素应用 F_0 翻转，不难得知有比例为 $(1 - \alpha/2)$ 的像素未翻转。下一步操作是对 LSB 隐写图像应用非负翻转，设其中 F_1 翻转的比例为 β，F_0 翻转就为 $1 - \beta$ 则非负翻转后有三类像素：

- 没有被翻转，灰度值未变，所占比例为：$(1 - \alpha/2)(1 - \beta)$;
- 经历一次翻转，灰度值变化 1，所占比例为：$(1 - \alpha/2)\beta + \alpha/2(1 - \beta) = \alpha/2 + \beta - \alpha\beta$;



(a) pic1. 未嵌入信息



(b) pic2. 嵌入秘密信息的图片

- 经历二次翻转，灰度值回到原始值，所占比例为：$\alpha\beta/2$.

相当于在原图像上有 $\alpha/2+\beta-\alpha\beta$ 像素被 F_1 翻转，即比隐写图像增加 $(1-\alpha)\beta$ 的像素被翻转 $(1-\alpha)\beta$ 随 α 增大而减小，这表明：R_M 与 S_M 的差距随 α 增大而减小；当 $\alpha=1$ 时，R_M 与 S_M 近似相等。图 1(a)中嵌入率为 100% 时就验证了这一点，图 1(b)是嵌入率为 100% 的图像四条曲线变化情况，也符合上述理论。

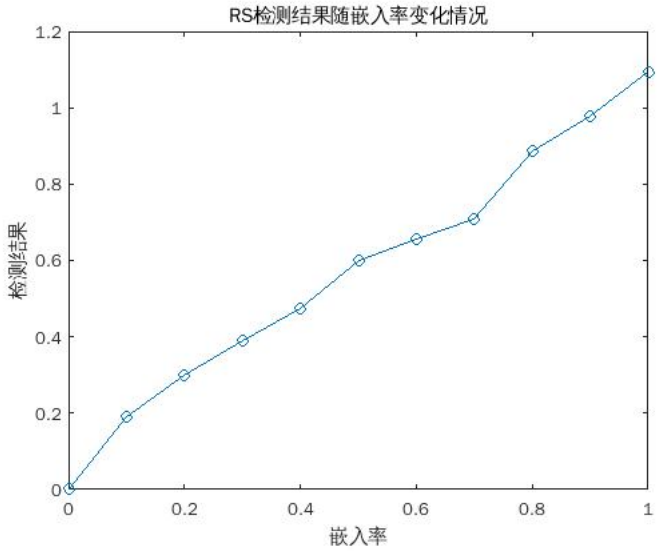


Figure 1: RS 隐写分析

要检测图像的隐写率，类似地，在不同隐写率的条件下计算 R_M、S_M 、R_{-M} 和 S_{-M}。得到的结果如图 2.1.1所示。设隐写率为 p，这个时候会得到 8 组数据，分别是 $R_M(p/2)$,$S_M(p/2)$,$R_M(1-p/2)$,$S_M(1-p/2)$ 以及 $R_{-M}(p/2)$,$S_{-M}(p/2)$,$R_{-M}(1-p/2)$,$S_{-M}(1-p/2)$。根据文献 [3]，通过大量的实验发现，如 1(a)所示，R_{-M} 和 S_{-M} 和嵌入率 α 是呈现出一种线性的关系的，R_M 和 S_M 和嵌入率 α 呈二次曲线关系。

令

$$\begin{aligned} d_0 &= R_M(p/2) - S_M(p/2) \\ d_1 &= R_M(1-p/2) - S_M(1-p/2) \\ d_{-0} &= R_M(p/2) - S_{-M}(p/2) \\ d\_1 &= R_M(1-p/2) - S_{-M}(1-p/2) \\ a &= 2(d_0 + d_1) \\ b &= d_{-0} - d_1 - d_{-1} - 3d_0 \\ c &= d_0 - d_{-0} \end{aligned}$$

构建一个二次方程 $ax^2 + bx + c = 0$，求解该二次方程 $ax^2 + bx + c = 0$，取其较小的解，那么嵌入率就为:

$$p = \frac{x}{x - 1/2}.$$

图 1是对 Lena 图像在不同嵌入率下的分析结果，算法 1给出了 RS 隐写分析算法的的描述。

算法 1: RS 隐写分析算法

输入：待检测图像；

计算 $R_M(p/2)$，$S_M(p/2)$,$R_M(1-p/2)$,$S_M(1-p/2)$；
if $R_M(p/2) \approx S_M(p/2)$ & $R_M(1-p/2) \approx S_M(1-p/2)$ **then**
    输出：没有嵌入秘密信息。
end
else
    将所有像素像素翻转；
    计算 $R_{-M}(p/2)$,$S_{-M}(p/2)$,$R_{-M}(1-p/2)$,$S_{-M}(1-p/2)$；
    求二次方程的根，取较小的一个解 x；
    输出：$p = x/(x - 0.5)$ 。
end

2.2 卡方隐写分析算法

卡方分析也是出现比较早的，但是不能用来估计隐写率。卡方隐写分析的思想是，比较待检测图像的分布和自然图像所期望的概率分布，差异比较大就认为嵌入过秘密信息。现在，设图像中灰度值为 j 的像素数为 h_j，其中 $0 \le j \le 255$，如果图像没有含有任何秘密信息，那么根据统计分析，h_{2i} 和 h_{2i+1} 的值则会相差得远一些。秘密信息在嵌入之前往往经过加密，可以看作是 0 和 1 随机分布的比特流，而且值为 0 与 1 的可能性都是 1/2, 如果秘密信息完全替代了载体图像的最低位，那么 h_{2i} 和 h_{2i+1} 的值会比较接近。于是，令

$$h^*_{2i+1} = \frac{h_{2i} + h_{2i+1}}{2}$$

Westfeld 等人在文献 [1] 中指出，像素只会在 2i 和 2i+1 之间变化，那么隐写前后 h^*_{2i+1} 是变化不大的，这时再构造一个统计量 q，

$$q = \frac{h_{2i} - h_{2i+1}}{2}$$

那么，当 h^*_{2i} 相当大的时候就会有

$$\frac{q}{\sqrt{h^*_{2i}}} = \frac{(h_{2i} - h_{2i+1})}{2\sqrt{h^*_{2i}}} = \frac{(h_{2i} - h^*_{2i})}{2\sqrt{h^*_{2i}}} \sim N(0,1)$$

构造一个统计量 r，它服从卡方分布

$$r = \frac{(h_{2i} - h^*_{2i+1})^2}{2}, \quad i \in \{0, \cdots, 127\}, r \sim \chi^2(127).$$

r 越小，隐写的概率就越大，结合卡方分布的概率密度函数，设 p 为嵌入秘密信息的可能性 (非隐写率)，按下列公式计算

$$p = 1 - \frac{1}{2^{\frac{k-1}{2}}\Gamma(\frac{k-1}{2})} \int_0^r e^{\frac{-t}{2}} t^{\frac{k-1}{2}-1} dt.$$

就得到了分析结果。下图 2是对 Lena 图像在不同嵌入率下进行卡方隐写测试的结果，算法 2 给出了卡方隐写分析算法的描述。

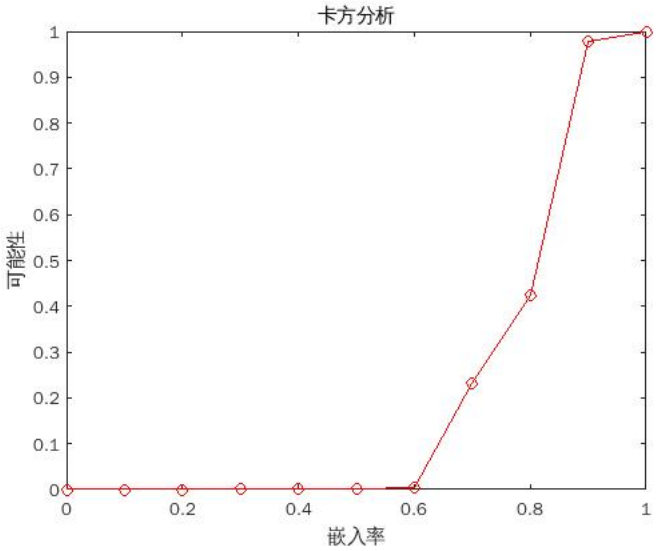


Figure 2: 卡方测试结果

算法 2: 卡方隐写分析算法

输入：待检测图像；

1 对图像进行统计，计算 h_{2i} 和 h^*_{2i}；

2 计算 r，并按公式计算 p；

输出：p 。

2.3 信息量估计法

最初信息量估计法 (Method Of Information Quantity Estimation，简称 IQE) 是用于音频隐写分析的，事实上也可用于图像隐写分析，它们的思想是相通的。假设载体图像嵌入率为 α，秘密信息嵌入信息的位置伪随机地分布于整个图像。事实上，如果载体图像嵌入率 α 接近于 0，可以认为没有经过隐写。设原始灰度图像中灰度值为 j 的像素数目为 f_j，密写后灰度值为 j 的像素数目为 h_j，文献 [5] 指出，秘密信息可以看作 0 和 1 等概率分布的随机码流，大约有个像素的灰度值由 $\alpha/2f_{2i}$ 个 2i 改为 2i +1, 也有大约有 $\alpha/2f_{2i+1}$ 个像素的灰度值由 2i +1 改为 2i。不难得知，

$$\begin{aligned} E(h_{2i}) &= f_{2i} - \alpha/2f_{2i} + \alpha/2f_{2i+1}, \\ E(h_{2i+1}) &= f_{2i+1} - \alpha/2f_{2i+1} + \alpha/2f_{2i}, \\ E(h_{2i+2}) &= f_{2i+2} - \alpha/2f_{2i+2} + \alpha/2f_{2i+3}. \end{aligned}$$

其中 $E(\cdot)$ 表示期望，那么上述两个公式相减，就会有

$$\begin{aligned} E(h_{2i+1} - h_{2i}) &= (1-\alpha)(f_{2i+1} - f_{2i}), \\ E(h_{2i+2} - h_{2i+1}) &= f_{2i+2} - f_{2i+1} + \frac{1}{2}\alpha(f_{2i+3} - f_{2i+2} + f_{2i+1} - f_{2i}). \end{aligned}$$

由此可知，h_{2i+1} 与 h_{2i} 之差与嵌入率有近似的线性关系，随着嵌入率的增大 h_{2i+1} 与 h_{2i} 趋近相等；而 h_{2i+2} 与 h_{2i+1} 则不一定随着嵌入率的增大而逐渐趋近。

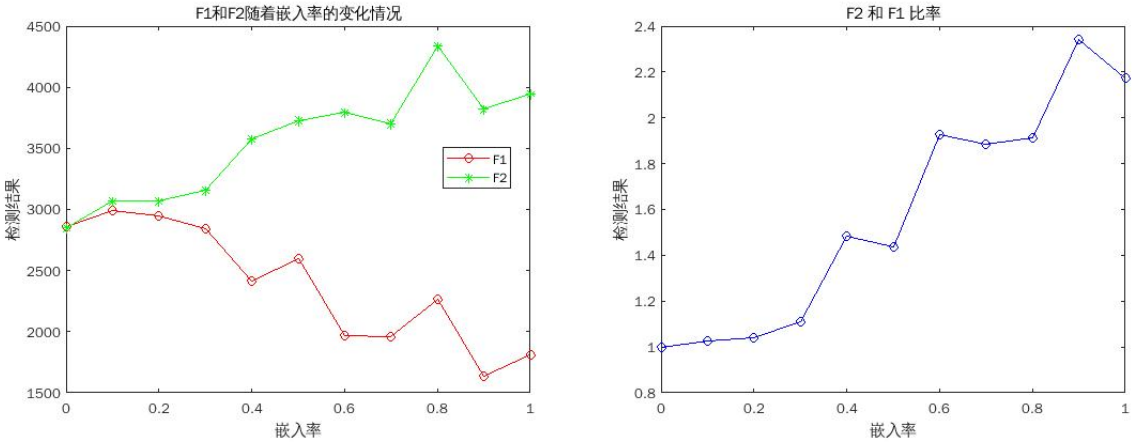
2.3.1 是否嵌入秘密信息的判断

有了上面的基础，就可以根据下面两个公式求出 F_1 和 F_2：

$$F_1 = |E(h_{2i}) - E(h_{2i+1})|$$

$$F_2 = |E(h_{2i+2}) - E(h_{2i+1})|$$

如果图像没有经过密写，F_1 和 F_2 在统计上并无不同，因此这两个参数的值应该很接近; 如果图像经过密写，F_1、F_2 与嵌入率近似呈线性关系，并且 F_1 随嵌入率增大而减小；而 F_2 不一定减小，即使下降，其速度也远小于 F_1 减小的速度。所得的结果如 3(a)和 3(b)所示，嵌入率为 0 的时候比值接近 1，而随着嵌入率的增大，比值越来越大，由此可以判断载体图像是否秘密信息。



(a) pic1.F_1 和 F_2

(b) pic2.F_1 和 F_2 比率 R

Figure 3: 信息量估计法

2.3.2 隐写率的估计

信息量估计法对密写嵌入率的估计，采用二次密写法。随机产生秘密信息，对载体图像最低有效位平面进行二次密写，嵌入率 β 逐步由 10%，20% 上升到 100%。对应于每一个 β 值分别计算 F_1 和 F_2。经二次密写后，有部分像素经历了两次改变又回到原始的值，而在两次密写中仅改变过一次的像素比例为 $(\alpha+\beta-\alpha\beta)/2$，即相当于经历了一次嵌入率为 $(\alpha+\beta-\alpha\beta)$ 的密写。也就是说，嵌入率为 β 的二次密写相当于将嵌入率提高了 $(1-\alpha)\beta$。因为二次密写线性地增加了嵌入率，由此得到的 F_1 和 F_2 会呈线出接近线性的变化。用二次密写的方法得到 F_1 ，F_2 曲线，此时对两条曲线做拟合，将新得到的 F_1 ，F_2 方向延长，得到交点 (x, y)，那么嵌入率的估计值为：

$$\alpha = \frac{|x|}{(|x|+1)}.$$

算法 3给出了信息量估计法的的描述。

算法 3: 信息量估计法

输入：待检测图像；

在不同嵌入率的对图像在进行二次密写，计算 F_1 和 F_2；
if $F_1/F_2 \leq T$ **then**
    输出：未嵌入信息。
end
else
    对得到的 F_1 和 F_2 的曲线做拟合求其交点；
    按公式计算嵌入率 α;
    输出嵌入率 α 。
end

2.4 GPC 分析法

考虑两个平行于 XY 平面的平面簇，平面簇 P_0 由 z=1.5, 3.5, 5.5,$\cdots$,255.5 组成，平面簇 P_1 由 z=0.5,2.5,4.5,$\cdots$,254.5 组成，记图像的三维曲面穿越平面簇 P_0 的次数为 N_0，图像的三维曲面穿越平

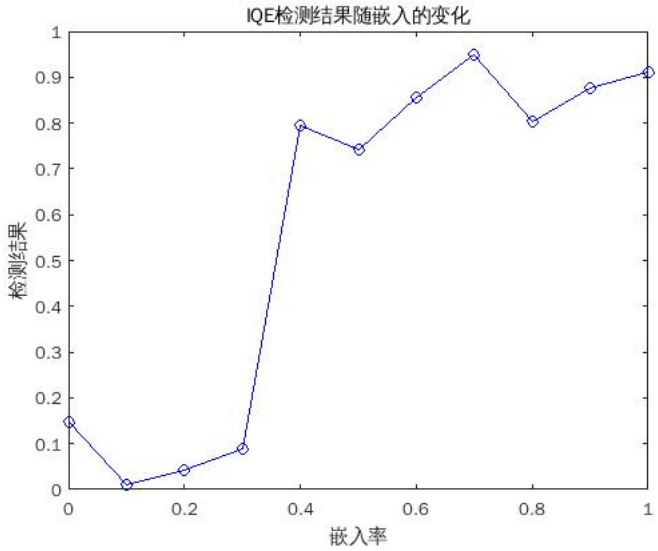


Figure 4: IQE 检测结果

面簇 P_1 的次数为 N_1。同样的，对于自然图像来说，N_0 近似等于 N_1，LSB 隐写图像，载体数据在 2i 和 2i+1 之间互变，不会穿越平面簇 P_0，但会穿越平面簇 P_1，N_0 不变，N_1 增大令 $R = R_1/R_2$，如果 R 大于某个阈值，则认为是隐写图像。事实上，该算法也可以估计隐写率。在待检测图像中进行

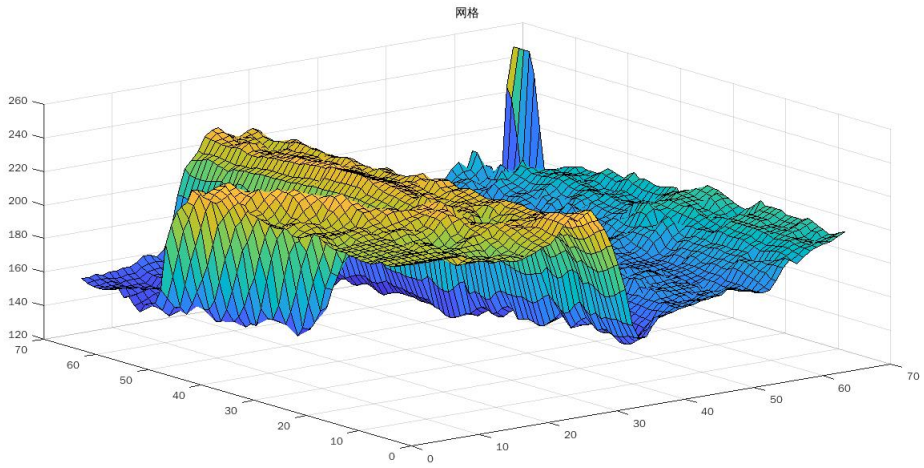


Figure 5: 网格图像

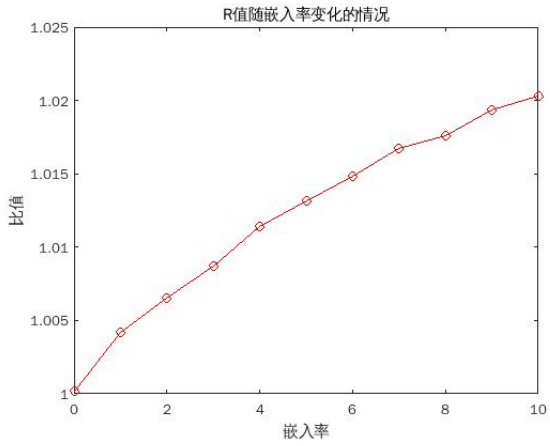
二次密写，嵌入率 β 为 100%。经二次密写后，有部分像素经历了两次改变又回到原始的灰度值，而在两次密写中仅改变过一次灰度值的像素占全部像素的 $(\alpha+\beta-\alpha\beta)/2$ ，即相当于经历了一次嵌入率为 $(\alpha+\beta-\alpha\beta)$ 的密写。也就是说，当二次嵌入率为 100% 时，仅经历过一次灰度值变化的像素正好占全部像素的 50%，也就相当于经历了一次嵌入率为 100% 的密写。因此，可以用二次全嵌入近似一次全嵌入时的情况。记待检测的图像初始的 R 值为 r_0, 二次随机隐写的 R 值为 r_1,α 为隐写率。接下来，建立如下方程组:

$$y_0 = b + c * 0$$
$$y_\alpha = b + c * \alpha$$
$$y_1 = b + c * 1$$

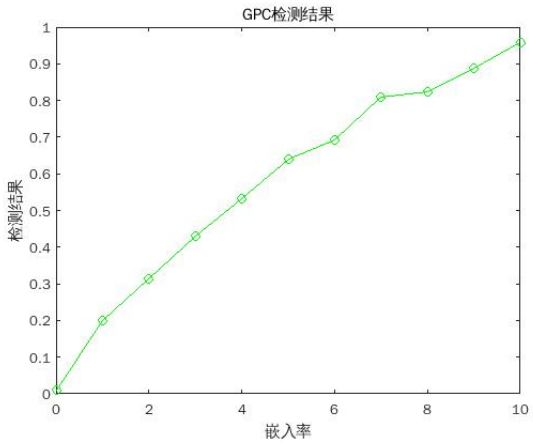
其中，$y_0 = 0$，$y_1 = r_1$，$y_\alpha = r_0$，求得

$$\alpha = \frac{y_\alpha - b}{c}.$$

图 6是对 Lena 这幅图像进行 GPC 隐写测试的结果，算法 4给出了信息量估计法的的描述。



(a) pic1. 检测的隐写率随嵌入率的变化



(b) pic2. 嵌入秘密信息的图片

Figure 6: GPC 隐写分析

算法 **4:** GPC

输入：待检测图像；

1 对载体图像计算 N_0 和 N_1，得到比值 r_0；
2 **if** $r_0 < T$ **then**
3 　输出：未嵌入信息。
4 **end**
5 **else**
6 　再对图像进行二次密写，计算 N_0 和 N_1, 得到比值 r_1；
7 　按求解方程组得到嵌入率 α；
8 　输出：嵌入率 α 。
9 **end**

2.5 基于相邻像素统计特性的隐写算法 (OEP)

记 S 为图像中所有像素构成的集合，$S = \{p_i | i = 0, 1, 2, \cdots, N\}$，$p_i^*$ 为 p_i 的相邻像素，(p_i, p_i^*) 为待检图像的相邻像素值对，按 p_i 和 p_i^* 的奇偶性及大小关系定义如下三个不相交的相邻像素对集合 S_1,S_2,S_3:S_1 是奇数像素值大于偶数像素值的相邻像素值对构成的集合，S_2 是偶数像素值大于奇数像素值的相邻像素值对构成的集合，S_3 是像素值同为奇像素或同为偶像素的相邻像素值对构成的集合。用集合表示就是：

$$S_1 = \{(p_i, p_i^*) | p_i (mod \quad 2) = 1, p_i^* (mod \quad 2) = 0, p_i > p_i^*, i = 1, 2, \cdots, N\}$$

$$S_2 = \{(p_i, p_i^*) | p_i (mod \quad 2) = 1, p_i^* (mod \quad 2) = 0, p_i < p_i^*, i = 1, 2, \cdots, N\}$$

$$S_3 = \{(p_i, p_i^*) | p_i (mod \quad 2) = p_i^* (mod \quad 2) = 0, i = 1, 2, \cdots, N\}$$

2.5.1 嵌入有无的判定

计算 $R = N_1/N_2$，选择一个合适的阈值 T(T 接近 1)，如果 $R \geq T$ 则认为嵌入了秘密信息，否则没有。下图 7是对 Lena 这幅图像进行基于相邻像素统计特性的隐写测试的结果。

2.5.2 隐写率的计算

类似于 RS 隐写分析，嵌入秘密信息后三个集合都会发生变化，记变化后的集合为 S_1^*,S_2^*,S_3^*。S_1 在可以分为两个子集，

$$S_{11} = \{(p_i, p_i^*) | p_i (mod \quad 2) = 1, p_i^* (mod \quad 2) = 0, p_i - p_i^* = 1, i = 1, 2, \cdots, N\}$$

$$S_{12} = \{(p_i, p_i^*) | p_i (mod \quad 2) = 1, p_i^* (mod \quad 2) = 0, p_i - p_i^* > 1, i = 1, 2, \cdots, N\}$$

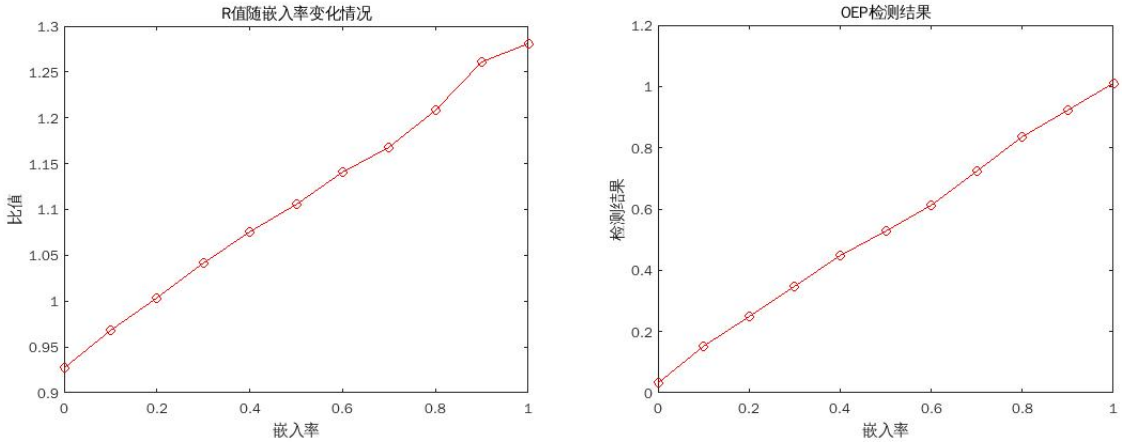
不难得出

$$\|S_2^*\| = \|S_2\|(1-\alpha/2) + S_{12}\alpha/2$$
$$\|S_{12}^*\| = \|S_{12}\|(1-\alpha/2) + S_2\alpha/2$$

根据文献 [5]，$\|S_2\| = 0.125$，于是隐写率 α 就为

$$\alpha = 2\frac{\|S_2\| - \|S_2^*\|}{2\|S_2\| - \|S_{12}^*\| - \|S_2^*\|}.$$

算法 5给出了信息量估计法的的描述。



(a) pic1. 检测的隐写率随嵌入率的变化　　(b) pic2.OEP 算法分析结果

Figure 7: OEP 隐写分析

算法 **5:** OEP

输入：待检测图像；

1 对载体图像统计，计算集合 S_1 和 S_2 元素的个数，得到比值 R；
2 **if** $R < T$ **then**
3　　输出：未嵌入信息。
4 **end**
5 **else**
6　　根据公式求出嵌入率 α；
7　　输出：嵌入率 α 。
8 **end**

3　实验结果以及分析

本文从互联网爬取 200 张 128*128 的彩色图片，然后把彩色图像灰度化，用不同的算法对其进行隐写分析，比较各种算法分析出来的结果。首先分别以嵌入率 α =0,0.1,0.2，$\cdots$,0.9,1.0 对上述图片嵌入秘密信息，生成 2200 张灰度图片，然后分别用上述五种算法进行分析。表 1给出了在不同嵌入率下平均的隐写分析的结果。

对于隐写分析的评估，主要有虚警率和漏警率。虚警率指的是，将没有隐藏信息的载体判断为含有秘密信息的载体的百分率，而漏警率是含密载体判定为未含有秘密图像的载体。从表中数据来看，RS、卡方分析、IQE 和 OEP 的虚警率都很高，只有 GPC 虚警率很低；另外一方面，只有 IQE 的漏警率比较高，其他的算法在这方面的表现都比较好的。

从表中的数据来看，RS 隐写分析分析在嵌入率为 0% 和 100% 最为准确，与实际嵌入率相比分析出的结果大概有 10% 的误差，这个结果是可以接收的。卡方隐写分析算法给出的分析结果是可能有的结果而非嵌入率，在嵌入率较小的时候检测为 0，在嵌入率约 70% 的时候开始检测出有嵌入，嵌入率越大分析结果约接近 1。信息量估计法得出的结果误差在嵌入率比较小的时候差距不大，在嵌入

Table 1: 五种算法在不同嵌入率下的分析结果

算法	RS	Chi	GPC	IQE	OEP
0	0.1151	0.6901	0.0379	0.3295	0.1610
0.1	0.1153	0.6943	0.0537	0.3251	0.1840
0.2	0.1279	0.7192	0.1029	0.342	0.2227
0.3	0.1473	0.7311	0.1535	0.3392	0.2940
0.4	0.209	0.7532	0.2396	0.3808	0.3300
0.5	0.2541	0.7789	0.3110	0.3621	0.4128
0.6	0.3464	0.8104	0.3895	0.4587	0.5027
0.7	0.439	0.8371	0.5074	0.5379	0.5826
0.8	0.5486	0.8656	0.6376	0.5866	0.7219
0.9	0.7204	0.9046	0.8203	0.6759	0.8481
1	0.989	0.9998	1.0241	0.7241	1.0101

率比较大的时候检测较为准，确事实上，作者在文献 [5] 中也指出，得出的概率一般作为判断是否嵌入了秘密信息的依据，而不可以当做实际的嵌入率。GPC 分析法在隐写率较大的时候比较准确，较小的时候估计不准确，得出的曲线比较稳定；最后 OEP 算法得出的效果是比较好的，它对于嵌入率的估计非常接近，误差非常小。

4 结束语

针对 LSB 隐写算法，除了上述隐写分析算法外，还有一些隐写分析算法例如 SPA 等等算法。这些算法本质上都是利用嵌入秘密信息前后载体图像的统计变化构建相应的算法来检测嵌入，有些算法本职上是一致的。为了抵抗这些隐写算法的攻击，保证秘密信息的安全性，需要改进 LSB 算法，使之统计特性不再这么明显。而另一方面，目前的的隐写分析算法仍需改进，使得能够准确估计嵌入率和获得更强的泛化性能，使得对其他信息隐藏算法也有效。

References

[1] P. J. Angeline, G. M. Saunders, and J. B. Pollack, "An evolutionary algorithm that constructs recurrent neural networks," IEEE Trans. Neural Netw., vol. 5, no. 1, pp. 54–65, Jan. 1994.

[2] Fridrich J, Goljan M, Du R. Reliable detection of LSB steganography in grayscale and color images. In: Dittmann J, Nahrstedt K, et al. eds. Proc. of the ACM Workshop on Multimedia and Security. Ottawa: ACM Press, 2001. 27 30.

[3] Fridrich J, Goljan M, Du R. Detecting LSB steganography in color and gray-scale images. IEEE Multimedia, 2001,8(4):22 28.

[4] Xinpeng Zhang, Shuozhong Wang, Kaiwen Zhang. Steganography with the Least Histogram Abnormality [C]// Computer Network Security, Lecture Notes in Computer Science 2776. Heidelberg,Berlin: Springer-Verlag, 2003: 395-406.

[5] Xia D . Analysis of MIDI Audio Files Based on Method of Information Quantity Estimation[J]. Applied Mechanics and Materials, 2015, 733:838-841.

[6] 秦姣华, 孙星明, 程小艳. 基于相邻像素统计特性的 LSB 隐写分析技术 [J]. 系统仿真学报, 2007, 19(24):5856-5860.

[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 [J]. 软件学报, 2004, 15(1): 151-158.

[8] Sorina Dumitrescu, Xiaolin Wu, Zhe Wang. Detection of LSB Steganography via Sample Pair Analysis [J]. IEEE Transactions on Signal Processing (S0018-9380), 2003, 51(7): 1995–2007.