

## 实验三 专用隐写分析（二）

姓名：王运韬

学号：201628018627123

### 一、实验目的：

理解图像原始快速匹配分析（*Raw Quick Pair, RQP*）的原理，并采用其攻击彩色空域图像的最低比特位替换（*LSBR*）隐写算法。

### 二、实验原理：

图像原始快速匹配分析能够有效检测基于彩色空域图像的 *LSBR* 隐写算法，其原理如下。

对于彩色空域图像而言，一幅图像中出现的实际颜色数目为  $U$ ，而图像像素总数  $MN$  代表中图像中可能出现的颜色数上限，定义  $U/MN$  为颜色复杂度。由于空域隐写通常使得图像中出现的颜色数目  $U$  增多，根据颜色复杂度特征，可以判断一幅图像是否隐写。

该方法具体计算过程如下：

（1）计算相邻颜色对的归一化值。若两个颜色值  $(R_1, G_1, B_1)$  和  $(R_2, G_2, B_2)$  满足：

$$|R_1 - R_2| \leq 1, |G_1 - G_2| \leq 1, |B_1 - B_2| \leq 1$$

则称它们为相邻颜色对。对于待分析的彩色图像  $I$ ，设其所含相邻颜色对的数量为  $P$ ，对  $P$  进行归一化

$$Q = P / \binom{U}{2}$$

其中， $\binom{U}{2}$  代表该图像的所有颜色对组合数。一般使用 *LSB* 空域隐写方法在图像中嵌入隐蔽信息会导致  $Q$  值增加，当图像已含有隐蔽信息时，再次嵌入新的隐蔽信息后  $Q$  值得增量小于在不含隐蔽信息的图像中首次嵌入的情况。

（2）进行基于 *RGB* 分量的 *LSB* 二次嵌入。首先，计算待分析图像的  $Q$  值，

记为 $Q_1$ 。接着，使用基于 $RGB$ 分量的 $LSBR$ 隐写算法在待测图像中嵌入 $3aMN$ 比特的测试数据，其中， $MN$ 为图像的像素总数， $a$ 代表嵌入率。计算二次嵌入后的隐写图像 $Q$ 值，记为 $Q_2$ 。

(3) 计算二次嵌入前后 $Q$ 值比例值。计算二次嵌入前后 $Q$ 值的比例 $R = Q_2 / Q_1$ ，并与某个分析阈值 $T$ 进行比较。当 $R < T$ 时判定为有隐写，当 $R \geq T$ 时判定为未见隐写。

### 三、实验步骤：

1. 进行分析参数配置：设置 $RQP$ 分析的消息嵌入率，以及检测窗口的大小和位置；

2. 进行 $RQP$ 分析：对于彩色空域图像的 $LSBR$ 隐写算法产生的载体和某种嵌入率的隐写样本，先随机选取其中的一部分图像计算 $RQP$ 分析值，并得到最佳判断阈值，再根据此阈值对剩余图像进行隐写分类判决，记录该嵌入率下的 $RQP$ 判断阈值和隐写分析的真阳性率、真阴性率和正确率。

### 四、实验结果：

$RQP$ 分析的 $matlab$ 函数接口为 $R = analysis(imgPath, var)$ ，其中 $imgPath$ 表示待分析图像的路径， $var$ 代表配置参数，其为结构体变量，成员包括 $rate$ ， $width$ ， $height$ ， $startX$ ， $startY$ ，依次表示：对待测图像进行二次嵌入的嵌入率（该值一般很小，建议取值区间为 $[0.01, 0.05]$ ），检测窗口的宽度、高度（不宜过大，否则影响算法执行时间，一般小于 $100$ 像素），检测窗口在图像中的水平、垂直偏移（以像素为单位）。

四个图像集 $BOSS\_LSBR\_000$ 、 $BOSS\_LSBR\_100$ 、 $BOSS\_LSBR\_300$ 、 $BOSS\_LSBR\_500$ ，文件夹命名格式为" $BOSS\_LSBR\_EmbedRate$ "， $EmbedRate$ 代表嵌入率，例如，文件夹" $BOSS\_LSBR\_300$ "包含的是采用彩色空域图像的 $LSBR$ 隐写算法，在 $0.3bpp$ 的嵌入率下制备的隐写样本。相应的 $R$ 值分别为 $R_1$ 、 $R_2$ 、 $R_3$ 、 $R_4$ 。

二次嵌入的嵌入率和检测窗口大小对最终 $RQP$ 分析均有影响，选取合适的配置参数对隐写分析十分必要，以下图示为对这两个参数的分析。

利用 $matlab$ 的 $tic, toc$ 函数对 $analysis(image, var)$ 函数运行时间进行统计，每张图片需运行 $2s$ ，4个图像集，每个图像集中有200张图片。对于窗口大小，只

考虑正方形窗口，窗口大小为10:10:100，共10个梯度；对于二次嵌入的嵌入率，0.01:0.005:0.05，共9个梯度。所以，共需计算72000次，即144000s，40h，耗时较长，此处可以考虑使用多核多线程或者减少分级梯度数和估计样本数，避免在同一个程序同时讨论窗口大小和二次嵌入率的影响，减少循环次数。其次，窗口大小也会影响程序运行速度。

据不完全统计，*BOSS\_LSBR\_100*、*BOSS\_LSBR\_300*、*BOSS\_LSBR\_500*三个隐写图像集的*RQP*分析值恒等于1，因此在实验中主要考虑未隐写图像集*BOSS\_LSBR\_000*的*RQP*分析结果，从而确定出合适的阈值，主要目的是为了降低未隐写样本的虚警率，从而提高隐写分析的准确率。

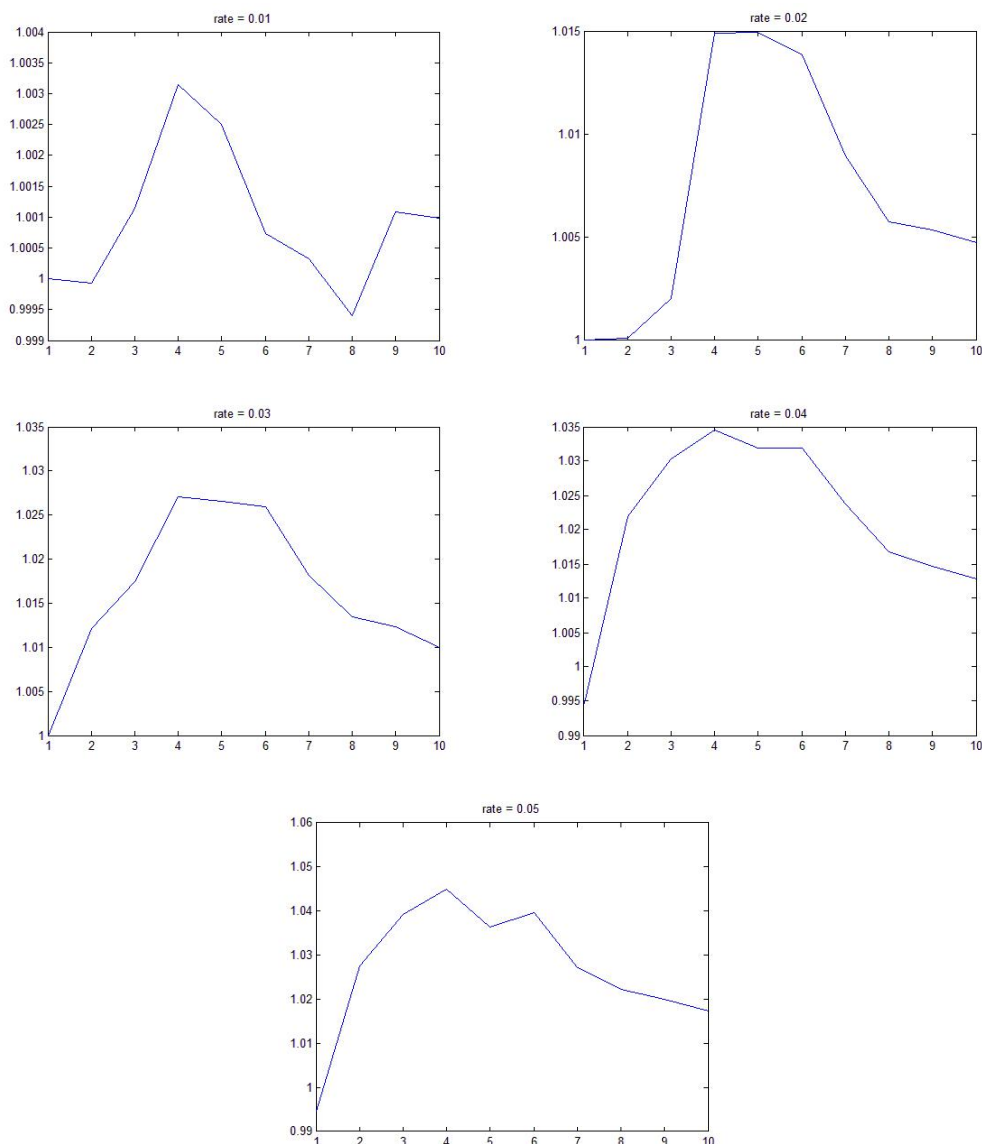


Fig 1. 二次嵌入率对*RQP*分析值得影响

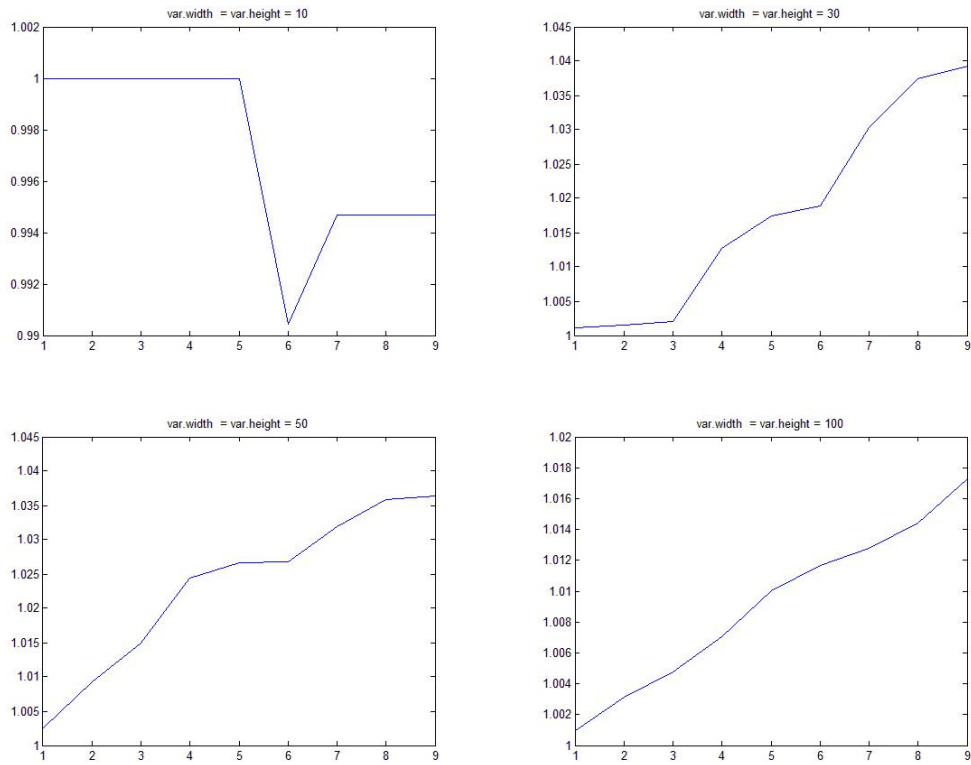


Fig 2. 窗口大小对  $RQP$  分析值得影响

最终选定分析参数为:  $var.rate = 0.04$ ,  $var.width = 100$ ,  $var.height = 100$ , 阈值设定为:  $T = 1$ , 当  $RQP$  分析值  $R \leq 1$  时即可判定为隐写。

以四个图像集  $BOSS\_LSBR\_000$ 、 $BOSS\_LSBR\_100$ 、 $BOSS\_LSBR\_300$ 、 $BOSS\_LSBR\_500$  为训练集, 获得准确率, 真分析阈值设定是否合理。

$BOSS\_LSBR\_000$  图像集:

虚警率:  $FP = 0.035$ , 准确率为:  $acc = 0.965$ 。下图是  $RQP$  分析值的散点图和分布直方图。

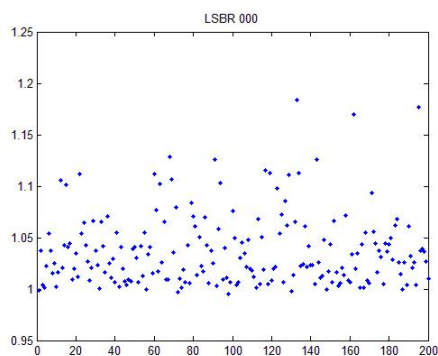


Fig 3.  $RQP$  分析值散点图

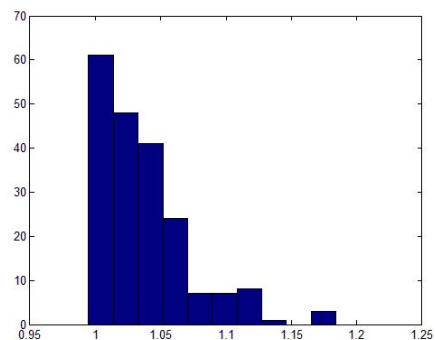


Fig 4.  $RQP$  分析值分布直方图

*BOSS\_LSBR\_100* 图像集:

真阳性率:  $FP=1$ ，准确率为:  $acc=1$ 。下图是 *RQP* 分析值的散点图和分布直方图。

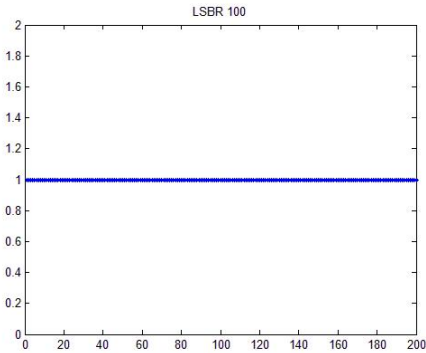


Fig 5. *RQP* 分析值散点图

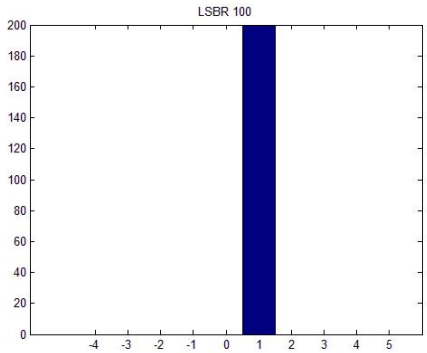


Fig 6. *RQP* 分析值分布直方图

*BOSS\_LSBR\_300* 图像集:

真阳性率:  $FP=1$ ，准确率为:  $acc=1$ 。下图是 *RQP* 分析值的散点图和分布直方图。

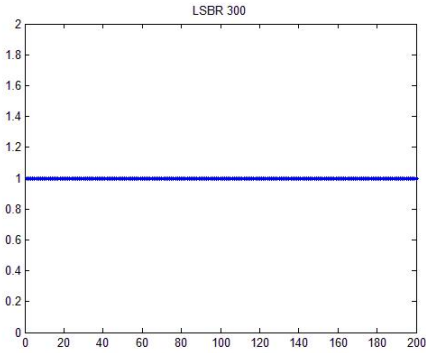


Fig 7. *RQP* 分析值散点图

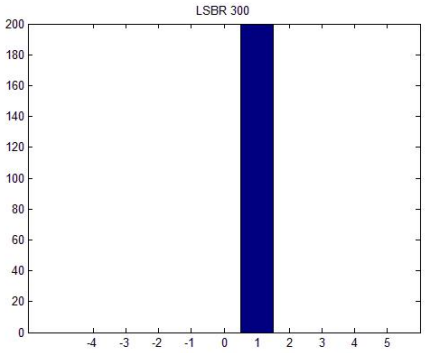


Fig 8. *RQP* 分析值分布直方图

*BOSS\_LSBR\_500* 图像集:

真阳性率:  $FP=1$ ，准确率为:  $acc=1$ 。下图是 *RQP* 分析值的散点图和分布直方图。

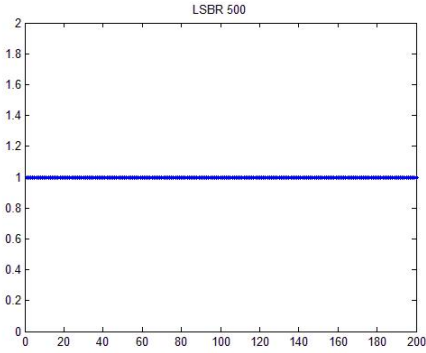


Fig 9. *RQP* 分析值散点图

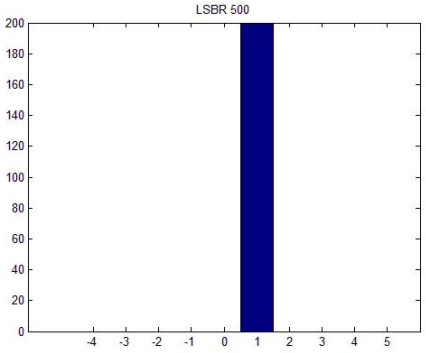


Fig 10. *RQP* 分析值分布直方图

测试集制备:

四个图像集 *BOSS\_LSBR\_000*、*BOSS\_LSBR\_100*、*BOSS\_LSBR\_300*、*BOSS\_LSBR\_500* 分别编号为  $i(i=1,2,3,4)$ ，每个图像集中有 200 张图片，分为 10 组 ( $j=1,2,3,\dots,10$ )。相应地，测试集也配置为相应的 10 组。测试集  $j(j=1,2,3,\dots,10)$  为每个图像集中的第  $j$  组图像，每组 20 张。因此，每个测试集有 80 张图片，且每个测试集的前 20 张图片均未发生隐写，方便计算准确率。

Workspace			
Name	Value	Min	Max
FP	[0,0,0,0,0,0,0,0,0,0]	0	0
R	<80x10 double>	0.9949	1.1841
TF	[0.9500,1,0.9500,...]	0.9500	1
TF_mean	0.9650	0.9650	0.9650
TF_num	19	19	19
TP	[1,1,1,1,1,1,1,1,1,1]	1	1
TP_mean	1	1	1
TP_num	60	60	60
accuracy	0.9825	0.9825	0.9825
file_path	'C:\\Users\\Adm...		
file_path_test	'C:\\Users\\Adm...		
file_path_tes...	'C:\\Users\\Adm...		
file_test_num	10	10	10
image_name	'9'		
image_num	80	80	80
image_num...	9	9	9
image_path	'C:\\Users\\Adm...		
image_path...	<80x1 struct>		
j	80	80	80
k	10	10	10
var	<1x1 struct>		

Fig 11. 仿真实验结果

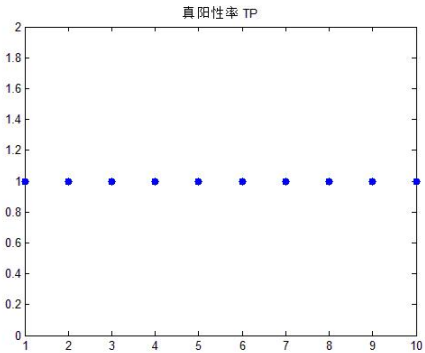


Fig 12. 测试集真阳性率 TP 散点图

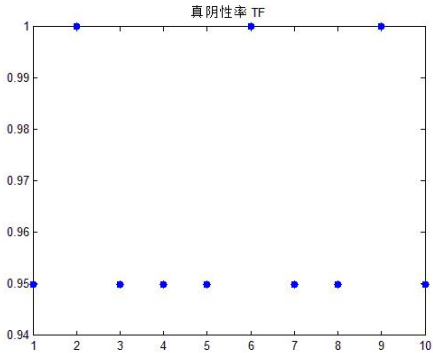


Fig 13. 测试集真阴性率 TF 散点图

$TP\_mean = 1$ ， $TF\_mean = 0.9650$ ，可得准确率为： $acc = 0.9825$ 。

*analysis.m* 脚本为核心代码，用于对图像进行 *RQP* 分析；

*threshold\_rate.m* 脚本用于分析二次嵌入的嵌入率 *var.rate* 对分析参数设定

的影响；

*threshold\_window.m* 脚本用于分析窗口大小 *var.height* 和 *var.width*（只考虑句型检测窗）对分析参数设定的影响；

*RQP\_analysis.m* 脚本用于对训练集进行分析，以选定合适的阈值  $T$ ；

*make\_testfile.m* 用于制备测试图像集；

*RQP\_test.m* 用于对测试图像集进行 *RQP* 分析，得到隐写分析准确率。

测试图像集过大，且制备速度快，因而可临时制作，并未附于附件中。

## 五、实验结论

一般使用 *LSB* 空域隐写方法在图像中嵌入隐蔽信息会导致  $Q$  值增加，当图像中已含有隐蔽信息时，再次嵌入新的隐蔽信息后， $Q$  值的增量小于在不含隐蔽信息的图像中首次嵌入的情况。

对图像进行 *RQP* 分析，即基于 *RGB* 分量的 *LSB* 二次嵌入，得到二次嵌入前后  $Q$  值的比例值，可有效实现对彩色图像 *LSBR* 的分析，准确率为 98.25%。

## 参考文献

[1] J. Fridrich, R. Du, and M. Long. Steganalysis of LSB encoding in color images. In Proc. ICME 2000, New York City, New York, USA, July 31 - August 2, vol.3: 1279-1282, 2000.