基于污染数据分析的隐写分析方法*

管　伟[1,2]，张 卫明[1,2]，刘文芬[1]
(1. 信息工程大学信息工程学院，河南 郑州 450002；
2. 中国科学院研究生院信息安全国家重点实验室，北京 100039)

摘　要：隐写分析是信息隐藏技术的重要分支，大量的隐写软件利用彩色及灰度图像的最低有效比特位（LSB）隐藏秘密消息。基于污染数据分析，针对图像空域 LSB 隐写算法提出了一种新的隐写分析方法，该方法不仅可以快速有效地检测出图像中秘密消息的存在性，同时还可以精确地估计出连续及随机间隔嵌入算法下所嵌入秘密消息的长度。

关键词：信息隐藏；隐写术；隐写分析；LSB；污染数据分析
中图分类号：TP309　　文献标识码：A　　文章编号：0529-6579（2004）S2-0062-05

信息隐藏技术作为信息安全领域的一种新技术，随着网络技术的发展得到了前所未有的重视和发展。隐写术和隐写分析是信息隐藏技术的两个重要分支。前者通过隐藏通信的存在性来获得隐秘通信的安全性，使得攻击者在不知道隐写密钥的情况下无法提取出秘密消息；后者则是对前者的攻击。

通过替换图像空域的最低有效比特位（LSB）来嵌入秘密消息的方法是隐写术中出现的较早也是较为经典的一种数据隐藏方法，针对此隐写算法已有一些研究结果。Westfeld 等[1] 最早提出通过分析像素值对（PoVs）建立卡方统计量，可以有效地检测出秘密消息的存在性。但这种算法只对连续 LSB 替换算法有效。Fridrich 等[2,3] 利用载密图像中的空域相关性，提出了 RS（regular singular）隐写分析算法。Dumitrescu 等[4] 基于基本集（primary set，PS）在自然图像模型的基础上建立了完整的 LSB 替换算法的数学模型。张涛等[5] 提出了一种基于差分直方图的隐写分析方法。以上 3 种方法不仅可以可靠地检测出图像中采用连续或随机间隔 LSB 替换方法嵌入的秘密消息，还可以较精确地估计秘密消息的大小。

本文运用污染模型的思想方法[6]，视被替换的 LSB 为污染点，将载密图像中像素值为奇数的像素点分为修改过的和未修改的两部分，把经过处理后获得的噪声数据看成来自污染分布的样本，从中寻找出某种优势，并利用这种优势建立新的统计量，这不仅可以有效地检测出连续或随机间隔替换算法下所嵌入秘密消息的存在性，并可精确地估计出秘密消息的长度，其结果优于 RS 和 PS 算法。

1　检测方法

1.1　符号说明

考虑一幅 $M\times N$ 大小的灰度载体图像（彩色图像类似）C，则 C 包含 $n=M\times N$ 个像素点。记图像中位置（k，j）的像素点对应的像素值为 x_{kj}，其中 $x_{kj}\in[0,255]$，$1\leqslant k\leqslant M$，$1\leqslant j\leqslant N$。为了分析方便，将像素值矩阵转化成一维序列，记为 $X=\{x_i\}_{i=1}^{n}$。那么采用空域 LSB 替换方法进行信息伪装的最大容量为 n bit。设嵌入了 qn 比特消息（$0\leqslant q\leqslant 1$，称为嵌入比例）后得到载密图像 $S=\{s_i\}_{i=1}^{n}$。即每个像素点被选来承载消息的概率是 q，进一步可假设消息与图像的 LSB 位独立（一般假定消息是经过加密的），从而消息和图像的 LSB 位相等的概率是 $1/2$，所以一个像素点被修改的概率是 $q/2$，则 X 中平均有 $qn/2$ 个点被修改。

1.2　方法的描述

设 $N(x_{kj})$ 表示 x_{kj} 及其相邻像素点。定义函数 T：$N(x)\rightarrow R$，其中 R 为实数集合，$x\in X$，其中 X 是任意一幅图像的像素值集合，称 $T(N(x))$ 为 x 的局部估计量。如 FIR 空间均衡滤波：

$$T(N(x_{kj}))=\sum_{r=-L}^{L}\sum_{s=-L}^{L}c_{rs}x_{k+r,j+s} \tag{1}$$

其中 c_{rs} 为常数，L 是一个取值较小的正整数。一种较为简单的形式是取 x_{kj} 及其相邻 8 个像素值的算术平均值：

* 收稿日期：2004－09－11
作者简介：管伟（1979 年生），男，硕士；E-mail：yzguanwei@yahoo.com.cn

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

$$T(N(x_{kj})) = \frac{1}{9}(x_{kj} + x_{k-1,j-1} + x_{k-1,j+1} + x_{k-1,j} + x_{k,j-1} + x_{k,j+1} + x_{k+1,j-1} + x_{k+1,j} + x_{k+1,j+1}) \quad (2)$$

为了检测图 S 中是否嵌有消息，即是否有 $q>0$，我们采用下面的方法，在 S 中再加嵌一段 0，1 随机序列 m，嵌入比例为 p，加嵌后得到的图像记为 $S_m = \{S_i^m\}_{i=1}^n$，其中加嵌消息的区域中由 0 修改到 1 的像素点的集合记为 S_m^0，其在 S 中对应位置的像素点的集合记为S^0；加嵌消息的区域中保持 1 不变（即未修改）的像素点的集合记为 S_m^1，其在 S 中对应位置的像素点的集合记为S^1。任给一幅图像 X，记其像素值与滤波值之差为

$$f(x) = x - T(N(x)), \qquad x \in X$$

当 X 分别为S 和S_m 时，记

$$F_{10}(x) = x - T(N(x)), \qquad x \in S_m^0 \quad (3)$$

$$f_{11}(x) = x - T(N(x)), \qquad x \in S_m^1 \quad (4)$$

$$f_{20}(x) = x - T(N(x)), \qquad x \in S^0 \quad (5)$$

$$f_{21}(x) = x - T(N(x)), \qquad x \in S^1 \quad (6)$$

当 x 取遍上述 4 个集合时，得到噪声数据 $\{f_{10}\}$，$\{f_{11}\}$，$\{f_{20}\}$，$\{f_{21}\}$，将这些数据四舍五入取整。显然，$f_{ij} \in [-255, 255]$，$i=1, 2$，$j=0, 1$。令 $t_{ij}(k)$ 等于 $f_{ij}(x)=k$ 的像素点的个数，即

$$t_{ij}(k) = |\{x: f_{ij}(x) = k\}| \quad (7)$$

相对应地求出 $f_{ij}(x)=k$ 的像素点在整幅图中出现的频率和均值，即有

$$p_{ij}(k) = \frac{t_{ij}(k)}{M \times N} = \frac{t_{ij}(k)}{n} \quad (8)$$

$$E_{ij} = \sum_{k=-255}^{255} k \cdot p_{ij}(k) \quad (9)$$

实验数据表明，$q=0$ 时，E_{11}，E_{20}，E_{21}为 10^{-3} ~ 10^{-4} 数量级的实数，其中 E_{20} 为负值，而 E_{10} 为 10^{-1}数量级的实数；$q>0$ 时，E_{11}，E_{20}，E_{21} 都比 $q=0$ 时相应高出 1 ~ 2 个数量级，E_{10} 则相同。这种现象就是我们需要利用的优势，于是考察统计量

$$R = \frac{|E_{10} - E_{11}|}{|E_{20} - E_{21}|} \quad (10)$$

将这种优势量化。

1.3　实验结果与分析

选用了 60 幅标准灰度图进行实验。在每幅图像中，分别取 $q=0 \sim 1.0$，步长 0.1 嵌入秘密消息。将序列 m 加嵌满图像的 LSB 平面，若 $q=0$，即图像为原始图像时，R 值比较大，从接近于 10 到几百甚至上千不等；反之，当 $q>0$ 时，则 R 较小，一般小于 10 到接近于零。两种情况下的结果不在一个数量级上，故可以凭借 R 的取值大小初步判断一幅图是否嵌有消息。根据现有的实验数据，将 19 作为判断是否嵌有消息的阈值：若 $R>19$，表明被检测对象为原始图像；反之，则说明图像中嵌有秘密信息。这种检测方法可以有效地检测出最低 0.05 ~ 0.08 bpp 嵌入比例的秘密消息。在对 60 幅嵌入比例大于 0.05 的载密图像的检测中，共成功检测出 58 幅。限于篇幅，这里只给出了 Lena 图像的部分实验数据：

表 1　Lena 图像的实验数据

Tab. 1　Experiment data of Lena

q	0	0.2	0.4	0.6	0.8	1.0
R	55.336	4.114	1.585	0.851	0.296	0.028

2　估计秘密消息的嵌入比例

2.1　噪声数据的污染模型

通过进一步的实验，发现在加嵌满整幅图像的情况下，R 值随着 q 的增加而递减，从而呈现出某种函数关系。而 ISB 方法本质上是把消息隐藏于图像的噪声部分，所以下面以载密图像的噪声数据为对象作进一步的分析研究。

设 ζ 服从正态分布 $N(0, \sigma^2)$，η 服从正态分布 $N(1, \sigma^2)$。令 $\zeta' = [\zeta]$，$\eta' = [\eta]$ 其中$[x]$ 为对 x 四舍五入取整。易证，$E[\zeta'] = E[\zeta]$，$E[\eta'] = E[\eta]$。记 ζ'，η' 的分布函数为$F(x)$ 和 $G(x)$。

经过 1.2 节中 4 个抽样操作可得到噪声数据。f_{10}，f_{11}，f_{20}，f_{21}。f_{10}，f_{20} 的样本容量是相同的，设为 N_1，f_{11}，f_{21}，的样本容量相同设为 N_2，记这 4 个抽样分别来自总体 ζ_{10}，ζ_{11}，ζ_{20}，ζ_{21}。若像素点的像素值为奇数，则称这个点为奇数点，否则称为偶数点。

首先是对 S_m 进行抽样：其中若 $s_i^m \in S_m$ 为奇数点，但对应的 $s_i \in S$ 为偶数点，由前面的分析知，这些点中大约有$\frac{q}{2}N_1$ 个点的像素值与载体图像 C 中对应位置上的像素值都为奇数，即又修改成 C 中的点，而另外大约$\left(1-\frac{q}{2}\right)N_1$ 个点是在嵌入消息的过程中，由偶数值加 1 得到的。一般可以认为，没有修改的像素点对应的噪声数据近似服从分布 $F(x)$，而作加 1 修改的像素点对应的噪声数据应近似服从分布 $G(x)$。所以总体 ζ_{10} 服从污染分布

$$F_{10}(x) = \frac{q}{2}F(x) + \left(1-\frac{q}{2}\right)G(x) \quad (11)$$

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

其污染率为$1-\frac{q}{2}$，则

$$E(\zeta_{10})=E_{10}=1-\frac{q}{2} \tag{12}$$

若 $S_i^m \in S_m$ 为奇数点，对应的 $S_i \in S$ 也为奇数点，类似地可得到总体 ξ_n 的分布和均值

$$F_{11}(x)=\left(1-\frac{q}{2}\right)F(x)+\frac{q}{2}G(x) \tag{13}$$

$$E(\xi_{11})=E_{11}=\frac{q}{2} \tag{14}$$

其次是对 S 进行抽样：其中若 $S_i \in S$ 为偶数点，但对应的 $s_i^m \in S_m$ 为奇数点，则大约有$\frac{q}{2}N_1$ 个点是嵌入过程中载体图像 C 中的奇数点减 1 得到的，另外$\left(1-\frac{q}{2}\right)N_1$ 个点与原来的相同，从而总体 ξ_{20} 服从分布

$$F_{20}(x)=\left(1-\frac{q}{2}\right)F(x)-\frac{q}{2}G(x) \tag{15}$$

$$E(\xi_{20})=E_{20}=-\frac{q}{2} \tag{16}$$

当 $s_i \in S$ 为奇数点，对应的 $s_i^m \in S_m$ 也为奇数点，则总体 ξ_{21}服从分布

$$F_{21}(x)=\left(1-\frac{q}{2}\right)F(x)+\frac{q}{2}G(x) \tag{17}$$

$$E(\xi_{21})=E_{21}=\frac{q}{2} \tag{18}$$

由公式（10）及（11）－（18）式即得

$$R=\frac{1-q}{q},\ q\neq 0 \text{ 及 } q=\frac{1}{1+R} \tag{19}$$

若 S 为载体图像，由（10）式求得的 R 值较大（远大于载密图像），故认为在 $q=0$ 时，（19）式也成立。据此可对秘密消息的嵌入比例进行估计。

2.2 实验结果

实验表明，公式（19）可以较为准确地估计出秘密消息的嵌入长度，由表 1 容易得到 Lena 图像中秘密消息嵌入比例的估计值。

表 2 Lena 图像的嵌入比例估计值

Tab. 2 The estimation of embedding rate in Lena

q	0	0.2	0.4	0.6	0.8	1.0
估计值	0.018	0.196	0.387	0.540	0.772	0.973

对上述 60 幅图像进行同样的计算，发现单幅图像之间对嵌入比例估计的精度有着较大的差异。一些图像的误差可以控制在 2%以内，而有些图像则超过 5%，甚至达到 10%以上。图 1 给出了 2 幅 windows 自带图像的实验数据作为对比。

估计值	q					
	0	0.2	0.4	0.6	0.8	1.0
Lilies	0.028866	0.17750	0.40201	0.61012	0.83082	0.97899
Hills	0.026991	0.25633	0.45782	0.64836	0.83082	0.98998

图 1 2 幅图像嵌入比例估计值的比较

Fig. 1 The comparison of estimation to two pictures

2.3 误差分析

考虑到加嵌的 0，1 序列 m 是随机选取的，这样求得的 R 值及 q 的估计值会受到 m 的影响。首先采用多次实验取平均的方法，分别加嵌 10 条不同的 0，1 序列，将 10 个估计值的均值作为 q 的最后估计值。

然而发现，图像 Lilies 中秘密消息嵌入比例的估计值可以较精确地逼近实际嵌入比例。而 hills 的估计值在取均值修正后仍然与实际值之间有较大的误差。比较 2 图，Lilies 色彩及纹理较为丰富，相对而言 Hills 则有着多个大块颜色相近的平滑区域，这些区域中像素值的变化很小。这类连续相近颜色的大块区域的结合处像素值的变化较大，造成这些位置上的噪声值出现异常。

考虑在实验中剔除这些异常值以达到对 q 估计值的进一步修正。经过实验发现采用像素点周围 4 个点像素值的平均值与其本身像素值的距离作为限制条件，对 q 估计值的修正效果最好，即根据

$$d(x_{ij})=\left|x_{ij}-\frac{x_{i-1,j}+x_{i,j-1}+x_{i,j+1}+x_{i+1,j}}{4}\right| \tag{20}$$

来判断是否把某个像素点的噪声数据舍弃。

保留 $d<3$ 那些像素点的噪声值，得出图像 Hills 进一步修正后秘密消息嵌入比例的估计值。

表 3 Hills 值修正后的估计值

Tab. 3 The estimation of Hills after value modified

q	0	0.2	0.4	0.6	0.8	1.0
估计值	0.029	0.2173	0.3999	0.5642	0.7961	0.994

3 性能比较与分析

为了比较我们的方法与 RS 隐写分析方法[3]、PS 基本集方法[4] 的可靠性和准确性，用这 3 种算

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

法对以上 60 幅图像做同样的实验。

对表 4 进行分析，若从估计值的均值来看，除嵌入比例 0 外，本文提出的算法整体上比 RS 及 PS 算法要优；从估计值的标准差来看，除嵌入比例为 0.2，0.4 外，此方法也与 RS 和 PS 算法性能相当。为了更直观地比较 3 种方法的性能，以 0.1 为步长，给出每个嵌入比例情况下估计值均值与真实值的差的绝对值和标准差的图表（图 2），均以百分比为单位。

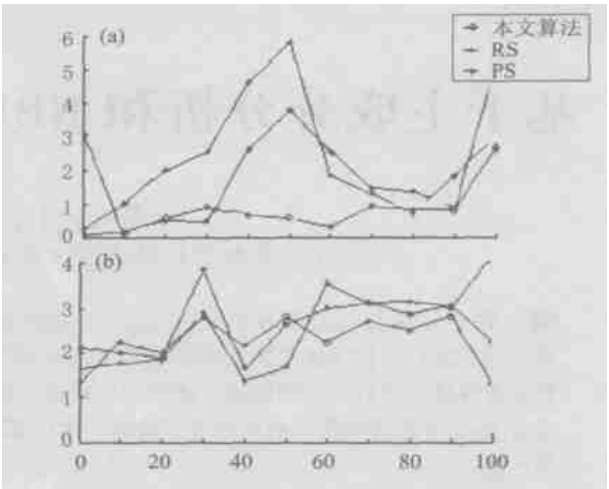

图 2　（a）绝对均值误差；（b）标准差

Fig. 2　（a）Absolute mean error；（b）Standard deviation

4　结　论

本文通过对基于 LSB 替换的信息伪装方法的深入研究，提出了一种基于污染模型的唯载密图像的检测方法，这种方法可以有效地检测出图像中是否嵌入了秘密消息。整个算法物理意义直观，实现简单，同样适用于彩色图像。我们注意到这种算法不仅可以对连续嵌入算法下的嵌入比例进行估计，而且同样能够给出随机间隔嵌入算法下的嵌入率，并有较高的精度。图像的检测与估计嵌入率之间函数关系简单明了，几乎可以同时实现。

表 4　与 RS 和 PS 算法的性能比较

Tab. 4　The comparison with RS and PS steganalysis technique

q	本文算法		RS		PS	
	$\mu(q)$	$\sigma(q)$	$\mu(q)$	$\sigma(q)$	$\mu(q)$	$\sigma(q)$
0	0.030665	0.013226	0.0024123	0.016271	0.00068108	0.021178
0.2	0.20622	0.020711	0.22008	0.018147	0.20477	0.018706
0.4	0.40712	0.022507	0.4459	0.013438	0.42601	0.016526
0.6	0.60303	0.02219	0.6182	0.035491	0.62512	0.029983
0.8	0.80804	0.024948	0.80686	0.031438	0.81349	0.028712
1.0	0.97356	0.013016	0.96975	0.042005	0.94851	0.022292

但是从实验结果来看，虽然得出的估计值的均值效果较好，但方差偏大。它与我们选取的加嵌序列，以及图像本身的特征性质有很大的关系。怎样选取加嵌的 0，1 序列，需要取多少次的平均值合适；对 d 的限制值，针对每幅图像应该如何控制；以及门限值 R_0 选为 19，存在着一定的取伪概率。这些问题我们还没有得到更为明确的答案，在后续的工作中有待进一步解决。并且如何运用本文的相关算法，对随机间隔嵌入的消息进行提取攻击也是我们关注的重要方向。此外，信息伪装的方法千变万化，也将关注其他类型的隐写算法，希望能够通过对其深入研究，给出其他有效的隐写分析方法。

参考文献：

[1]　WESTFELD A，PFITZMANN A. Attacks on steganographic systems // Lectures Notes in Computer Science[C]. New York：Springer-Verlag，2000：61—75.

[2]　FRIDRICH J，GOLJAN M，DU R. Detecting LSB steganography in color and gray-scale images[J]. Magazine of IEEE Multimedia，2001：22—28.

[3]　FRIDRICH J，GOLJAN M，DU R. Reliable detection of LSB steganography in grayscale and color images // Proceedings of ACM Workshop on Multimedia and Security[C]. 2001：27—30.

[4]　DUMITRESCU S，WU X L，WANG Z. Detection of LSB steganography via sample pair analysis // Lectures Notes in Computer Science[C]. New York：Springer-Verlag，2002：355—372.

[5]　ZHANG T，PING X J. A new approach to reliable detection of LSB steganography in nature images[J]. Signal Processing，2003，83(10)：2085—2094.

[6]　郑祖康，吴雪明，饶刚. 污染数据的处理[J]. 应用概率统计，1998，14(3)：307—312.

（下转第 70 页）

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

A Universal Blind Steganalysis Based on Principal Component Analysis and RBF Network

CHEN Dan, ZHANG Jian-hong, WANG Yu-min
(State Key Laboratory on ISN, Xi'an University of Electronical Technology, Xi'an 710071, China)

Abstract: On the basis of the limitation of Farid's universal blind steganalysis that image features for the classification are too many and have correlations, a new universal blind steganalysis scheme is presented. Decorrelation preprocessing using principal component analysis on the image statistics features and steganalysis classifier using RBF network are proposed. The scheme not only reduces the dimension of the feature vector enormously so that the speed of steganalysis detection is increased, but also the performance of the detection is improved as well. Applying this scheme and Farid's scheme to detecting the stego images produced by JSteg, EZStego, and S-Tools respectively, the comparison of these simulation results shows that this scheme is quite efficient for different size of messages embedded.

Key words: steganalysis; universal blind steganalysis; principal component analysis; RBF network

---

（上接第 65 页）

Steganalysis Based on Contamination Data Analysis

GUAN Wei[1,2], *ZHANG Wei-ming*[1,2], *LIU Wen-fen*[1]
(1. Institute of Information Engineering, Information Engineering University, Zhengzhou 450002, China;
2. State Key Laboratory of Information Security,
Graduate School of the Chinese Academy Sciences, Beijing 100039, China)

Abstract: Steganalysis is the important branch of information hiding, a large number of steganographic programs use the least significant bit (LSB) embedding as the method of choice for message hiding in color images, and grayscale images. In this paper, based on contaminated data analysis, a new steganalytic technique capable of a reliable detection of spatial LSB steganography is proposed. And this method can not only fast and reliably detect the existence of hiden message embedded in images, but also accurately estimate the amount of hiden message embedded by sequentially or randomly scattered algorithm.

Key words: information hiding; steganography; steganalysis; LSB; contamination data analysis

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net