应用机器学习于 Chi 平方密写分析的研究

葛 屾 高 阳 周如益

(南京大学计算机软件新技术国家重点实验室 南京 210093)

摘 要 回顾当前的密写和密写分析技术,并将机器学习方法应用到经典的 χ^2 密写分析上。给出了方法的框架并进行了实验,然后对实验结果进行分析和比较,得出了应用机器学习方法的 χ^2 密写分析优于简单 χ^2 密写分析的结论,从而肯定了机器学习方法的有效性。最后指出了在密写分析技术中进一步应用机器学习方法的方向。

关键词 机器学习,密写技术,Chi 平方密写分析

Applying Machine Learning to Chi-square Steganalysis: A Case Study

GE Shen GAO Yang ZHOU Ru-Yi

(National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093)

Abstract In this paper, conventional steganography and steganalysis techniques are reviewed. In order to get satisfactory accuracy of hidden information detection, machine learning methods are applied to chi-square steganalysis. The experiment results are promising, and further analysis gives the reason for the effectiveness of the machine learning based methods. Finally, we provide some possible future work of applying machine learning methods to steganalysis.

Keywords Machine learning, Steganography, Chi-square steganalysis

1 引言

信息隐藏技术在最近的十几年来成为了信息技术的研究热点。随着互联网的飞速发展,以及数字媒体的广泛应用,信息隐藏技术中的隐蔽通信和数字水印等技术得到了长足的发展,同时针对信息隐藏技术的攻击技术也随之发展起来,形成了今天相互对抗相互推动的局面。

信息隐藏技术包括了密写(Steganography)和密写分析(Steganalysis)技术。密写(Steganography)一词来源于希腊语的字根 στεγανος 和 γραφειν[1],其含义就通过表面上无害的文件来传递秘密信息[2]。传统的密码学(Cryptography)通信的目的只保证传输信息的秘密性,并不掩盖通信的事实,而密写的目标是保证通信的过程的隐蔽性,使人无法察觉到通信正在进行,从而保证信息的安全。为实现这个目标,大多数算法修改了媒体文件中的冗余位来实现信息的嵌入。密写是一种隐蔽通信的艺术,从而提供了对隐秘通信的否认[3]。由于密写技术很可能会被误用来进行犯罪活动,因此基于安全性的考虑,密写分析的技术是必要的。密写分析的目的在于揭示出文件中的隐蔽信息的存在性,或估计嵌入信息的长度,甚至提取出嵌入的信息[4]。密写分析主要利用了媒体文件在嵌入信息前后之间的差异,通过某些算法来发现这些差别,从而达到目的。

传统的 χ^2 密写分析可以用来估计顺序 LSB 嵌入的信息量,但是如果用它来对图像中隐藏信息的存在性进行检测,效果却不是很理想。本文将机器学习方法应用到 χ^2 密写分析上来检测图像中嵌入信息的存在性,并对结果进行分析,得出了机器学习方法的有效性。本文第 2 节回顾主要的图像密写和密写分析方法,然后着重介绍 LSB 密写分析、χ^2 方法以及我们自己的机器学习方法。第 3 节描述我们的实验并对结果进行分析以得出结论。最后是总结以及展望。

2 密写和密写分析技术

2.1 主要的密写和密写分析技术

密写技术是信息隐藏技术的一种,信息隐藏技术还包括数字水印等版权保护技术等。水印技术与密写技术在通信内容、稳健性、隐蔽性和嵌入数据量的方面有很大的不同[5]。图像密写技术大体上可以分为两大类,针对位图进行嵌入和针对变换域图像(如 JPEG 等)。

对位图进行直接嵌入的算法主要是 LSB 方法及其改进。LSB(Least Significant Bit,最低有效位)指的是对图像灰度值贡献最小的位平面。原始的 LSB 方法就是用欲嵌入的秘密信息代替图像的最低有效位平面,因为最低有效位对灰度的影响很小,所以人的肉眼无法察觉到隐藏信息的存在。经过原始的 LSB 方法嵌入的图像有残留的统计特征,所以出现了很多 LSB 的改进,如广义 LSB[6]等,最新的改进有 PSP[7]等。

现在研究得比较多的就是针对 JPEG 文件的密写技术,因为 JPEG 是非常流行的图像文件格式标准。Jsteg[8]密写将信息嵌入在量化后的 DCT 系数的 LSB 上,实现简单但安全性不好。OutGuess[9]利用频率计数保留了一些统计信息,使一些基于统计量检测的密写分析无效。F5[10]保留了 DCT 系数直方图,并应用混洗和矩阵编码技术以提高密写的性能。在 JPEG 上最新的密写方法有基于模型的嵌入(Model based embedding)[11]和扰动量化嵌入(Perturbed quantization embedding)[12]等。

与密写技术的分类类似,密写分析技术也可以根据算法所针对的图像表现形式来分类,但是也可以按照是否使用实例进行训练而分为两类:一类使用实例构造分类器,另一类由启发式参数模型来计算图像的数值特征,然后由这些值进行

葛 屾 研究生,主要研究方向:人工智能、机器学习、信息隐藏。

判断。这种方法没有训练过程，训练隐含在预指定的参数模型之中。

基于实例的方法一般包含了对训练图像进行特征提取的过程，以及后续训练过程，最后构造分类器，把图像当作输入，分类的结果当作输出。如下的几种方法均使用了实例训练的模式：Avicibas 等人提出的方法使用了 IQM(Image Quality Metric，图像质量度量)[13]。他们使用 ANOVA 算法来选择一系列 IQM，使用多元回归来训练分类器并进行分类。Lyu 和 Farid[14] 提出的检测隐藏信息的方法使用小波变换，得到图像变换后的高阶统计量，然后通过训练一个支持向量机(SVM，Support Vector Machines)来实现分类。他们扩展了自己的工作以使用色彩统计信息和非线性 SVM[15]。Fridrich[16] 提出了一种方法，即使用基准特征(Calibrated Feature)来估计无隐藏信息的原始图像，然后用了 23 个一、二阶统计量来构建分类器，取得了满意的效果。

非基于实例的方法采用预定义参数模型来挖掘图像的统计特征，并且用预定义的规则进行判断。下面的方法均为此类方法：Pfitzmann 和 Westfeld[17] 提出了基于值对的 χ^2 密写分析，他们应用了统计学理论，我们将在下一节详细描述该方法。Fridrich 等人[4] 提出了 RS 密写分析：他们根据翻转操作改变区别函数的值的方式定义了三个像素组：R，S 和 U，然后通过 RS 图上曲线之间交点的计算来估计隐藏信息的长度。Fridrich，Goljan 和 Du[4] 提出了一种基于 JPEG 兼容性的判别方法。他们指出经 JPEG 压缩后再保存为非压缩格式的文件在经过嵌入后会有破坏 JPEG 的特征存在，所以他们检查每个 8×8 块来检查这种不兼容性。他们的方法可以检测短至 1 比特的信息，所以他们建议不要采用 JPEG 图像转换后的图片当成嵌入源。

2.2 针对顺序 LSB 嵌入的 χ^2 密写分析方法

针对一幅待嵌入的图像，LSB 方法在图像的灰度值的最低位嵌入秘密信息位，如果与隐藏的信息位与图像的最低位相同，则无需改变，反之要改变图像的最低位，因此这是一个位翻转的过程。设图像的灰度值为 j，则 $j\in[0,255]$，如果 $j=2i$，则翻转后变为 $2i+1$，绝不会变成 $2i-1$，如果 $j=2i+1$，则翻转后变成 $2i$，而不会变成 $2i+2$。这样的话 LSB 的隐藏操作就在图像中留下了一些可察觉的统计痕迹。

Pfitzmann 和 Westfeld[17] 由此提出了值对(PoV，，Pairs of Values)的概念，将其它位相同但是最低位不同的两个灰度值组成为值对，由 LSB 密写方案可以知道，位翻转的操作仅在值对之间进行转换，不同的值对之间没有影响。设灰度值为 $2i$，$2i+1$ 的值出现的频度为 h_{2i}，h_{2i+1}。因为秘密信息在嵌入前往往经过加密等操作，所以可以看成是随机的比特流，而且 0，1 出现的频率接近 1/2。这样的话图像在经过 LSB 嵌入后，每个值对的两个灰度值的出现频率会趋向于相等。为了判定 h_{2i}，h_{2i+1} 的出现次数分布是否显著不同，我们可以应用统计分析中的 χ^2 检验，计算如下的 χ^2 检验统计量(嵌入后 h_{2i} 的期望值为 $h_{2i}^* =(h_{2i}+h_{2i}+1)/2$)：

$$\chi^2_{k-1}=\sum_{i=1}^{k}\frac{(h_{2i}-h_{2i}^*)^2}{h_{2i}^*}$$

然后用 χ^2 分布的密度函数便可以计算图像被密写的可能性 p：

$$p=1-\frac{1}{2^{\frac{k-1}{2}}\Gamma(\frac{k-1}{2})}\int_0^{\chi^2_{k-1}} e^{-\frac{1}{2}x}x^{\frac{k-1}{2}-1}\mathrm{d}x$$

如果 p 接近于 1，则说明图像可能存在隐藏信息。

χ^2 密写分析可以估计嵌入信息的长度，但是也可以用作隐藏信息的检测。用于隐藏信息检测时，一般采用的方法是预先定义阈值，然后对 χ^2 密写分析的分段概率系数进行检验，如果有大于阈值的系数出现则判断为存在隐藏信息，否则就判定不存在，我们在后文中称这种方法为简单的 χ^2 系数方法。这种方法简单易行，但是由于图像本身的最低位的随机性程度的影响，实际分类的准确性还不很令人满意。为了进一步提高分类的准确率，我们将机器学习方法应用到 χ^2 隐藏信息检测当中，期望能提高检测的准确率与稳定性。

2.3 基于机器学习的密写分析

我们将机器学习引入 χ^2 隐藏信息检测，采用文[18]中的 pov3 算法计算 χ^2 概率系数。pov3 算法将图像位平面像素集合划分为 100 段，对每一段，计算从第一段开始到该段结束所有像素的 χ^2 概率值，这样每个颜色分量平面对应 100 个系数，通过这些值来进行密写检测。我们采用了彩色图片嵌入，它与灰度图像嵌入的唯一区别是嵌入容量的增加。我们应用机器学习的方式如下：用 pov3 计算彩色位图的 RGB 三个平面各一百个 χ^2 系数，然后加上一个表示分类的标记，其中 0 表示不含隐藏信息的图像类，1 表示经过嵌入的图像类。在这里，机器学习方法可以任意选择，但是为了测试，我们会在下一节选择不同的方法来检验方法不同对结果产生的影响。我们给出我们方法的框架如图 1，同时也列出简单的 χ^2 系数方法做对比。

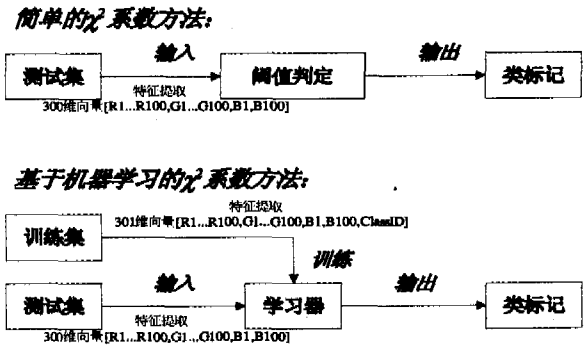

图 1 简单 χ^2 方法和基于机器学习的 χ^2 方法的框架比较

3 实验内容和分析

3.1 实验平台

我们收集网站上的图片建立了测试图像库，用 LSB 方法进行顺序嵌入。我们采用 WEKA 作为实验工具，来测试不同的机器学习方法所能得到的结果。我们选择了 Naïve Bayes，Bayes Net，J48 决策树，kNN，SVM(RBF 核的 SMO 方法)以及 BP 神经网络进行分类器的训练学习，并对结果做了 10-fold 交叉验证，同时选定 99% 和 95% 作为简单的 χ^2 概率系数中选择的阈值进行对比。为了研究图像本身复杂程度以及嵌入信息的长度对结果的影响，计算了图像的 HSV 颜色空间下的 H 分量离散化后的熵来作为图像复杂度的度量，然后将图像按复杂度(也就是位平面的随机程度)的不同分成五个等级，同时我们按照 10%，20%，50%，100% 四个等级分别进行嵌入。我们对等级和嵌入率进行汇总混合，以便对应真实情况。实验的数据集构成如下，对每种单独的复杂度等级和嵌入率对应的数据由没有嵌入的源图的系数向量和相应嵌入图的系数向量组成，这样经过嵌入的和没有嵌入的向量个数之比为 1∶1。对复杂度汇总混合时将所有的不同复杂度上

的图像数据集进行合并，但是对嵌入率的汇总混合不同，它由没有经过嵌入的系数和相应图像的四个不同嵌入率的系数构成，经过嵌入的和没有嵌入的系数之比为 4∶1。实验选用了 989 幅图片。

3.2 实验结果和分析

实验的部分结果数据如表 1。

表 1 部分实验结果数据表

Naïve Bayes	Level 0	Level 1	Level 2	Level 3	Level 4	All Mixed
Embed 0.1	97.7273%	97.9167%	91.6309%	85.8000%	85.4938%	81.5976%
Embed 0.2	100.0000%	98.6111%	96.7811%	96.3000%	94.7531%	95.6016%
Embed 0.5	100.0000%	99.3056%	99.1416%	98.8000%	98.7654%	98.8372%
Embed 1.0	100.0000%	99.3056%	99.1416%	99.6000%	99.3827%	99.1911%
All Mixed	100.0000%	96.1111%	77.9399%	73.9200%	72.4691%	71.8301%
J48	Level 0	Level 1	Level 2	Level 3	Level 4	All Mixed
Embed 0.1	97.7273%	97.9167%	97.2103%	95.7000%	93.5185%	97.3711%
Embed 0.2	100.0000%	99.3056%	97.8541%	99.0000%	96.2963%	98.5844%
Embed 0.5	100.0000%	100.0000%	98.9270%	99.8000%	99.0741%	99.4944%
Embed 1.0	100.0000%	100.0000%	99.3562%	99.7000%	99.6914%	99.6967%
All Mixed	99.0909%	98.8889%	99.0558%	98.0800%	96.4198%	97.9980%
chi 99%	Level 0	Level 1	Level 2	Level 3	Level 4	All Mixed
Embed 0.1	95.4545%	91.6667%	89.9142%	80.0000%	71.6049%	82.1537%
Embed 0.2	97.7273%	96.5278%	91.4163%	80.1000%	71.9136%	83.0131%
Embed 0.5	100.0000%	97.9167%	91.4163%	80.2000%	71.9136%	83.2154%
Embed 1.0	100.0000%	97.9167%	91.6309%	80.2000%	71.9136%	83.2659%
All Mixed	97.2727%	96.1111%	95.7940%	91.9600%	88.6420%	92.7401%
chi 95%	Level 0	Level 1	Level 2	Level 3	Level 4	All Mixed
Embed 0.1	100.0000%	96.5278%	90.1288%	77.5000%	67.9012%	80.7887%
Embed 0.2	100.0000%	97.9167%	90.5579%	77.5000%	67.9012%	80.9909%
Embed 0.5	100.0000%	97.9167%	90.5579%	77.5000%	67.9012%	80.9909%
Embed 1.0	100.0000%	97.9167%	90.5579%	77.5000%	67.9012%	80.9909%
All Mixed	100.0000%	98.6111%	96.0515%	91.0000%	87.1605%	92.3155%

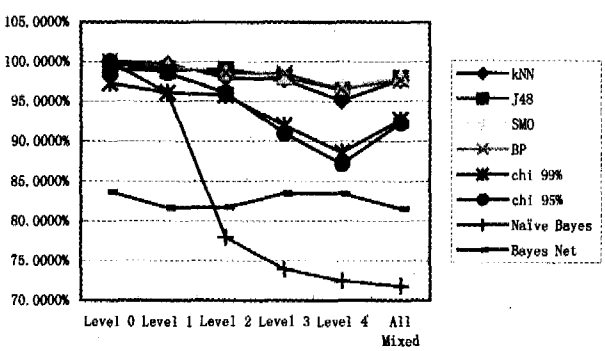

图 2 在不同图像复杂度下的结果(所有嵌入率混合)

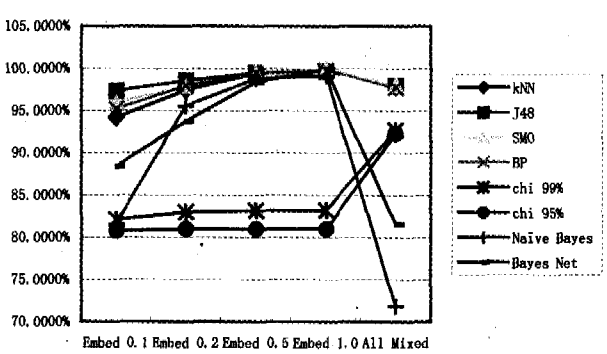

图 3 在不同嵌入率下的结果(所有复杂度混合)

3.2.1 实验数据总体趋势

由整个数据结果来看，分类的精确度随着原始图像的复杂程度的升高而递减，并随着图像的嵌入比例的升高而升高。其解释比较直观，因为图像本身的复杂程度高的话，会和由嵌入而引起的复杂性相混合，从而造成 χ^2 方法的计算出来的系数偏高，引起误判。而在相同的复杂性下，嵌入比例越高，由嵌入带来的复杂性就越大，从而会使 χ^2 方法的计算出来的系数偏高，易于判断。

由图 2，图 3 可以看出，在嵌入率混合或复杂度混合的情况下，除了 Naïve Bayes 和 Bayes Net 这两种方法之外，其他的传统学习方法 kNN、J48 等方法均达到或高于简单应用 χ^2 方法进行判定的准确程度。图 4 表示了传统机器学习方法的最好结果与简单 χ^2 方法的最好结果的差值。

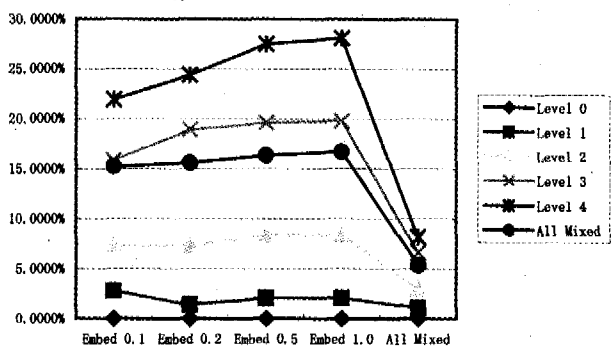

图 4 机器学习方法的结果与简单 χ^2 方法结果的差值

由图 4 可以看出，图像本身的复杂程度越高，机器学习方法优于简单 χ^2 方法的程度就越高。由于图像分布集中于 Level2～4 之间，因此机器学习方法在一般情况下要比简单 χ^2 方法更为有效。这样我们就可以得出初步结论，机器学习方法的引入，能比较有效地提高 χ^2 方法的分类准确率。

3.2.2 对实验数据的进一步分析

首先来看简单 χ^2 方法，由表 2 我们可以看出，简单的 χ^2 方法对判定正例(也就是图像中存在隐藏信息的情况)比较有效，而对反例的正确判断率远远低于正例，也就是会出现大量的伪正例。我们同时可以看到提高该方法的阈值可以减少伪正例的出现，但是正例的判定正确率会有少许下降。我们通过引入机器学习的方法，可以减少单纯利用 χ^2 系数阈值的局限性，从而提高判定的精确程度。

表 2 chi99%和 chi95%方法中真正例和真反例的判断正确率

	Level 0	Level 1	Level 2	Level 3	Level 4	All Mixed
Chi 95% Truc Positive	100.0000%	99.3056%	99.7854%	100.0000%	100.0000%	99.8989%
Chi 99% Truc Positive	96.5909%	96.1806%	98.9270%	99.8500%	99.8457%	99.2922%
Chi 95% Truc Positive	100.0000%	95.8333%	81.1159%	55.0000%	35.8025%	61.9818%
Chi 99% Truc Positive	100.0000%	95.8333%	83.2618%	60.4000%	43.8272%	66.5319%

另外我们注意到，对简单应用 χ^2 系数的方法，对嵌入率汇总的趋势与机器学习方法不同，汇总后的结果高于前面单独每个等级的值(如图 5)。我们认为这样的现象是由于 χ^2 方法的特点和数据集的构成决定的。由于不同的嵌入率下 χ^2 方法正例和反例的判定正确率均基本相同(没有列出)，且

正例的判定正确率几乎是100%，反例的判定正确率小得多，所以最终的结果取决于数据集内反例的比率。由数据集的构造，嵌入率的汇总中反例的比例只有20%，比单独等级中的反例比例50%小，所以会出现汇总时结果正确率提高的现象。

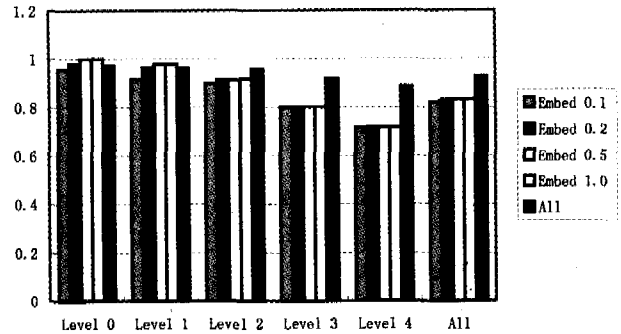

图5 不同图像复杂度下简单Chi方法的结果比较

我们发现Naive Bayes和Bayes Net(数据未列出)的精度结果明显低于其他的方法，我们认为这和具体算法有关，在Weka软件包的算法中对连续数据分布拟合使用了正态分布，但是通过观察原始数据我们认为 χ^2 方法所得出的系数并不符合以平均值居多的正态分布，而是倾向于两端分布，平均值的出现频率比较少。这样Bayes方法计算后验概率的误差就会很大。所以Bayes方法在本例中不是一个非常合适的方法。

我们观察J48方法的结果(也就是算法生成的决策树)发现，对不同的嵌入率，决策树所得出的规则的根节点，往往就是在当前嵌入率的边界系数上所作的判断。比如，对于嵌入率10%(顺序嵌入)的话，那么根节点往往就是对第10个系数(RGB任意一个位平面均可)所作的判断。所以，我们发现如果需要一种简单而又有效的基于 χ^2 系数的顺序嵌入检测方法的话，可以定义一个能被检验出来的最小嵌入率，然后在此边界上对系数进行判断即可。

当图像的复杂度和嵌入率变化时，可以看出传统的学习方法kNN、J48、SMO和BP方法都达到了很不错的效果，对不同复杂度和不同嵌入率的精度变化都很平稳。变化趋势是随嵌入率的升高，不同复杂度之间的差异减小，且准确度升高。我们从数据可以看出嵌入率的变化对简单应用 χ^2 系数的方法的结果几乎没有影响。所以我们可以得出结论，简单应用 χ^2 系数的方法主要由图像本身的复杂程度决定，和嵌入比例的关系比较弱。我们还发现kNN在相同复杂度下的变化比其他机器方法大，我们推断，这可能与kNN方法的内禀随机性有关，因为随机数的不同对这类方法可能有很大的影响。

结论和展望 我们从前面的数据结果和分析中可以看出，机器学习方法的应用对顺序LSB密写的 χ^2 系数密写分析是相当有效的，比较好地弥补了原始 χ^2 系数密写分析的缺陷。但是问题同样存在，我们没有分析当LSB为非顺序嵌入时的结果，以及应该如何对其他的密写方法应用机器学习来构造有效的特征量。同样机器学习的方法应用在不同密写分析中的效果也值得进一步深入的探讨。总之，要将机器学习进一步推广到其它一般的密写分析中的方法，还需要进一步的讨论和研究。

参考文献

1 Petitcolas F, Anderson R J, Kuhn M G. Information Hiding: A Survey. In: Proceedings of the IEEE, special issue on Protection of Multimedia Content, 1999,87:1062～1078

2 Berg G, Davidson I, Duan M Y, et al. Searching for Hidden Messages: Automatic Detection of Steganography. In: Riedl J, Hill Jr. RW, eds. Proceedings of the Fifteenth Conference on Innovative Applications of Artificial Intelligence. Acapulco, Mexico: AAAI, 2003. 51～56

3 Di Y M, Liu H, Ramineni A, Sen A. Detecting Hidden Information in Images: A Comparative Study: [Technical Report]. PPDM, 2003

4 Fridrich J J, Goljan M. Practical Steganalysis of Digital Images: State of the Art. Proceedings of the SPIE Photonics West, Security and Watermarking of Multimedia Contents, 2002,4675: 1～13

5 王朔中，张新鹏，张开文. 数字密写和密写分析——互联网时代的信息战技术. 北京：清华大学出版社，2005

6 Celik MU, Sharma G, Saber E, Tekalp AM. Reversible Data Hiding. In: Proceedings of the 2002 International Conference on Image Processing (ICIP 2002), 2002,2: 157～160

7 Böhme R, Westfeld A. Exploiting Preserved Statistics for Steganalysis. In: Fridrich JJ, ed. Information Hiding 6th International Workshop (IH 2004) Revised Selected Papers, Lecture Notes in Computer Science Vol. 3200. Toronto, Canada: Springer, 2004. 82～96

8 Hsu C T, Wu J L. Hidden Digital Watermarks in Images. IEEE Transactions on Image Processing, 1999, 8(1): 58～68

9 http://www.outguess.org/

10 Westfeld A. F5-A Steganographic Algorithm. In: Moskowitz, ed. Proceedings of Information Hiding 4th International Workshop (IHW 2001), Lecture Notes in Computer Science Vol. 2137. Pittsburgh, PA, USA: Springer, 2001. 289～302

11 Sallee P. Model-Based Steganography. In: Kalker T, Cox IJ, Ro YM, eds. Digital Watermarking Second International Workshop (IWDW 2003) Revised Papers, Lecture Notes in Computer Science Vol. 2939. Seoul, Korea: Springer, 2004. 154～167

12 Fridrich J J, Goljan M, Soukal D. Perturbed Quantization Steganography. ACM Multimedia&Security Journal, 2005, 11(2): 98～107

13 Avcibas I, Memon N, Sankur B. Steganalysis Using Image Quality Metrics. IEEE Transactions on Image Processing, 2003, 12 (2): 221～229

14 Lyu S, Farid H. Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines. In: Petitcolas FAP, ed. Information Hiding 5th International Workshop (IH 2002) Revised Papers, Lecture Notes in Computer Science Vol. 2578. Noordwijkerhout, The Netherlands: Springer, 2003. 340～354

15 Lyu S, Farid H. Steganalysis Using Color Wavelet Statistics and One-Class Support Vector Machines. In: Delp EJ, Wong PW, eds. Security, Steganography, and Watermarking of Multimedia Contents VI, Proceedings of SPIE Vol. 5306. San Jose, California, USA: SPIE, 2004. 35～45

16 Fridrich J J. Feature-Based Steganalysis for JPEG Images and Its Implications for Future Design of Steganographic Schemes. In: Fridrich JJ, ed. Information Hiding 6th International Workshop (IH 2004) Revised Selected Papers, Lecture Notes in Computer Science, Toronto, Canada: Springer, 2004,3200: 67～81

17 Westfeld A, Pfitzmann A. Attacks on Steganographic Systems. In: Pfitzmann A, ed. Proceedings of Information Hiding Third International Workshop (IH'99), Lecture Notes in Computer Science Vol. 1768. Dresden, Germany: Springer, 2000,1768: 61～76

18 Stanley C A. Pairs of Values and the Chi-squared Attack. MSCC, Department of Mathematics, Iowa State University, 2005