

Steganalysis of LSB Matching Based on Mean Co-occurrence Matrix of the Sum Features Using Image Estimate

Qian XIAO, Xiangwei KONG, Bo WANG

Information Security Research Center, Dalian University of Technology, 116024, Dalian, China

Abstract: According to LSB matching, from the co-occurrence matrix, excavates the relevance of pixels of the LSB matching stego image, a new steganalysis LSB matching scheme for gray image was proposed in this paper. In order to strengthen the difference between the cover image and the stego image to increase the performance of the steganalysis, a total of 16 dimensional mean co-occurrence feature vectors extract separately from the test image and the estimate image are used to steganalysis with the FISHER linear classification. Experimental results prove that the detection accuracy of this method is above 90% with the embedded rate of 25%, even when the embed rate of 10%, the detection accuracy reaches 80%, and compares with other methods.

Key words: LSB matching; steganalysis; co-occurrence matrix; image estimate

利用图像估计的平均共生矩阵和特征的 LSB 匹配隐密分析

肖倩, 孔祥维, 王波

大连理工大学信息安全研究中心, 大连, 中国, 116024

摘要:针对 LSB 匹配隐密方法, 从灰度共生矩阵的角度分析图像像素之间的相关性, 提出一种基于共生矩阵和的 LSB 匹配的隐密分析方法。为了强化载体图像和载密图像之间的差异, 借助图像去噪恢复载体图像得到相应的估计图像, 分别从待测图像和估计的图像中提取共 16 维平均共生矩阵和特征, 结合 FISHER 线性分类器进行分类检测。实验结果表明嵌入率在 25% 以上时, 检测率达到 94%; 即使嵌入率在 10% 左右, 检测率也到达 80% 左右, 比其他方法提高 10% 左右。

关键词: LSB 匹配; 隐密分析; 共生矩阵; 图像估计

引言

随着互联网的不断普及和壮大, 人们的生活每时每刻都在与互联网发生各种联系, 使得互联网上的信息安全变得举足轻重。加强对网络上信息安全的控制, 是保证网络正常运行, 维护国家安全以及保护个人隐私的重要保证。数字隐密技术作为信息隐藏的一个重要分支, 隐藏了秘密信息存在, 增强了信息的安全性, 同时给网络上信息安全的监管增加了难度。数字隐密分析技术作为数字隐密技术的攻击技术, 研究隐密分析为监管网络安全, 截获不安全的数字多媒体信息提供了很好的检测手段。

数字隐密技术的载体种类繁多, 例如文本、音频、视频、图像等等。由于图像具有较大冗余量、较小存储容量的特点, 被广泛用于隐密技术。针对图像的隐密方法繁多, 涉及空域、频域、变换域等等。而空域隐密方法因操作简单, 信息嵌入量大的特点备受关注, 尤其是基于 LSB 的改进算法——LSB 匹配隐密方法成为业界的研究热点。

有关 LSB 匹配隐密分析的研究从 2003 开始, Harmse 等提出图像直方图特征函数质心

$C[H] = \sum_{k=0}^{N/2} k |H(k)| / \sum_{k=0}^{N/2} |H(k)|$, 根据载体图像的 $C[H]$ 大于等于载密图像的 $C[H]$ 的性质进行隐密判断, 但是这种方法只对彩色 BMP 图像有效, 而对于灰度 BMP 图像则失效。此后 Ker^[1]等对 Harmen 的方法提出了两点改进: 一、采用下采样对待测图像进行处理; 二、用邻接直方图特征代替普通直方图特征, 经过改进后对灰度图像 LSB 匹配在信息满嵌时的检测正确率达到了 96%。文献[2]提出一种基于像素差分相关的分析方法, 将图像直方图、差分直方图的平滑度、梯度能量、像素差分一维和高维统计分布作为分类特征。Fridich 提出一种性能优良的 LSB 匹配的分析方法, 在小波域提取噪声统计共 27 维特征, 称作 WAM 分析方法。WAM 隐密分析方法在信息满嵌时检测正确率高达 99%。然而在低嵌入率时, 这些方法的检测率较低, 本文方法是为了解决低嵌入率检测率低的问题提出的。

1 图像 LSB 匹配隐密模型

2001 年 Toby Sharp 提出了 LSB 匹配的思想, 但是没有正式的提出 LSB 匹配这一概念。2004 年 Andrew D. Ker 在第六届信息隐藏国际会议上正式提出“LSB matching”这一概念。LSB 匹配 (LSB matching) 与 LSB 替换 (LSB replacement) 是基于 LSB 的两种广泛应用的隐密方法。LSB 匹配相对于 LSB 替换而言, 具有更

基金项目: 国家自然科学基金资助项目 (60971095), 中央高校基本科研业务费专项资金资助

高的安全性。它对数据的 LSB 位随机的加减 1，消除了 LSB 替换的对效应，更好的保持相邻像素之间的关联。随机±1 的 LSB 隐密具体匹配嵌入过程可以用式(1)表示：

$$I_s = \begin{cases} 1 & b \neq LSB(I_c) \& I_c = 0 \\ I_c \pm 1 & b \neq LSB(I_c) \& 0 < I_c < 2^t \\ I_c & b = LSB(I_c) \\ 2^t - 2 & b \neq LSB(I_c) \& I_c = 2^t - 1 \end{cases} \quad (1)$$

提取过程则直接提取 LSB 比特，利用相应的密钥解密即可得到秘密信息。自然图像的 LSB 位平面图像类似于随机噪声，0、1 随机分布，概率分别近似为 1/2。LSB 匹配隐密的模型可以表示为式(2)

$$I_s = I_c + \eta \quad (2)$$

η 为嵌入的秘密信息。隐密过程建模为添加噪声的过程，那么相应的分析方法可以建模为去噪的过程。LSB 匹配隐密算法的作用域为空域，直接对图像的像素值进行处理，可以看作是图像经过一个低通滤波器，滤除了高频分量；同时削弱了图像原有的相关性，像素之间的关联减弱。

2 平均共生矩阵和特征的隐密分析

由于隐密算法对图像的平滑区域较边缘区域的改变程度明显，因而可以利用共生矩阵描述图像局部区域像素差异，构造平均共生矩阵和特征作为隐密判别依据。为了强化载体图像和载密图像之间的差异，提高低嵌入率时的检测正确率，采用了结合图像去噪得到的估计图像进行隐密分析的方法。

2.1 图像像素之间相关性分析

在研究自然图像时，研究者把它看作是一个局部区域平稳的信源，因为物体具有相近的电磁波反射特性，使得局部区域内的像素值之间存在很强的相关性。在进行图像分析时，两种假设被承认，一是马尔科夫假设即某一像素的取值与其一定空间邻域像素的取值相关；二是平移不变性假设即邻域内像素的分配不依赖于邻域在图像中的绝对位置。

从图像的内容角度看，一幅有意义有内容的图像，各个像素点的分布都具有一定的规律性，这种规律性构成了图像所呈现出的事物内容。从宏观上看，图像表现为可视的景象，从微观上看，它表现为一系列的点集或是点对。这些像素点对或是具有相同的灰度值抑或具有接近的灰度值。图像的平滑区域中的像素点的像素值存在一个较小的界，平滑区域中像素点对的值接近；而图像中的边缘部分的像素值分布起伏很大，这部分像素点对的值就相差较大。LSB 匹配隐密方法对像素值的改变范围为[-1,0,1]，它对边缘区域的像素点对的值的改变程度相对于原本像素点对的差异不明显，而对平滑区域的点对的改变较明显。因此，在隐密分析时，着重考虑平滑区域的变化。在平滑区域，像素之间的相关性变化突出，更适合作为隐密分析的

特征。对于像素点 $f(x, y) = p$ 的相对平滑的邻域内，其邻域内的像素值分布在 p 附近可以用式(3)表示：

$$\begin{aligned} p - \varepsilon < (f(x, y) | x \in [x - \Delta x, x + \Delta x], \\ y \in [y - \Delta y, y + \Delta y]) < p + \varepsilon \end{aligned} \quad (3)$$

ε 的取值为整数。为了更加准确的刻画隐密前后的差异，本文 ε 取 3。

2.2 灰度共生矩阵

灰度共生矩阵(GLCM)(简称为共生矩阵)是在 1973 年由 Haralick 提出的。最早被应用于纹理特征提取，在纹理分析中具有优越性。在图像隐密分析方面有不少研究者利用共生矩阵作为手段进行隐密分析，Sullivan^[4]等首次将共生矩阵用于扩频隐密分析，选取共生矩阵中的部分元素构成 129 维特征用于分类。G.Xuan^[5]等将共生矩阵用于 LSB 和 DCT 域嵌入分析，选用主对角线和其上方的两条对角线共 1029 个元素作为特征，为了降维采用 CNPCA 分析方法实现分类检测。下面对灰度共生矩阵简要介绍：

GLCM 描述的是在角度为 θ 方向上相距 d 的两个像素，它们的像素值分别为 i 和 j 的频率相关矩阵 $p(i, j, d, \theta)$ 。 θ 为 0° 、 45° 、 90° 和 135° 。GLCM 不仅反映了亮度的分布特性，同时反映了具有同样亮度或接近亮度的像素之间的位置分布特性，包含了关于方向，相邻间隔，变化幅度的综合信息。一幅二维数字图像的像素用 (x, y) 表示，大小为 $M \times N$ ，灰度级别最大为 255，满足一定空间关系的灰度共生矩阵的表达式为式(4)：

$$p(i, j, d, \theta) = \begin{cases} [(x, y), (x + \Delta x, y + \Delta y)] f(x, y) = i, \\ f(x + \Delta x, y + \Delta y) = j, \\ x = 0, 1, 2, \dots, N_x - 1, y = 0, 1, 2, \dots, N_y - 1 \end{cases} \quad (4)$$

定义 1

平均共生矩阵定义为 θ 分别 0° 、 45° 、 90° 和 135° ，距离为 d 的四个灰度共生矩阵和的平均，记为：

$$C_d = (C_{d0} + C_{d1} + C_{d2} + C_{d4}) / 4 \quad (5)$$

将灰度共生矩阵看成是由 N 条对角线元素组成，如图 2 所示。 D_i 表示第 i 条对角线的元素。 $i \in (-255, 255)$ ， i 的大小代表了两个像素值之间的差值。由于本文计算得到的灰度共生矩阵的主对角线上方元素与下方元素是对称的，因而可以只考虑 $i \in (0, 255)$ 的情况。

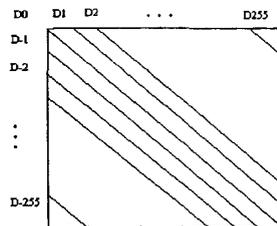


图1 灰度共生矩阵的对角线元素

定义 2:

平均共生矩阵和特征定义为灰度共生矩阵各条主对角线方向的所有元素的和, 记为:

$$G_i = \sum D_i, \quad i = 0, 1, 2, \dots, 255 \quad (6)$$

式(6)中的 i 表示了两像素灰度值之间的绝对差值, 当 i 较小时, 表示两像素灰度值较接近, 这两个像素位于图像的平滑区域, 它们之间的相关性较强; 随着 i 值的不断增大, 这两个像素则位于图像的边缘区域, 它们之间的相关性减弱, 图像隐密前后的改变在边缘区域不显著, 假若是一幅纹理复杂的图像, 边缘信息可能会对隐密特征造成干扰。因此结合公式(3), 式(6)改写为式(7):

$$G_i = \sum D_i, \quad i = 0, 1, 2, 3 \quad (7)$$

利用平均共生矩阵和特征对经过 LSB 匹配隐密图像和原始载体图像进行分类检测, 在嵌入率为 100% 时检测率正确率接近 100%, 然而当嵌入率低于 20% 时, 检测率正确率在 70% 左右, 因此误判概率较大。

2.3 估计原始载体图像

以上利用平均共生矩阵和特征的隐密分析方法在低嵌入率时检测正确率较低。为了解决这个问题只有增强低嵌入率载密图像和载体图像之间的差异, 才能够获得较高的检测率。假若能够较好的从载密图像中还原出载体图像, 借助于还原的载体图像去检测隐密必然能达到很好的效果。这里引进了一种能够保留图像细节的图像滤波技术——自适应中值滤波复原^[6]。具体实现步骤有以下 4 步:

- (1) 确定点 $P(x, y)$ 为中心大小为 $N \times N$ 区域, 标记最小值 Z_{min} , 最大值 Z_{max} , 中值 Z_{med} ;
- (2) If $Z_{med} - Z_{min} > 0$ && $Z_{med} - Z_{max} < 0$, 转到步骤(3), else 增大窗口的尺寸, 重复步骤 2;
- (3) If $Z_{xy} - Z_{min} > 0$ && $Z_{xy} - Z_{max} < 0$, 输出 Z_{xy} , else 输出 Z_{med} ;
- (4) 逐点循环至结束。

为了验证经过滤波处理的估计图像和原始图像的相似度, 用重构信息错误比特率 BER 表示, BER 为原始图像和估计图像对应像素点具有不同像素值的个数与图像像素点总数的比值。分别对原始载体图像和载密图像进行滤波处理, 载体图像滤波后的 BER 大约在 0.1 左右, 载密图像滤波后的 BER 大约在 0.3 左右。使用估计图像作为辅助图像提取平均共生矩阵和特征, 能够提高隐密分析的检测率。图像估计的准确率是提高隐密检测的关键。

2.4 特征提取

本小节详细列出了本文特征提取方法, 并绘出图像隐密分析检测流程图, 如图 3 所示。

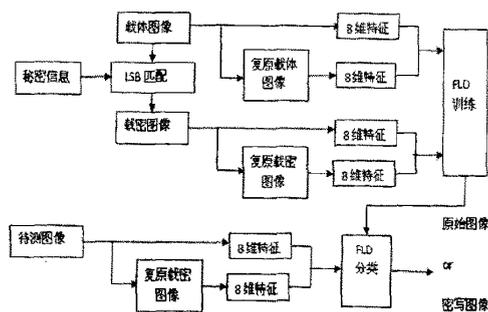


图2 LSB 匹配隐密分析检测流程

第 1 步: 对待检测图像进行自适应中值滤波得到估计图像;

第 2 步: 依据公式(5)对待测图像计算平均共生矩阵 C_1, C_2 , 分别对 C_1, C_2 提取平均共生矩阵和特征 $G_i, i = 0, 1, 2, 3$, 这样得到 8 维特征;

第 3 步: 依据公式(5)对估计图像计算平均共生矩阵 C_1^*, C_2^* , 分别 C_1^*, C_2^* 提取平均共生矩阵和特征 $G_i^*, i = 0, 1, 2, 3$, 这样得到 8 维特征;

第 4 步: 合并前两步的特征, 得到 16 维平均共生矩阵和特征。

第 5 步: 利用 FISHER 线性分类器对特征进行分类。

3 实验结果

为了更好的验证隐密分析方法的检测性能, 实验图像选自未压缩 TIFF 格式的图像, 共 485 幅分辨率为 1440×960 的图像。所有图像转换为灰度图像构成载体图像库, 对图像应用随机 LSB 匹配隐密方法分别嵌入 100%、75%、50%、25%、15%、10% 的秘密信息, 分别组成载密图像库。随机从每个图像库中挑选 200 幅图像作为训练集, 其他用于测试。

为了比较方法的性能, 对训练集中的图像分别提取本文 16 维特征, 未使用估计图像的 8 维特征, WAM27 维特征以及改进的 HCF 特征利用 FISHER 进行分类进行比较。在不同嵌入率和不同虚警率的检测结果通过 ROC 曲线表示。图 4 为嵌入率为 100% 时的不同方法的 ROC 曲线。表 1 详细列出了各个方法在虚警率为 0.1 时的不同嵌入率下的检测率。

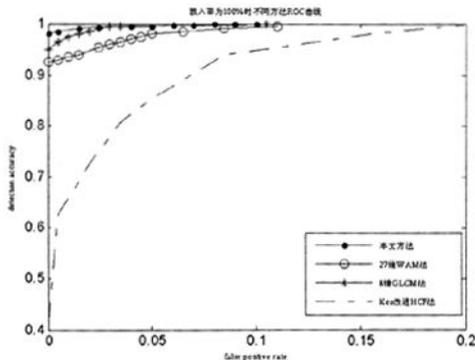


图3 嵌入率为100%时不同方法下的ROC曲线

从图4中可以看出,在信息满嵌的情况下,本文的方法较其他三种方法,在虚警率相同的情况下,隐密检测的正确率是最高的。Ker改进的HCF方法的检测正确率最低。结合图像估计的平均共生矩阵和特征比单一的8维平均共生矩阵和特征具有显著的优势。从表1中可以看出,本文的方法在不同嵌入率的检测率是最高的。特别是在嵌入率低于25%以下时,比Ker改进的HCF方法提高了20%以上,比单一8维共生矩阵和特征提高10%左右,比WAM方法提高了13%左右。因此,本文的方法能够较好的实现LSB匹配隐密检测。总而言之,检测率的提高依赖于图像估计技术的准确率,估计图像越接近原始载体图像,提取的特征越能够体现载密图像与原始载体图像的差异,隐密检测的正确率越高。

表1 不同嵌入率和方法下的检测性能比较

嵌入率%	未经过图像复原	自适应中值复原	WAM ^[9] 方法	改HCF方法
100	99.30	99.65	99.47	96.90
75	98.42	98.95	98.25	96.84
50	97.72	97.89	96.14	81.05
25	91.40	94.39	90.70	62.46
15	76.49	87.72	72.09	60.42
10	73.51	82.81	70.65	58.42

4 结论

本文从图像空域的相关性出发,将图像看作局部平稳的信源,基于LSB匹配模型,提出一种结合图像估计的平均共生矩阵和特征的隐密检测方法。实验证明,本文在满嵌时,正确率接近100%;在低嵌入率的情况下,也达到了80%以上。

参考文献

- [1] Ker A D. Steganalysis of LSB matching in grayscale images [J]. IEEE Signal Processing Letters, 2005, 12(6): 441-444.
- [2] T Zhang, W Li, Y Zhang, et al. Steganalysis of LSB matching based on statistical modeling of pixel difference distributions[J]. Information Sciences, 2010,180(23):4685-4694.
- [3] M Abolghasemi, H Aghaieinia, K Faez, et al. Detection of LSB ± 1 steganography based on co-occurrence matrix and bit plane clipping[J]. Journal of Electronic Imaging 19(1) (Jan-Mar 2010).
- [4] K Sullivan, U Madhow, S Chandrasekaran. Steganalysis of spread spectrum data hiding exploiting cover memory[C]. SPIE,2005, 5681, :38-46.
- [5] G Xuan, Y Q Shi, C Huang, et al. Steganalysis using high-dimensional features derived from co-occurrence matrix and class-wise principal components analysis CNPCA[C]. IWDW2006.
- [6] 王晓凯,李峰.改进的自适应中值[J].计算机工程与应用. 2010,46(3):175-176. Wang Xiao kai, Li Feng. Improved adaptive median filtering[J]. Computer Engineering and Applications,2010,46(3):175-176.

作者简介

孔祥维(1963-),女,博士生导师,E-mail:kongxw@dlut.edu.cn