

doi:10.3969/j.issn.1673-9833.2013.05.020

基于统计特征的二值图像信息隐藏检测

滕峻林, 云 航

(四川工程职业技术学院, 四川 德阳 618000)

摘 要: 目前在各种形式的数字信息传输中加强信息的保密和安全性问题日益引起人们的关注, 考虑信息隐藏技术是为信息传递者将私有信息嵌入掩饰信息中来传递消息, 因此对隐藏信息的检测在信息安全方面非常重要。为了检测二值图像的隐藏信息, 提出一种基于统计特征的二值图像信息隐藏检测方法。该方法利用二值图像的统计特征, 找出符合二值图像信息检测需要的统计特征进行隐藏分析, 通过不同的隐写方法对统计特征的影响确定用不同的特征对具体的隐写方法进行隐藏信息的检测, 从而保持对载体图像含有隐藏信息检测的准确性。试验结果表明: 所提出的方法检测率在90%以上, 可解决当前检测系统对于海量数据的安全性测试分析。

关键词: 统计特征; 二值图像; 信息隐藏; 信息检测

中图分类号: TP309

文献标志码: A

文章编号: 1673-9833(2013)04-0090-05

Binary Image Information Hiding Detection Based on the Statistical Characteristics

Teng Junlin, Yun Hang

(Sichuan Engineering Technical College, Deyang Sichuan 618000, China)

Abstract: At present the problem to strengthen information confidentiality and security in various digital information transmission increasingly causes the attention of people. The information hiding technology transfers message by means of embedding private information in concealing information, so the hidden information detection is very important in the field of information security. In view of binary image information detection, presents a binary image information hiding detection method based on statistical characteristics. The method applies the statistical characteristics of binary image to find those in accordance with binary image information detection needed for hiding analysis, and through the influence of different steganalysis method on statistical characteristics, determines different characteristics applied for hidden information detection of specific steganography, thus maintain the detection accuracy of carrier image with hidden information. The experimental results show that the proposed method has the detection accuracy above 90% and can be used for the current test system in security testing analysis of huge amounts of data.

Keywords: statistical characteristics; binary images; information hiding; information detection

0 引言

随着计算机和通信网技术的发展与普及, 信息

媒体的数字化以及其他电子出版物的传播和交易变得越来越便捷, 人们借助于计算机、数字扫描仪、打印机等电子设备, 可以方便、迅捷地实现数字信息

收稿日期: 2013-03-23

基金项目: 中央企业电动车产业联盟共性技术基金资助项目(JS-314)

作者简介: 滕峻林(1984-), 男, 湖南麻阳人, 四川工程职业技术学院教师, 硕士, 主要研究方向为智能控制理论及其工程应用, E-mail: junlinteng@163.com

的传输。与此同时,如何加强信息的保密和安全问题也日益引起人们的关注,特别是军方、各级政府和情报部门很重视隐秘通信和信息隐藏技术。信息隐藏技术就是传递者将私有信息嵌入表面上看来无关紧要的掩饰信息中进行传送的各种方法^[1],其主要目的是避免传送秘密消息的时候引起监测者的怀疑,从而达到在正常、合法的渠道下传递“隐蔽”信息的目的。对于网络和传送渠道的监控者和信息安全与保密的从事者而言,希望可以判断出在海量的信息传输中,是否存在带有某些“隐秘”信息的文件被用于传递危及网络和其他保密信息的安全^[2-3]。信息检测技术就是根据载体的已知特征,通过一定的算法判断出媒质中是否含有隐藏信息。

本文针对二值图像隐藏信息检测,在分析不同的隐写方法对统计特征的不同影响基础上,提出一种基于统计特征的二值图像信息隐藏检测方法,通过实验验证了算法的有效性。

1 二值图像信息检测

信息隐藏的检测是目前学术界研究的一个热点,目前的检测方法主要基于灰度图像或者彩色图像,而对二值图像的信息隐藏检测的研究较少。通过对二值图像隐藏算法的研究发现:任何信息的嵌入,都会改变图像的某些自然特征,这些特征可以表现为图像的统计特征^[4-7],这些图像统计特征又是能很方便地检测出来的。首先对二值图像的统计特征进行分析,然后找出符合二值图像信息检测需要的统计特征进行隐藏检测。

二值图像统计特征有很多种,这里只对二值图像信息隐藏敏感的特征进行讨论。首先讨论的几种特征都是用于信息隐藏(二值图像信息隐藏或灰度图像信息隐藏),在信息隐藏后,这些统计特征一般都会发生改变。下面介绍用于二值图像信息隐藏分析的一些统计特征。

1.1 黑白边界复杂度

黑白边界复杂度最早应用于位平面复杂度分割(bit-plane complexity segmentation, BPCS)信息隐藏^[4-7],BPCS用这种特征区分哪些图像块可以用来嵌入隐秘信息。信息隐藏后,一般会带来二值图像噪声,这种噪声特征其实就反映了二值图像信息隐藏的存在的可能。

黑白边界复杂度的定义如下:在图像平面上定义一个图像块,这种图像块可以是 8×8 或者其他大小,然后统计图像块中黑白像素变化数,统计的方向是先水平后垂直进行的。最后黑白边界复杂度由

下式计算

$$\alpha = \frac{k}{\max\{I(x,y)\}}, \quad (1)$$

式中: k 为总的黑白像素变化数; $\max\{I(x,y)\}$ 是所有可能黑白像素变化数中的最大值。这样黑白边界复杂度 α 的范围是 $0 \leq \alpha \leq 1$ 。

1.2 游程不规则度

游程不规则度是Hirohisa^[3]提出的,最早用于BPCS的信息隐藏。如果黑白像素的分布具有周期性的特点,使得使用黑白边界复杂度的度量值很高,这种图像块可能被列为噪声区域,在信息隐藏检测中认为这样的噪声区域可能有隐秘信息存在,但实际上可能只是二值图像的细节部分。为了避免这种情况,采用Hirohisa的游程不规则度来作为信息隐藏检测的另一种可行的统计特征。游程不规则度主要度量黑白像素的非均匀分布情况,当游程不规则度的度量值很高的时候,那么图像中周期性的黑白变化和全部是黑色像素或者白像素的情况就不会存在。游程的频率定义为

$$p_i = \frac{h[i]}{\sum_{i=1}^n h[i]}. \quad (2)$$

式中: n 为最长的游程; $h[i]$ 表示游程出现的个数。

游程的熵定义为

$$h_s = -\sum_{i=1}^n h_i \log_2 p_i. \quad (3)$$

令图像块的大小为 $n \times n$, r_i 和 c_j 为图像块中的第 i 行和第 j 列。色像素或者白像素的情况就不会存在。游程不规则度 β 定义为

$$\beta = \min\{\overline{h_s(r)}, \overline{h_s(c)}\}, \quad (4)$$

式中: $\overline{h_s(r)} = \{\overline{h_s(r_0)}, \dots, \overline{h_s(r_{n-1})}\}$;

$$\overline{h_s(c)} = \{\overline{h_s(c_0)}, \dots, \overline{h_s(c_{n-1})}\};$$

$$\overline{h_s(r_i)} = h_s(r_i) / \max(h_s(r_i));$$

$h_s(\bullet)$ 是所有 $h_s(\bullet)$ 元素的平均值。

1.3 图像块噪声

图像块噪声也是Hirohisa^[5]提出来的用于BPCS的信息隐藏。图像块噪声 γ 定义为

$$\gamma = \frac{1}{n} \min\{E_f(p_x(r)), E_f(p_x(c))\}, \quad (5)$$

式中: $p_x(r) = \{\rho(r_0 \oplus r_1), \dots, \rho(r_{n-2} \oplus r_{n-1})\}$,

$$p_x(c) = \{\rho(c_0 \oplus c_1), \dots, \rho(c_{n-2} \oplus c_{n-1})\}, \quad (7)$$

其中, \oplus 表示异或, $\rho(x)$ 统计是二值序列 x 中1的个数的函数;

$$E_f(x) = \left(1 - V(X) / \max\{V(X)\}\right) \bar{X}, \quad (8)$$

其中, $X = \{x_0, x_1, \dots, x_{m-1}\}$, $V(X) = D(X)$, $\bar{X} = E(X)$ 。

图像块光滑度是建立在邻域光滑度的基础上的, 最早由 $wu^{[2]}$ 提出出来的一种度量二值图像复杂度的方法。邻域光滑度是沿不同的方向定义的, 定义如下。

水平方向:

$$N_h(i, j) = \sum_{k=-1}^1 \sum_{l=-1}^0 I\{\{p_{i+k, j+l} \neq p_{i+k, j+l+1}\}\}; \quad (9)$$

垂直方向:

$$N_v(i, j) = \sum_{k=-1}^1 \sum_{l=-1}^0 I\{\{p_{i+l, j+k} \neq p_{i+l+1, j+k}\}\}; \quad (10)$$

对角线方向:

$$N_{d1}(i, j) = \sum_{k=-1}^0 \sum_{l=-1}^0 I\{\{p_{i+k, j+l} \neq p_{i+k+1, j+l+1}\}\}; \quad (11)$$

反对角线方向:

$$N_{d2}(i, j) = \sum_{k=0}^1 \sum_{l=-1}^0 I\{\{p_{i+k, j+l} \neq p_{i+k-1, j+l+1}\}\}; \quad (12)$$

以上式中 $I(\cdot)$ 是一个是非判断函数, 取值范围是 $\{0, 1\}$, $p_{i, j}$ 为图像中的第 i 行和第 j 列的像素值。因而图像块光滑度 η 定义为:

$$\eta = N(i, j) / \max_{i, j} (N(i, j)), \quad (13)$$

其中,

$$N(i, j) = \sum_{l=0}^{row-1} \sum_{j=0}^{col-1} (N_h(i, j) + N_v(i, j) + N_{d1}(i, j) + N_{d2}(i, j)),$$

式中 row , col 分别是图像块的高度和宽度。

1.4 图像块粗糙度

图像块粗糙度主要用于定义图像的边界不规则性, 是通过对二值图像信息隐藏后对图像边界的影响性而提出的一种测量方法。主要通过定义图像块中使图像边界显得粗糙的像素数和图像块中总的像素数比值来进行度量。粗糙度指检查像素和周围 8 连通像素联结情况的一种度量, 定义为

$$\xi = \sum_{i=1}^{row-2} \sum_{j=1}^{col-2} I(i, j) / (row \times col). \quad (14)$$

2 二值图像信息隐藏检测

对于扫描图像、传真文档资料或者卡通图片等二值图像, 仍然有各种视觉冗余空间, 可以利用这些冗余来进行信息隐藏, 而不会引起人眼的视觉上的差异。但任何信息隐藏都会引起一些图像特性发生改变, 前面所分析的各种二值图像特征可能发生改变, 从而可以用这些特征来进行信息隐藏检测。如大多数扫描图像, 传真的文档资料或者卡通图片这

种二值图像黑色区域和白色区域具有比较清晰的边界, 这种边界可以使用前面的二值图像统计特征来度量, 当信息隐藏后, 这种统计特征值将显著增大, 可以利用这种二值图像特征来区分原始图像和嵌入了信息的图像。

广义地定义原来二值图像特征为 U , 信息隐藏于二值图像后, 图像特征的值增大为 U' 。这样, 可假定原始图像的 U 肯定小于嵌入信息后的图像的统计特征值 U' 。但在实际应用中发现, 二值图像千差万别, 有的原始图像 U 的值甚至比嵌入信息后的 U' 都要大, 很难找到一个能区分所有二值图像为原始图像还是嵌入信息图像的阈值。然而 Fridrich 等人^[4]发现, 如果二值图像中已经隐藏了隐秘信息, 第二次嵌入信息后, U 的变化不是很明显; 而对原始图像, 嵌入隐秘信息后, U' 变化很显著。可以使用对 U 的相对比值作为判断标准, 设定阈值, 用来区分原始图像和嵌入了信息的图像。

信息隐藏检测方法大致如下: 对于任何给定的二值图像, 首先把图像分成 8×8 的图像块, 分块计算其二值统计特征 $(\alpha, \beta, \gamma, \eta, \xi)$, 然后, 对所有的图像块的统计特征求均值, 均值就是前文所说的 U 。

接下来, 通过二值图像信息隐藏方法, 对这些图像嵌入一定量的隐秘信息 (这些隐秘信息一般为随机二值序列)。然后重新采用第一步所说的计算方法计算图像统计特征, 得到 U' 。这样, 如果图像已经包涵了隐秘信息, U 和 U' 应该大致相等; 如果图像不包涵隐秘信息, U' 一定会比 U 大很多。采用比值 R 来区分这两类图像。如果图像已经嵌入了隐秘信息, R 大致等于或者略微大于 1 (这时候用 R_s 表示 R); 而如果图像中没有隐秘信息, 一定比 1 大得多 (这时候用 R_0 表示 R)。如此一来, R 可以看成一条件随机变量, 如果图像已经嵌入了隐秘信息, 均值是 R_s ; 而如果图像中没有隐秘信息, 均值是 R_0 。可以使用假设检验的方法来判断图像中是否含有隐秘信息。假设检验的零假设和备择假设分别是:

H_0 图像中没有隐秘信息;

H_1 图像中已经嵌入了隐秘信息。

为了判断 R 是否是所要的判断标准, 假设 R 符合正态分布, 采用 t 检验法来检测。对图像库 (图片是从 IFNENIT 网站上下载的, 共有 5 578 幅图像, 包括各种卡通、人物以及扫描文件图像等) 中图片先使用前面介绍的 3 种二值图像信息隐藏的方法进行了隐秘信息嵌入, 得到含隐秘信息的图片库。在给定显著水平 $\alpha=0.05$ 的情况下, 发现 R 值有很好的区分性能。因篇幅所限, 这里不详细讨论检验过程。

在实际的信息隐藏检测中, 可以对检测量设定一个阈值 T , 如果 $R > T$, 认为图像中没有隐秘信息; 否则就认为图像是“干净”的图像, 没有隐秘信息存在。当然判断中也存在 2 类错误:

Error 1 图像中没有信息而错误地判断有信息;

Error 2 图像中有信息而没有发现。

阈值 T 可以根据等错误率而设定为

$$T = (R_0\sigma_s + R_s\sigma_0) / (\sigma_0 + \sigma_s), \quad (15)$$

式中: σ_0 是图像中没有隐秘信息时 R 的标准差; 而 σ_s 为图像已经嵌入了隐秘信息后 R 的标准差。当然在具体的信息隐藏检测中, 也可以根据自己的实际情况, 调整阈值 T 的大小, 从而适应不同的信息隐藏情况。

3 试验结果

试验所用数据库来源网络下载的二值图像^[8]。图像文件格式是 GIF 格式。图 1 显示的是样本图像。



图1 测试数据库样本图像

Fig. 1 Test database sample images

试验中选取了 3 种二值图像信息隐藏来测试的算法, 包括 ZK 信息隐藏嵌入方法、CPT 方法 (correlated phase threshold, 相粒相关阈值方法) 和游程编码嵌入方法。

ZK 信息隐藏方法是基于图像块中白色像素与整个图像块大小的比值的隐藏方法。嵌入过程中黑白对比很大的边界部分和杂乱的区域得到充分的考虑, 嵌入后图像质量很好。信息隐藏检测方法如下:

首先, 对图像库中每幅图像都用 ZK 信息隐藏方法嵌入隐秘信息, 信息容量为全部图像信息容量, 即对所有的数据块都嵌入隐秘信息, 得到一个信息图像库。接着在原始图像库和新的图像库上, 采用 ZK 信息隐藏方法嵌入同时再次嵌入隐秘信息, 容量大致为 2/3 总的嵌入容量。嵌入后分别采用上面所叙述的信息隐藏检测方法进行信息隐藏检测。

检测对所有的图像统计特征进行, 得到试验结果如图 2 所示, 图中 ‘o’ 隐藏了信息的图像库中每幅图像的统计特征值 R_s , ‘x’ 代表的是原始图像库中每幅图像的统计特征值 R_0 (图中所示的统计特征是黑白边界复杂度)。表 1 为各种不同的统计特征情况

下计算的阈值和第一类、第二类错误率。从表中所见, 第二类错误相对于第一类错误较小, 这也说明对于一般二值图像来说, 统计特征值变换是比较大的, 但嵌入信息后, 其特征值都会变得相对原始图像来说要大很多。

表 1 ZK 算法检测的结果

Table 1 ZK algorithm testing results

统计特征	阈值	第一类错误	第二类错误
块平滑度	1.028 4	0.244 1	0.110 2
块粗糙度	1.081 3	0.270 1	0.031 5
黑白像素比例	1.084 3	0.173 1	0.047 2
块不规则度	1.120 5	0.204 7	0.078 7
块噪声	1.130 7	0.126 0	0.078 7

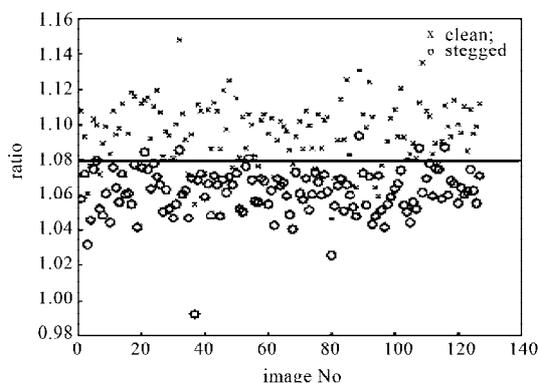


图 2 ZK 算法的检测

Fig. 2 ZK Algorithm detection

第二个试验对 CPT 信息隐藏方法进行检测。CPT 方法也是基于图像块的。把一幅二值图像分成大小为 $m \times n$ 的图像块, 应用 CPT 方法可以最多嵌入 r 位长度的隐秘信息, 其中 $2^{r+1} \leq mn$ 。而对每一图像块最多改变 2 个 bit, 因而这种嵌入方法对人眼的视觉效果改变很小, 但也会引起一些统计特征的变化。对这种信息隐藏方法的检测与前面 ZK 的检测一样, 首先, 对图像库中每幅图像都用 CPT 信息隐藏方法嵌入隐秘信息, 信息容量为全部图像信息容量, 即对所有的数据块都嵌入隐秘信息, 得到一个信息图像库。然后在这 2 个库, 对所有的图片分别计算各种二值图像特征的值, 计算对应的比值。试验结果显示为图 3 和表 2。试验表明, 检测方法可以发现所有的嵌入了隐秘信息的图像。

表 2 CPT 算法检测的结果

Table 2 CPT algorithm testing results

统计特征	阈值	第一类错误	第二类错误
块平滑度	1.873 0	0.157 0	0.196 9
块粗糙度	1.501 2	0.015 7	0.000 0
黑白像素比例	1.954 3	0.078 7	0.000 0
块不规则度	1.914 1	0.047 2	0.000 0
块噪声	1.873 0	0.157 8	0.193 9

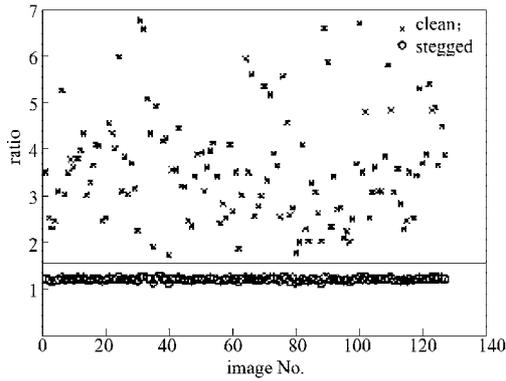


图3 CPT算法的检测

Fig. 3 CPT algorithm detection

第三个试验对基于游程编码的信息隐藏方法进行检测。首先统计图像中游程的长度,由于信息隐藏改变的是游程的长度,改变的象点都处于图像的黑白边界部分,对人眼的视觉影响很小,对图像的黑色或者白色区域没有变化,图像质量很好,非常适合卡通图像的信息隐藏。但对于大片的黑色或者白色区域的图像来说,游程数目非常少,因而嵌入容量很小。

对这种信息隐藏方式进行检测。由于这种信息隐藏方式只改变边界区域,对于前面提到的图像统计特征如黑白边界复杂度、游程不规则度或者图像块噪声的影响很小,但图像块光滑度和图像块粗糙度会发生较大的变化。试验中采用图像块粗糙度来检测信息隐藏。试验结果见图4。

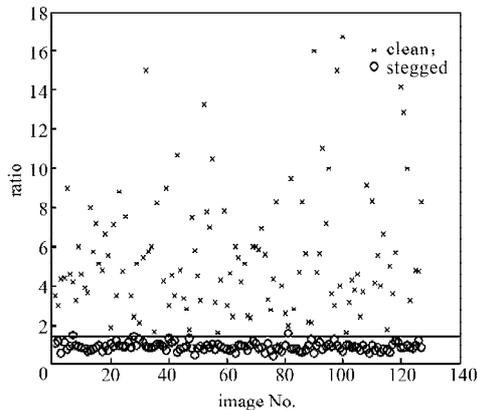


图4 游程编码算法的检测

Fig. 4 The run-length encoding algorithm dection

试验中使用的阈值是1.7652,第一类和第二类错误都大致是0.0157。从实验结果发现,基于二值图像统计特征的信息隐藏检测分析方法是非常有效的,对于基于图像块的信息隐藏方法来说,应用各种二值图像统计特征检测方法的检测可靠性都很高;而

对于基于边缘的方法(基于游程编码),图像粗糙度是个较好的二值图像统计量,当这种统计量与二值图像分解相结合,甚至可以确定信息隐藏所试验的具体的隐藏方法。

4 结论

在分析当前信息隐藏技术中普遍使用的几种二值图像信息典型隐藏算法的原理和性能的基础上,针对几种算法隐藏的二值图像,提出一种基于统计特征的二值图像隐藏检测算法,通过分析不同的隐写方法对统计特征的影响确定用不同的特征对具体的隐写方法进行隐藏信息的检测。该算法在大量的试验证明下是一种行之有效的方法,它在判决信息隐藏与否环节的正确率能够达到90%以上。可解决当前检测系统对于海量数据的安全性测试分析。

参考文献:

- [1] Lampso B W. A Note on the Confinement Problem[J]. Communications of the ACM, 1973, 16(10): 613-615.
- [2] Wilkins J. Mercury, Or, The Secret and Swift Messenger: Sewing, How a Man May with Privacy and Speed Communicate His Thoughts to a Friend at Any Distance [M]. London: [s.n.], 1707: 141-143.
- [3] Murray A H, Burchfield R W. The Oxford English Dictionary: Being a Corrected Re-Issue[M]. Oxford, England: Clarendon Press, 1933: 1228-1240.
- [4] Kawaguchi E, Eason R O. Principle and Applications of BPCS-Steganography[J]. In Proceedings of SPIE: Multimedia Systems and Applications, 1998(3528): 464-472.
- [5] Tseng Y C, Pan H K. Secure and Invisible Data Hiding in 2-Color Images[C]//Proceedings of INFOCOM 2001, Anchorage: IEEE Press, 2001: 887-896.
- [6] Fridrich J, Long Meng. Steganalysis of LSB Encoding in Color Images[C]//In Proceedings of ICME. New York: IEEE Press, 2000: 1279-1282.
- [7] Hirohisa H. A Data Embedding Method Using BPCS Principle with New Complexity Measures[J]. In Proceeding of Pacific Rim Workshop on Digital Steganography, 2002(7): 30-47.
- [8] [ANON]IFN/ENIT-Database[EB/OL].[2012-08-03]. [http : www.ifnenit.com/download.htm](http://www.ifnenit.com/download.htm).

(责任编辑:申剑)