

基于模型的信息隐藏检测研究^{*}

于小亿, 谭铁牛, 王蕴红

(中国科学院自动化研究所, 北京 100080)

摘要: 提出一种模型的信息隐藏检测方法, 这种方法是对要分析的图像建立一个统计特征模型, 而信息隐藏将改变该特征, 那么将能检测到信息隐藏。对 JSteg 和 QIM 两种信息隐藏为例, 通过对系数进行建模, 估计出原始图像系数分布, 然后将其与实际图像系数分布进行比较, 可以很方便的检测出图像是否隐藏了信息, 并且估计出隐藏的信息量。试验证明, 该方法非常有效。

关键词: 信息隐藏; 信息隐藏检测; DCT; QIM

中图分类号: TP751 **文献标识码:** A **文章编号:** 0529-6579 (2004) S2-0184-05

信息隐藏是往多媒体数据(如图像、声音、视频信号等)中添加要传送的信息而不影响原数据的视听效果, 然后利用公共信道传送出去, 以达到隐蔽通信的目的。与信息隐藏相对应, 信息隐藏检测是指分析互联网上的公开网站和多媒体数据流, 检测可疑的隐蔽通信渠道。“9·11”恐怖袭击以后, 美国媒体先后报道本·拉登利用互联网图片进行隐蔽通信, 引起美国政府的高度重视。在国内, “法轮功”分子和境外“民运分子”利用网页上图像文件隐藏信息。开发出具有实际意义的信息隐藏检测技术日益变得迫切和必要。

近来的文献报道了一些关于信息隐藏检测技术的工作。Johnson 等^[1]提出了基于信息隐藏软件嵌入后所特有的软件同特征来判断是否有信息隐藏。Fridrich^[2]提出了一系列的信息隐藏检测方法, 并对以前的一些常见的信息隐藏检测方法作了很好的总结。通用的信息隐藏检测方法主要有 Memon^[3]的基于图像质量评价和 Farid 等^[4]的基于高阶图像统计量的方法。但针对具体一个或者一类信息隐藏方法, 检测率高, 准确, 但实用性差, 对新的嵌入方法将工作不可靠; 通用盲信息隐藏检测, 能调整到新的方法, 但检测率低。因此如何开发更加通用、准确的信息检测技术仍是一个挑战。

本文试图提出一种新的信息隐藏检测方法, 该方法可以作为信息隐藏检测的一种通用性的框架, 以适应检测各种信息隐藏过程。第二部分对基于模型的信息隐藏检测方法的进行一般性描述; 第三部分对 Jsteg 类和 QIM 类的信息隐藏进行简单说明;

接下来是对该两类信息隐藏的检测, 试验和结论在最后两部分。

1 基于模型的信息隐藏检测方法描述

本文的思想是: 如果对所有的多媒体能建立一个统计特征模型, 而信息隐藏将改变该特征, 那么我们将能检测到信息隐藏。假设 x 是原始载体的一个实例, 可以是图像的灰度值或者各种变换(如 DCT 变换)后的系数。如果我们把 x 看作随机变量 X 的一个实例, 那么随机变量 X 可以用一个概率模型 $P_X(x)$ 来描述。检测一个可疑图像是否含有隐藏信息可以采用假设检验来判断, 如 x 服从 $P_X(x)$, 则该图中不含有隐秘信息, 否则可以认为含有可疑信息。进一步, 在已知嵌入方法的情况下, 可以估计出嵌入信息的容量。估计出嵌入信息的容量的方法如 JPEG 类型 DCT 系数符合高斯分布, 那么该高斯模型就可以用于信息隐藏检测。

Fridrich^[5]总结出估计信息量的技巧可以用在我们的基于模型的信息隐藏检测方法中。Fridrich 的想法是: 对大多数的信息隐藏方法来数, 如果我们能确定一个统计量 $S(m)$, 而这个量将随着嵌入信息量 m 的多少而发生变化, 那么我们就可计算出嵌入信息量的多少。通常来说, $S(m)$ 带有未知参数, 这些参数可以通过边界条件而求解, 如 $S(0)$ (原始载体)。本文对 JPEG 类型文件 DCT 系数进行统计建模, 估计出原始图像系数分布, 然后将其与实际图像系数分布进行比较, 建立一统计量, 该统计量随着嵌入数据发生变化以检测出图像是否

* 收稿日期: 2004-09-11

基金项目: 国家 863 计划资助项目 (2003AA144080); 国家自然科学基金资助项目 (60172054, 60121302, 60335010)

作者简介: 于小亿 (1972 年生), 男, 博士生; E-mail: xyu@nlpr.ia.ac.cn

隐藏了信息，并且估计出隐藏的信息量。试验证明，该方法非常有效。我们将以 JSteg 和 QIM 为例进行说明。

2 JSteg 和 QIM 算法

JSteg 是 JPEG 格式文件中最常应用的信息隐藏软件，是由 Upham^[6] 最早发明，而后很多类似的软件采用该软件类似的信息隐藏方法。信息隐藏的嵌入过程如下：首先将 BMP 格式文件转换成 JPEG 文件，然后将 JPEG 中量化 DCT (discrete cosine transform) 系数的最低位 (LSB, least significant bit) 替换成要嵌入的二进制信息。替换过程中保持系数为 0 和 1 的系数不变，并对直流系数也保持不变。系数的选择有两种方式：一是随机选择，另一方式为顺序选择。提取过程和嵌入过程相反，量化 DCT 系数的最低位就是嵌入的二进制信息。QIM (quantization index modulation)^[7] 是另外一种有效的信息隐藏方法，最简单的信息隐藏方法嵌入过程是这样的。选择量化 DCT 系数，然后 QIM 化，系数的选择与 JSteg 类似。本文所叙述的即采用这种简单的 QIM 方法。

信息隐藏后，对载体 (JPEG 文件) 的量化 DCT 系数将发生改变。图 1 显示的是一个测试图像的 8×8 分块位置 (3, 3) 的 DCT 系数直方图。图 1 (a) 是原始图像的 DCT 系数直方图，图 1 (b) 和图 1 (c) 分别是使用 JSteg 和 QIM 信息隐藏方法后的 DCT 系数直方图。从图上看，系数发生变化很明显。许多方法^[8] 可以检测 JSteg 类信息隐藏，这些方法有的不能估计嵌入的信息量。对于 QIM，仅仅张文^[9] 提出一种有效的检测并破坏信息隐藏的方法。在接下来的，将介绍我们的信息隐藏检测方法，不仅能有效检测信息隐藏，而且能估计信息量的多少。

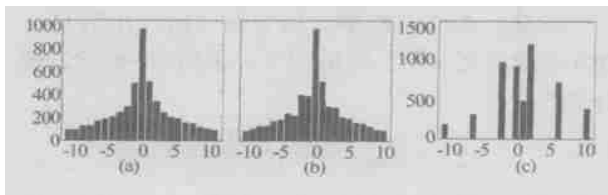


图 1 (a) 原始; (b) JSteg; (c) QIM 直方图

Fig 1 (a) Original; (b) Jsteg; (c) QIM histogram

3 统计模型

一般认为图像经过 DCT 变换后，除直流系数以外的 8×8 块 DCT 系数分布服从 Laplacian 分布。本文为了计算方便和考虑到 JPEG 压缩后的量化效

应，采用 Cauchy 分布的特殊形式：

$$P(x) = \frac{P-1}{2s} \left[\left| \frac{x}{s} \right| - 1 \right]^{-p} \tag{1}$$

采用分布拟合试验检验该模型能很好的反应量化 DCT 系数的统计分布，对此模型的参数的估计采用极大似然估计方法得到。

对其他嵌入方法，可以采用其他的分别模型。如将信息隐藏在小波域系数上的方法，我们可以采用马尔可夫模型等。

在我们第二部分的说明中，认为统计量 S (m) 中 S (0) 的求取依赖于原始载体，仅仅知道嵌入信息后的载体，如何对原始载体进行建模便成了我们目前需要解决的问题。假设 h (d), d= ..., -2, -1, 0, 1, 2, ..., 为原始图像量化 DCT 系数直方图，H (d) 为嵌入一定信息量后的图像量化 DCT 系数直方图，数据嵌入率定义为：

$$2\alpha = \frac{m}{\sum_{i \neq 0, i \neq 1} h(i)}$$

将分别对 JSteg 和 QIM 两种信息隐藏方法进行分析。

3.1 JSteg 信息隐藏模型分布估计

在已知具体的嵌入算法和嵌入了数据的多媒体的情况下，根据多媒体的统计模型，可以估计出模型参数和模型的具体分布。Jsteg 通过用待嵌入的信息替换量化 DCT 系数的最低位置嵌入信息，这种替换排除系数 ‘0’ 和 ‘1’。那么原始图像量化 DCT 系数的直方图和嵌入了一定信息后的量化 DCT 系数的直方图之间具有如下关系：

$$H(0) = h(0), H(1) = h(1) \tag{2}$$

$$H(2i) = h(2i) - \alpha [h(2i) - h(2i+1)] \tag{3}$$

$$H(2i+1) = h(2i+1) + \alpha [h(2i) - h(2i+1)] \tag{4}$$

式中，i=±1, ±2, ..., α= $\frac{m}{2\sum_{i \neq 0, i \neq 1} H(i)}$ ，为嵌入 m bits 信息后，需要改变的系数数。得：

$$H(2i) + H(2i+1) = h(2i) + h(2i+1)$$

这说明相邻系数值 (2i 和 2i+1) 的总数目在信息嵌入前后保持不变，这个特性对所有的系数值都成立。

令

$$H\alpha(i) = H(2i) + H(2i+1)$$

使用 H(0), H(1), Hα(i) 去拟合式 (1) 所示意的统计模型。事实上 Hα(i) 是量化 DCT 系数的粗略直方图。拟合的方法采用极大似然估计方法，那么模型参数 s 和 p 就可以估计出来 (图 2)。

令 ĥ(d) 为估计量化 DCT 系数的直方图，根据嵌入特性有：

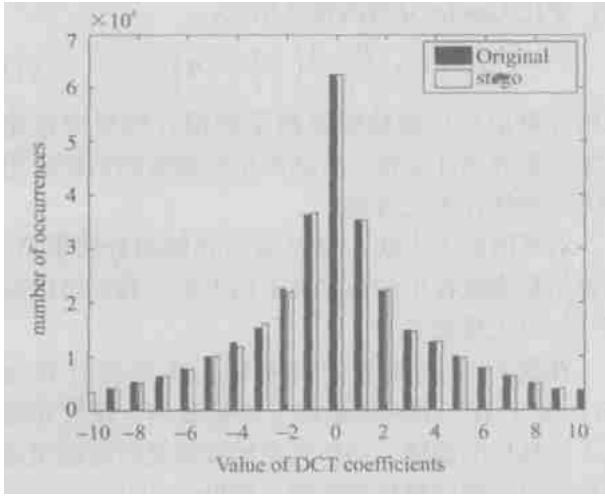


图 2 原始图像直方图和估计直方图

Fig. 2 A comparison of coefficient histograms between original histogram of cover image and estimated histogram of Steg image

$$\hat{h}(0) = H(0), \hat{h}(1) = H(1) \tag{5}$$

$$\hat{h}(2i) = H_a(i) \frac{P(2i)}{P_a(i)} \tag{6}$$

$$\hat{h}(2i+1) = H_a(i) \frac{P(2i+1)}{P_a(i)} \tag{7}$$

式中

$$P_a(i) = \int_{2i-0.5}^{2i+1.5} P(x) dx, P(2i) = \int_{2i-0.5}^{2i+0.5} P(x) dx,$$

$$P(2i+1) = \int_{2i+0.5}^{2i+1.5} P(x) dx$$

所以原图像量化 DCT 系数的直方图可以采用以上方法估计出来。图 2 显示原图像量化 DCT 系数的直方图和估计量化 DCT 系数的直方图。系数包括所有的 AC 系数。从图上看两者没有显著的差异。

3.2 QIM 信息隐藏模型分布估计

QIM 方法嵌入信息后，量化 DCT 系数转移具有如图 3 所示关系：图中黑竖条代表的是嵌入数据后的量化 DCT 系数。量化器 0，使得原 DCT 系数为 $x \in [A - \Delta/2, A + \Delta/2]$ 变成 A ，量化器 1，使得原 DCT 系数为 $x \in [A, A + \Delta]$ 变成 $A + \Delta/2$ 。当嵌入一定信息量后，那么原始图像量化 DCT 系数的直方图和嵌入了一定信息后的量化 DCT 系数的直方图之间具有如下关系：

$$H(x) = h(x)(1 - 2\alpha) \tag{8}$$

$$H(A + k\Delta/2) = h(A + k\Delta/2) +$$

$$\alpha \sum (h(x) + h(y)) + \alpha (h(A + (k - 1)\Delta/2) - h(A + (k + 1)\Delta/2)) \tag{9}$$

式中

$$A + k\Delta/2 \leq x < A + (k + 1)\Delta/2,$$

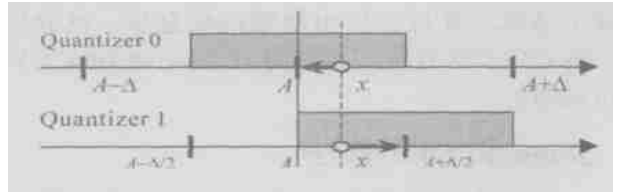


图 3 QIM 嵌入后系数变化

Fig. 3 Coefficients transition after QIM

$$A + (k - 1)\Delta/2 \leq y < A + k\Delta/2$$

从上式看，原始图像量化 DCT 系数的直方图和嵌入了一定信息后的量化 DCT 系数的直方图之间关系复杂，很难从嵌入了一定信息后的量化 DCT 系数的直方图估计出原始图像量化 DCT 系数的直方图。但从 (8) 式看，这两者直接具有一定的比例关系，如果仅仅使用 $H(x)$ 来估计模型 (1) 中的参数，这样我们得到一个“缩小的”直方图，这个直方图和原始直方图具有比例系数为 $1 - 2\alpha$ 的关系。从这个“缩小的”直方图中我们可以估计被嵌入信息的信息量。我们将在下一小节进行说明如何估计信息量，这里需要说明的是：在估计“缩小的”直方图的过程中，我们需要首先估计出一个参数量化步长，而这个可以通过对嵌入了一定信息后的量化 DCT 系数的直方图进行滤波而得到。

4 基于模型的信息隐藏检测和估计

4.1 对 Jsteg 的基于模型的信息隐藏检测和估计

在已知原始多媒体的统计模型参数或者模型具体分布的情况下，检测多媒体是否含有隐藏信息和估计嵌入信息量就变得相对容易了。实际上是寻找一个合适统计量，该统计量与模型有关，数据嵌入使该统计量发生变化，或者嵌入数据量与该统计量成一定的函数关系。对 Jsteg 嵌入法的信息隐藏检测和估计采用如下方法。

检测：在已知模型具体分布（估计出的量化 DCT 系数的直方图）的情况下，采用 χ^2 检验，统计量为：

$$\chi^2 = \sum_{i=-1}^{+k} \frac{(\hat{h}(i) - H(i))^2}{\hat{h}(i)}$$

在显著性水平 α 情况下，作自由度 $2k - 1$ 的 χ^2 检验。

嵌入信息量的估计：在已知模型具体分布（估计出的量化 DCT 系数的直方图）的情况下，从 (3)、(4) 式我们可以推导出以下方程式：

$$\alpha = \operatorname{argmin}_{\alpha} \sum_{i=-1}^{+k} (H(2i) - (1 - \alpha)\hat{h}(2i) - \alpha\hat{h}(2i+1))^2 \tag{10}$$

利用最小二乘法, (10) 式可写成:

$$\alpha = \frac{\sum_{i=\pm 1}^{\pm k} (H(2i) - \hat{h}(2i)) (\hat{h}(2i) - \hat{h}(2i+1))}{\sum_{i=\pm 1}^{\pm k} (\hat{h}(2i) - \hat{h}(2i+1))^2} \tag{11}$$

式中 k 是最大量化 DCT 系数。

4.2 对 QIM 的基于模型的信息隐藏检测和估计

从图 1 (c) 中可以看出嵌入信息后对量化 DCT 系数的影响, 所以对 QIM 的对检测将很是可行的。检测过程中首先需要估计量化步长, 这个量化步长实际上是对嵌入了一定信息后的量化 DCT 系数的直方图进行 FFT 滤波后的半周期^[9]。然后根据已经得到的半周期估计“缩小的”直方图, 从“缩小的”直方图中可以得到 $h'(0)$ 和 $h'(1)$ 根据嵌入过程我们知道, 由于排除了“0”和“1”系数, 这样 $H(0)$ 和 $H(1)$ 在嵌入前后将保持不变。这样利用 (8) 式可以推出嵌入信息量:

$$\alpha = \left\lfloor \frac{1-h'(m)}{H(m)} \right\rfloor, m = 0, 1 \tag{12}$$

5 实验结论

使用 CBIR 数据库 (共 824 JPEG 幅图像) 测试我们的方法^[10, 12]。实验中, 我们对图像库中每一幅图像都生成 8 幅图像 (4 幅采用 JSteg 嵌入方法, 4 幅采用 QIM), 每种嵌入方法嵌入的信息量分别为 20%, 50%, 80% 和 100% 的最大嵌入信息容量, 对应 $2\alpha = 0.2, 0.5, 0.8, 1.0$ 。实验结果如图 4 所示 (因篇幅限制, 本图中只显示部分估计的结果)。图 4 中, ‘△’, ‘*’, ‘o’, ‘+’ and ‘☆’估计的对应 $2\alpha = 1.0, 0.8, 0.5, 0.2, 0$ 的嵌入量。从图 4 看, 实验结果令人满意。

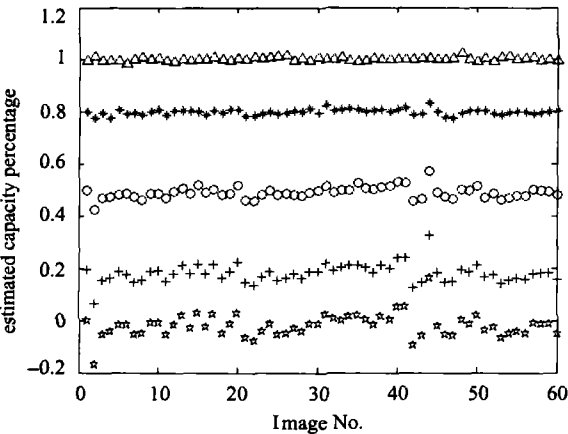


图 4 估计的嵌入信息量

Fig 4 Estimates of embedding ratio on part images from ICBR image database

本方法的主要贡献和特点在于: ①提出一种基于模型的信息隐藏检测方法, 使之能在已知未嵌入信息多媒体的统计模型的情况下, 有效检测出多媒体中是否有可疑信息嵌入; ②在已知多媒体统计模型和嵌入方法的情况下, 可以估计出嵌入信息的容量; ③本方法是一个框架性质的方法, 能应用于其他信息隐藏方法 (如小波变换后嵌入信息) 或者其他类型载体 (如 JPEG2000); ④应用本文所述的技术, 还可以证明其他经验性信息隐藏方法的正确性; ⑤该技术仍然属于有针对性的信息隐藏检测, 但它是一个框架性质的方法, 有一定意义的进步, 推进了信息隐藏检测技术的开发和应用。

参考文献:

[1] JOHNSON N F. Sushil jajodia. steganaly sis of images created using current steganography software[C] . LNCS Vol. 1525, Springer-Verlag, 1998: 273—289.

[2] FRIDRICH J, GOLJAN M. Practical steganalysis: state of the art[J] . In Proceeding of SPIE, 2002, 4675: 1—13.

[3] FARID H. Detecting steganographic message in digital images [J] . Report TR2001—412, Dartmouth College, Hanover, NH, 2001.

[4] MEMON N D, AVCIBAS I, SANKUR B. Steganalysis based on image quality metrics[J] . Proceeding of SPIE, 2001, 4314.

[5] FRIDRICH J, GOLJAN M, HOGEA D, et al. Quantitative steganalysis of digital images: estimating the secret message length[J] . ACM Multimedia Systems Journal, Special issue on Multimedia Security, 2003: 288—302.

[6] JPEG — JSteg — V4, <http://www.funet.fi/pub/crypt/steganography/jpeg-jsteg-v4.diff.gz>.

[7] CHEN B, WORNELL G. Quantization index modulation: a class of provably good method for watermarking and information embedding[J] . IEEE Transactions on Information Theory, 2001, 47(4): 1423—1443.

[8] WESTFELD A, PFITZMANN A. Attacks on steganographic systems[J] . IHW 99, Dresden, Germany, 1999.

[9] ZHANG K, WANG S, ZHANG X. Detection and removal of hidden data in images embedded with quantization index modulation[J] . MMM-ACNS, 2003: 360—370.

[10] CBIR Image Database, University of Washington, <http://www.cs.washington.edu/research/magedatabase/roundtruth/>.

[11] SMOOT S R, ROWE L A. Study of DCT coefficient distributions[C] . Proceedings of the SPIE Symposium on Electronic Imaging, 1996, 2657.

[12] SALLEE P. Model based steganography[J] . Proceedings of Intemational Workshop on Digital Watermarking, 2003.

(下转第 191 页)

Blind Audio Authentication Algorithm Based on SVD and Amalgamation

WANG Xiao, BAI Sen, ZHAO Bo

(Chongqing Communication Institute, Chongqing 400035, China)

Abstract: To verify the authenticity and integrity of audio, we present an audio authentication method, in which singular value decomposition (SVD) is used to extract singular value as invariable features of host audio. Combining with amalgamation, we insert these extracted features into the host audio at the sender. Then at the receiver, we extract inserted features as authentication code, and compare them with those features from SVD. After setting proper threshold, we can make our judgments at last. Experimental results show that the proposed method not only can distinguish Gaussian noise from malicious tampering but also can detect where tampered regions are.

Key words: audio authentication; singular value decomposition (SVD); amalgamation; watermarking

(上接第 187 页)

Model based Steganalysis

YU Xiao yi, TAN Tie niu, WANG Yun hong

(NLPR, Institute of Automation, CAS, Beijing 100080, China)

Abstract: To consider a new method for performing steganalysis using a statistical model of the cover medium. Using model based methodology, examples of detecting secret message and estimating the secret message length of bit streams embedded using JSteg-like steganography and quantization index modulation are proposed. This steganalysis technique is based on the model of statistical distribution of quantized DCT coefficients. The histogram of cover image and “shrinkage histogram” are estimated from stego image using the statistical model. Then the secret message is detected and the secret message length is estimated. The methodology described in this paper is a framework which can also be applied to virtually any type of media such as JPEG2000 file format embedding.

Key words: steganalysis; DCT; QIM