

数字水印算法检测标准的研究

彭 静

(重庆师范大学物理学与信息技术学院 重庆 沙坪坝区 400047)

【摘要】对数字水印检测技术中的相关系数进行研究分析,得到了水印检测器误码率与相关系数之间的关系式,推导出误码率与错误概率之间的关系式,能够直接通过控制关系式中的误码率取值来确定检测阈值和错误概率值。并据此提出一种新的计算检测系数的方法,对比实验结果表明,该方法增强了检测器对正确水印数据的响应。

关键词 相关系数; 检测阈值; 数字水印; 错误概率
中图分类号 TP391 文献标识码 A

A Brief Analysis on Detection Standard of Digital Watermark

PENG Jing

(College of Physics and Information Technology, Chongqing Normal University Shapingba Chongqing 400047)

Abstract The correlation coefficient is researched in digital watermark algorithm. The relation between the correlation coefficient and the bit error rates of watermark detection is found. The detection threshold and wrong probability can be calculated out directly by the relation. Furthermore, a method to calculate detection coefficient is proposed. This method can increase the threshold of detection while keeping the probability of the false detection not to change and can enable the detection to strengthen the responses of the correct watermark data.

Key words correlative coefficient; detective threshold; digital watermark false-detection probability

当今因特网络技术和多媒体信息处理技术发展迅速,为数字图像媒体的传输和获取提供了十分便利的条件,同时也暴露出越来越严重的安全问题。人们在通过网络发布软件、多媒体作品时,希望能够有效地保护自己的作品版权,防止他人对信息进行非法复制、传播以及篡改。而数字水印技术^[1-3]的出现给版权保护、防止非法复制、图像认证、电子商务中网页保护和票防伪,提供了比传统的密码技术更有力的保护,已得到了广泛的关注和发展。

在数字水印技术中,还有许多关键问题尚待解决,如良好的嵌入提取算法、水印图像可视性效果的评价标准、水印检测标准等。而水印检测通常采用相关检测的方法,水印检测系统不可避免要产生错误,常见错误包括虚警错误和漏警错误。虚警错误是指水印检测器检测一件没有嵌入水印的作品,结果报告发现水印;漏检错误是指水印检测器检测一件已经嵌入了水印的作品,结果报告没有发现水印。因而可将漏警概率和虚警概率作为衡量检测性能的指标^[4-5]。一个好的水印检测器应该具有高的可靠性,特别是在版权保护系统中将检测结果提供给

法庭作为证据时,要求虚警概率很低,因为如果水印检测器的虚警概率很高,会降低证据的说服力,所以对水印系统的虚警概率要求在 10^{-5} 以下。

采用相关检测器作检测时,关键问题是如何确定一个合理的检测阈值。检测阈值与人们所期望的检测器错误概率紧密相关。为确定错误率与检测阈值之间的数值关系,一个直接的方法是对大量的图像和水印做大量的实验,以求得不同检测阈值下的错误率,但实际上这种方法并不可行。本文在一种鲁棒水印算法的基础上,研究了检测系数的性能,计算出了相关系数与误码率之间的关系,提出一种新的检测标准。

1 检测系数

1.1 相关系数

由水印检测器将提取出的数据同指定的水印数据进行相关系数计算,再根据计算值与某个预先设定的阈值进行比较。相关系数值大于阈值时,可以判定待测图像中加入了水印数据,受版权保护;若相关系数值小于阈值时,则可说明待测图像中未含

有水印信息。

根据不同的水印数据形式及各种水印嵌入和提取算法,参与计算相关系数的数据对象也不尽相同。一般水印数据主要有两种形式:(1)由某一范围的实数或整数构成,如伪随机序列产生的水印、二维灰度图像水印;(2)将水印预处理成为二值序列数据串,每位数据取1或-1。在变换域水印算法中,如果提取得到的仅是修改过后的系数值 C' ,其中包含水印含量,则将该系数值与指定水印的数据 w 进行相关系数计算 $N_{NC}(w, C')$ 。如果直接提取出了水印的数据值 w' ,则可以将之与 w 进行自相关系数计算 $N_{NC}(w, w')$ 。本文讨论时采用后者的形式,可使检测更准确,并且其他形式都可以转换成二值序列形式。

1.2 相关系数与误码率的关系

设水印数据 w 包含 N_w 个二值序列数据位, $w(i)$ 在 $\{-1,1\}$ 中取值, $i=1,2,\dots,N_w$ 。在从待测图像中提取出的数据 w' 中假设有 n 个数据值与 w 中对应位上的值不相同,即是出错的数据位,则定义误码率 ρ 来表示两组数据间的差异性,其表达式为:

$$\rho = \frac{n}{N_w} \quad (1)$$

因此, w' 中有 $m=N_w-n$ 个数据与 w 中相应数据位的值一致,即是提取正确的数据位。相关系数:

$$N_{NC} = \frac{\sum w(i)w'(i)}{\sqrt{\sum w^2(i)}\sqrt{\sum w'^2(i)}} \quad (2)$$

又由 $\sqrt{\sum w^2(i)} = \sqrt{\sum w'^2(i)} = \sqrt{N_w}$ 和 $\sum w(i)w'(i) = m - n = (N_w - n) - n = N_w - 2n$,所以相关系数:

$$N_{NC} = \frac{N_w - 2n}{N_w} = 1 - 2\rho \quad (3)$$

式(3)表明了相关系数 N_{NC} 与误码率 ρ 之间的关系,同时也可知如果 ρ 越小, N_{NC} 越大,则在检测时可设置较高的阈值 T ,就可降低虚警概率。在通常的水印检测中,特别要注意降低虚警概率。

1.3 错误概率与误码率的关系

由式(1)、(3)知 ρ 的取值范围为 $[0,1]$ 之间的实数,而 N_{NC} 就取 $[-1,1]$ 之间的实数。阈值 T 一定是大于0的正数。因此对数字水印检测进行判定也可采用其他的检测标准,本文定义检测系数:

$$E = 1 - k\rho \quad (4)$$

式中 ρ 为误码率; k 取正实数。下面来讨论其虚警概率 $P=P\{E < T | \text{无水印}\}$ 。为便于计算,可将式(4)表示为:

$$E = \frac{N_w - k\rho N_w}{N_w} = \frac{N_w - kn}{N_w} \quad (5)$$

则虚警概率为 $P=P\{N_w - kn < N_w T | \text{无水印}\}$ 。即当 n 取小于 $N_w(1-T)/k$ 时,将出现虚警,此时有:

$$P = \sum_{n=0}^{N_w(1-T)/k} P\{N_w - kn | \text{无水印}\} = \sum_{n=0}^{N_w(1-T)/k} \binom{N_w}{n} \times p^n (1-p)^{N_w-n} = \sum_{n=0}^{N_w(1-T)/k} \binom{N_w}{n} 0.5^{N_w} \quad (6)$$

式中 p 为某数位提取时出错的概率,因其数据值取1或-1,所以 $p=0.5$ 。文献[2]中计算出的虚警概率为:

$$P = \sum_{m=N_w(1+T)/2}^{N_w} \binom{N_w}{m} 0.5^{N_w} \quad (7)$$

式中由于 m 是提取正确的数据个数,有 $m=N_w - n$,又有:

$$\binom{N_w}{m} = \binom{N_w}{N_w - n} = \binom{N_w}{n} \quad (8)$$

因此文献[2]中的虚警概率为:

$$P = \sum_{n=0}^{N_w(1-T)/2} \binom{N_w}{n} 0.5^{N_w} \quad (9)$$

即式(6)中当 $k=2$ 时,就可得到文献[2]中的虚警概率表达式。在进一步判定时,若水印检测器的最大误码率为 ρ_m ,则阈值 T 至少应取 $T = 1 - k\rho_m$,代入式(5)可得到:

$$P = \sum_{n=0}^{\rho_m N_w} \binom{N_w}{n} 0.5^{N_w} \quad (10)$$

从式(10)中可知,虚警概率 P 与误码率 ρ 有关,若提取水印时误码率 ρ 越小,则 P 越小,对水印检测准确性越高。当 ρ 固定时,则 P 固定,但 T 的取值会随 k 的增大而变小。但是检测时往往希望阈值较大,可以借助改变 k 的值来选择不同的检测系数。在式(4)中若 $k=1$,则 $E=1-\rho$,刚好是检测器的解码的正确率,因此也可用它作为检测系数。

1.4 改进检测系数计算方法

为了使检测时更能突出正确的水印检测值,抑制错误的检测值,用如下改进方法计算检测系数:先依次按式(4)计算每个待测数据 E_i 值;再求出所有 E_i 中的最小值;然后将每个 E_i 值代入式中进行计算:

$$S_i = lE_i \frac{E_i}{(1 + |E_{\min}|)} \quad (11)$$

按此方法计算时,一般用未嵌入水印的图像提出的数据,计算出的 E_i 值很小,则其 $E_i/(1+|E_{\min}|)$ 值也很小,因此 S 会更小;反之若待测数据是从嵌入了水印的图像中提取出的,则 E_i 的值均较大,因此得到的 S 值会远大于其他的值。乘上一个系数 l 是为了将结果放大使于观察,根据所选的阈值进行取值,再与

给出的阈值进行比较,得出的检测结果会更准确。

2 实验结果

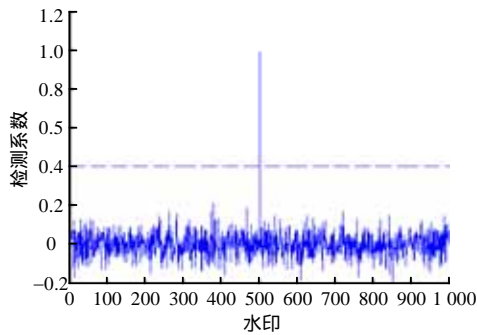


图1 文献[2]的检测图

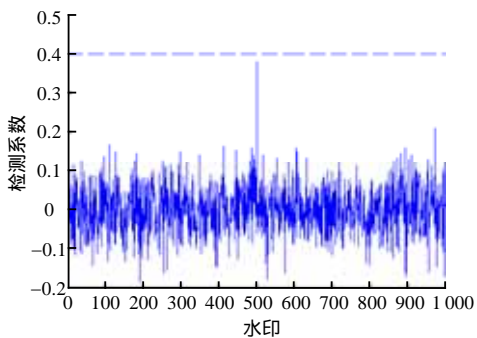


图2 文献[2]压缩后检测测图

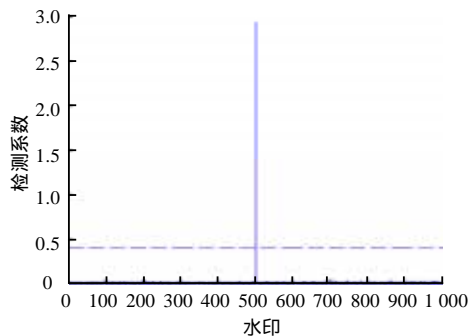


图3 改进方法的检测图

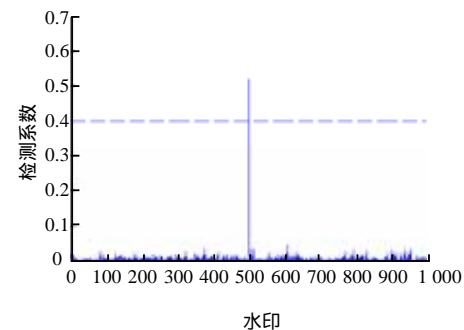


图4 压缩后改进方法检测图

对 512×512 的标准LENA图像,用文献[2]的鲁棒水印算法嵌入二值序列的水印,其长度为 32×32 。再提取水印进行检测,设水印检测器的最大误码率 ρ_m 为0.3,由文献[2]的相关检测系数 M_{NC} 和式(3),其检测阈值为0.4,检测结果如图1所示。但在受到攻击时,文献[2]的方法已不能检测到水印的存在,如JPEG压缩率在10时,水印相关系数小于阈值,如图2所示,而此时水印图像仍具有较好的可视效果。

用本文提出的改进式(11)对文献[2]嵌入的水印进行检测,并与上述结果相比较。在计算 E 的式(4)中 k 取2,式(11)中 r 取4,仍然用前面的检测阈值0.4,其水印检测器的响应如图3所示。在对水印图像进行JPEG压缩时,可见到检测器对水印数据的响应仍然非常高,明显地显示水印的存在,如图4所示。

3 结束语

在水印检测时由于检测阈值和错误概率互相影响,确定一个合理的阈值往往比较复杂。本文通过在水印检测时对相关系数的研究,得到了水印检测器的误码率与相关系数的关系,同时也得到了虚警概率与误码率的关系,进而可对水印检测器的判断阈值与虚警概率的取值由误码率来确定,即控制误码率的最大取值就可以方便地确定阈值和虚警概率的值。本文还对检测系数的计算方法做了改进,提高水印检测器对受攻击的水印图像的响应。

参 考 文 献

- [1] COX J. Secure spread spectrum watermarking for multimedia[J]. IEEE Trans-Image Processing, 1997, 6(2): 1673-1687.
- [2] KUNDUR D, HATZINAKOS D. Digital watermark using multiresolution wavelet decomposition[C]//IEEE International Conference on Acoustics, Speech and Signal Processing. Seattle, Washington: [s. n.], 1998: 2969-2972.
- [3] MEI Shi-chun, LI Ren-hou, FANG Hai-jian. An adaptive image watermarking algorithm based on neural-networks[J]. Journal of China Institute of Communications. 2002, 123(12): 47-53.
- [4] PIVA A. Threshold selection for correlation based watermark detection[C]//Proc COST254 Workshop on Intelligent Communications. L' Aquila, Italy: [s. n.], 1998: 5-6.
- [5] 刘 彤, 裴正定. 数字水印相关检测的可靠性研究[J]. 电子学报, 30(5): 2002, 685-688.

编 辑 孙晓丹