第 59 卷 第 8 期 2010 年 8 月
1000-3290/2010/59(08)/5870-10

物 理 学 报
ACTA PHYSICA SINICA

Vol. 59, No. 8, August, 2010

数字水印的鲁棒性评测模型*

曾高荣† 裘正定

(北京交通大学信息科学研究所, 北京 100044)

(2009 年 10 月 31 日收到;2009 年 11 月 23 日收到修改稿)

数字水印的鲁棒性是水印技术实用化的一个重要指标. 与通过 StirMark 测试和各种仿真测试不同,定义互信息作为代价函数,建立水印系统鲁棒性描述和度量的一般模型. 以扩频水印和量化水印为范例,推导出评测鲁棒性的互信息度量计算模型,并仿真分析了高斯噪声和 JPEG 压缩条件下互信息函数对鲁棒性的评估结果. 实验以统计误比特率的方法计算图像 DCT 域中低频系数为载体的扩频水印误码率和一段音频数据三级小波细节系数为载体的量化水印误码率. 互信息函数和误码率之间的匹配关系验证了互信息度量模型的有效性,互信息函数可以作为代价函数预测误码率的变化趋势.

关键词: 数字水印, 鲁棒性, 互信息, 误码率
PACC: 8710, 0250

1. 引 言

数字水印是不可感知地在载体中嵌入信息的操作行为[1]. 数字水印以不可感知的方式在数字内容中永久性隐藏水印信息的技术特性弥补了加密技术的不足,使得数字水印在版权保护、拷贝控制、数字指纹、交易跟踪、内容标识、隐秘通信、视频广播监控等领域成为一种重要的解决方案[1].

与加密技术不同,数字水印技术并不严格要求内容无失真传输,只要人们在视觉或听觉上感知无变化,允许内容使用时和原始内容存在一定程度的失真. 这就要求隐藏在数字内容中的水印信息在传输和处理后能幸存,这种幸存能力称为数字水印的鲁棒性. 要设计出符合鲁棒性要求的水印算法必须首先确定鲁棒性的描述方式和评价标准. 已有的鲁棒性评测基准一般采用国际标准的测试软件来测量不同水印算法的鲁棒性. 用得较多的是 StirMark 软件序列. StirMark 是 Petitcolas 等开发的首个用于数字图像水印系统鲁棒性测试的通用工具[2], Checkmark[3] 和 Optimark[4] 是它的扩展和改进版本. 该类测试工具指定嵌入信息的有效载荷容量(一般 80 bit)、必须保证的保真度(由专门的感知模型来衡量). 保持这两个特性不变,然后用 StirMark 程序来测试系统对抗各种失真的鲁棒性,最后将测试出的系统性能综合成一个分数作为鲁棒性度量值. 该基准工具虽可作为一般水印系统的公共测试平台,衡量算法平均鲁棒性. 但是,由于水印系统是与应用密切相关的,某些测试条件并不是应用背景所严格要求的,如,1)水印系统设计的载荷容量未必要求很高,无需达到 80 bit. 2)在一些应用中存在的攻击,在另一些应用中可能不存在. 如拷贝控制中的视频水印可能要考虑小角度的旋转攻击,而在广播监视中却未必需要考虑. 况且它只是个实验测试平台,所提供的测试也不能代表安全性所要求的普遍测试. 通过这种公共平台测试出来的分数值用于评价不同应用背景下水印系统的鲁棒性显然意义不大.

除了上述 StirMark 系列测试基准外,研究人员也针对具体算法采用具体的鲁棒性度量方式,如简单加性水印,采用线性相关检测水印,用检测统计量本身作为算法鲁棒性度量尺度[1],固定鲁棒性将相关系数降至阈值以下时的噪声幅度[5],量化索引调制水印中的量化器之间最小距离度量算法抗噪声干扰的鲁棒性能[6,7]. 更多的研究者是通过选择一些仿真攻击条件对设计的水印算法进行鲁棒性

* 国家科技支撑计划(批准号:2008BAH33B01),国家自然科学基金(批准号:60773015),国家高技术研究发展计划(批准号:2007AA01Z460)资助的课题.

† E-mail: loch. zeng@ gmail. com

仿真测试，如文献[8—10]. 这些一种算法一种度量的描述方式，虽然可以粗略评估算法的鲁棒性，但适用面过于狭窄，不便于扩展到其它算法，而且也不利于理论分析.

与上述鲁棒性评测工具不同，本文试图建立水印系统鲁棒性描述和度量的一般模型，可用于刻画任意的水印系统. 后续部分首先提出水印系统鲁棒性的描述和度量模型，以两类经典的算法：扩频水印和量化水印为范例，对水印鲁棒性进行详细分析，并通过统计误比特率的方法测量实验误码率，验证互信息度量模型的有效性，最后给出本文的结论.

2. 水印系统鲁棒性的描述和度量模型

水印过程类似数据通信过程，水印检测类似通信中从接收信号中提取出传输的信号. 根据通信理论，要从接收内容中提取水印信息，则接收内容必须包含足够的水印信息. 本文中接收内容用随机量 Y 表示，水印用随机量 M 或 m 表示. 水印的信息量由水印嵌入前检测器对水印的不确定性决定，可用水印的信息熵表示：$H(M)$. 检测时，接收内容对检测器是已知的，此时检测器对水印的不确定性为条件熵 $H(M \mid Y)$，从接收内容获得的关于水印的信息量为 $H(M) - H(M \mid Y)$. 根据互信息函数的定义

$$I(M;Y) = H(M) - H(M \mid Y), \tag{1}$$

即检测器从接收内容中获得的关于水印的信息量为接收内容和水印之间的平均互信息. 当数字内容在分发阶段没有经受任何处理，根据嵌入和检测之间一一对应的关系，M 和 Y 具有确定关系，$H(M \mid Y) = 0$，$I(M;Y) = H(M) - 0 = H(M)$，检测器获取了水印 M 的全部信息量，可以无差错提取水印. 当数字内容在分发阶段受到某种处理或攻击，由此产生了随机污染，对检测来说是种噪声干扰，使得检测器在已知接收内容时，关于水印 M 出现了不确定性，$H(M \mid Y) > 0$，导致互信息量下降，即

$$I(M;Y) = H(M) - H(M \mid Y) < H(M), \tag{2}$$

水印检测时将出现错误. 错误率有多大，或者说检测器还能多大程度上提取水印，由信号处理或攻击信道决定. 当攻击强度足够大，极限状态 $H(M \mid Y) = H(M)$，$I(M;Y) = 0$，即 Y 不包含 M 的任何信息，表示水印已经从内容载体中完全被删除了，此时理论上检测不到水印.

由以上分析，我们提出用表征互信息量的互信息函数即接收内容 Y 和水印 M 之间的互信息函数 $I(M;Y)$ 来描述水印系统的鲁棒性，其度量模型为

1) $I(M;Y)$ 越大，系统鲁棒性越强，检测误码率越小. $H(M \mid Y) = 0$ 时，$I(M;Y)$ 取最大值，系统鲁棒性最强，可以实现无差错检测.

2) $I(M;Y)$ 越小，系统鲁棒性越弱，检测误码率越大. $H(M \mid Y) = H(M)$ 时，$I(M;Y)$ 取最小值 0，系统已经没有鲁棒性，此时无法检测到水印.

对该模型从嵌入空间和攻击策略进行拓展，可以得到两种极限状态的性能限.

固定攻击策略，即指定信号处理条件，满足感知失真限定的嵌入空间内，选择某种嵌入算法使得 $I(M;Y)$ 最大，此最大值可称为该种信号处理/攻击下的水印容量

$$C = \max_{\text{嵌入空间}} \{I(M;Y)\}. \tag{3}$$

容量也反映该攻击条件下允许的水印传输率.

对确定的嵌入算法，在满足感知失真限制的攻击信道集合中，可以找到某个信道，使得水印检测达到应用标准的最小 $I(M;Y)$，此最小值为信息率失真函数 R 为

$$R = \min_{\text{攻击策略}} \{I(M;Y)\}. \tag{4}$$

率失真函数反映该嵌入算法所能承受的最大攻击条件.

由于水印技术和应用密切相关，不同的应用其要求相差很大，算法实现也相应不同，没有一个确切的表达式来涵盖所有的水印算法. 但是，具有里程碑意义上的水印算法主要包括两类：扩频水印和以 QIM 水印为代表的量化水印. 下面以这两类算法为范例采用互信息度量模型来分析它们的鲁棒性.

3. 水印系统的鲁棒性分析

3.1. 扩频水印鲁棒性分析

扩频水印是 Cox[11] 发明的一种鲁棒多媒体水印技术. 其基本思路是借鉴扩频通信以高传输带宽换取低的传输信噪比的思想，将 1 bit 水印信息隐藏在多个载体系数中，达到扩频和降低传输信噪比的目的.

扩频水印的嵌入函数可以简单表示为

$$\boldsymbol{X} = \boldsymbol{S} + \boldsymbol{W}, \tag{5}$$

其中 $\boldsymbol{S}$ 是由载体系数 $\{s_i, i = 1, \cdots, N\}$ 组成的向量，

$\boldsymbol{X}$ 是嵌入水印后载体系数 $\{x_i, i = 1,\cdots,N\}$ 组成的向量,其中 N 表示扩频向量长度. $\boldsymbol{W}$ 表示水印向量. 出于安全考虑,$\boldsymbol{W}$ 是由密钥生成并受水印 m 调制的伪随机向量. 对于二元水印,有 $\boldsymbol{W} = (-1)^m\boldsymbol{K}$, $\boldsymbol{K}$ 为密钥生成的伪随机系列 $\{k_i, i = 1,\cdots,N\}$ [12]. 这里不妨设 k_i 独立同分布,均值为 0.

经过噪声攻击或一般信号处理后的数字内容,本文称之为接收内容,接收内容向量用 $\boldsymbol{Y}$ 表示,其各维特征系数用 $\{y_i, i = 1,\cdots,N\}$ 表示,形式上可以得到

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{A} = \boldsymbol{S} + (-1)^m\boldsymbol{K} + \boldsymbol{A}, \tag{6}$$

其中向量 $\boldsymbol{A}$ 表示各种信号处理或攻击噪声,其各维元素为 $\{a_i, i = 1,\cdots,N\}$, 本文统称为噪声.

为了保证数字内容的可用性,嵌入水印后和遭受攻击后失真应满足一定限制. 采用数学度量 $D_w = \frac{1}{N}E[\|W\|^2]$ 评估每维载体系数的平均嵌入失真,$\|W\|^2 = \sum_i w_i^2$. 内容水印比 DWR 表示载体和水印之间的相对大小,度量嵌入水印引起的相对失真

$$\mathrm{DWR} = 10\log_{10}\frac{\frac{1}{N}E[\|S\|^2]}{D_w}. \tag{7}$$

对攻击噪声的失真限定可用水印噪声比 WNR 表示,定义为

$$\mathrm{WNR} = 10\log_{10}\frac{D_w}{\frac{1}{N}E[\|\mathrm{A}\|^2]}. \tag{8}$$

理论分析中涉及的信号被建模为随机变量. 不失一般性,可假定 $\{s_i, i = 1,\cdots,N\}$ 和 $\{a_i, i = 1,\cdots,N\}$ 为独立同分布的随机量,均值为 0,方差分别为 σ_s^2 和 σ_n^2 ;$m \in \{0,1,\cdots,p-1\}$ 并且均匀分布的. 讨论二元符号,即 $m \in \{0,1\}$ 等概率取值 0 或 1.

根据伪随机数发生器的原理,同一密钥生成的伪随机系列可以完全一致,对检测器而言密钥是已知的,即 $\boldsymbol{K}$ 就是已知的. 盲检测时原始载体对检测器不可知,扩频水印的鲁棒性度量函数 I_{R} 为

$$I_{\mathrm{R}} = I(m;Y|\mathrm{key}) = I(m;Y|k), \tag{9}$$

其中 k 为给定密钥下伪随机向量的一个实现.

根据检测与估计理论[13]

$$\begin{aligned} I(m;Y|k) &= I(m;Y^{\mathrm{T}}k|k) \\ &= h(Y^{\mathrm{T}}k|k) - h(Y^{\mathrm{T}}k|m,k). \end{aligned} \tag{10}$$

对于独立同分布的载体系数 $\boldsymbol{S} = \{\mathrm{s_i}, \mathrm{i} = 1,\cdots,\mathrm{N}\}$、噪声系数 $\boldsymbol{A} = \{a_i, i = 1,\cdots,N\}$, 和等概率分布的水印信息位 $m \in \{0,1\}$,有

$$\begin{aligned} Y^{\mathrm{T}}k &= (S + (-1)^m k + \mathrm{A})^{\mathrm{T}}k \\ &= S^{\mathrm{T}}k + (-1)^m\|k\|^2 + A^{\mathrm{T}}k \\ &= \sum_{i=1}^{N} s_i k_i + \sum_{i=1}^{N} a_i k_i + (-1)^m\|k\|^2. \end{aligned} \tag{11}$$

根据中心极限定理和李雅普诺夫定理,当 N 足够大时,$\frac{1}{N}\sum_{i=1}^{N} s_i k_i$ 服从高斯分布,其均值为 0,方差为 $\frac{1}{N^2}\sigma_s^2\|k\|^2$, $\frac{1}{N}\sum_{i=1}^{N} a_i k_i$ 服从均值为 0,方差为 $\frac{1}{N^2}\sigma_n^2\|k\|^2$ 的高斯分布. 所以有条件概率分布

$$Y^{\mathrm{T}}k|m,k \propto N((-1)^m\|k\|^2, (\sigma_s^2 + \sigma_n^2)\|k\|^2), \tag{12}$$

$$\begin{aligned} Y^{\mathrm{T}}k|k \propto &\frac{1}{2}N(\|k\|^2, (\sigma_s^2 + \sigma_n^2)\|k\|^2) \\ &+ \frac{1}{2}N(-\|k\|^2, (\sigma_s^2 + \sigma_n^2)\|k\|^2). \end{aligned} \tag{13}$$

根据信息论基础[14],

$$h(Y^{\mathrm{T}}k|m,k) = \frac{1}{2}\log(2\pi e(\sigma_s^2 + \sigma_n^2)\|k\|^2). \tag{14}$$

为书写简便,记 $T = Y^{\mathrm{T}}k|k$, 则

$$h(Y^{\mathrm{T}}k|k) = h(T) = -\int_R f(t)\log f(t)\,\mathrm{d}t, \tag{15}$$

$$\begin{aligned} I(m;Y|k) = &-\int_R f(t)\log f(t)\,\mathrm{d}t \\ &- \log(2\pi e(\sigma_s^2 + \sigma_n^2)\|k\|^2). \end{aligned} \tag{16}$$

其中 $f(t)$ 为变量 T 的概率密度函数,属于混合高斯分布.

对于非盲检测,即原始载体信息对检测器是已知的. 此时检测器从接收内容中获得的关于水印的信息量为 $I_{\mathrm{R}} = I(m;Y|s,\mathrm{key}) = I(m;Y|s,k)$. 令 $Y_A = Y - S = (-1)^m k + A$,根据信息论基础[14],$I(m;Y|s,k) = I(m;Y_A|s,k) = I(m;Y_A|k)$. 与盲检测类似,有

$$\begin{aligned} I(m;Y_A|k) &= I(m;Y_A^{\mathrm{T}}k|k) \\ &= h(Y_A^{\mathrm{T}}k|k) - h(Y_A^{\mathrm{T}}k|m,k), \end{aligned} \tag{17}$$

其中

$$Y_A^{\mathrm{T}}k|m,k \propto N((-1)^m\|k\|^2, \sigma_n^2\|k\|^2),$$

$$\begin{aligned} Y_A^{\mathrm{T}}k|k \propto &\frac{1}{2}N(\|k\|^2, \sigma_n^2\|k\|^2) \\ &+ \frac{1}{2}N(-\|k\|^2, \sigma_n^2\|k\|^2). \end{aligned}$$

为简便令 $T_1 = Y_A^{\mathrm{T}}k|k$,则非盲检测时的互信

息量

$$I_R = I(m;Y \mid s,k) = I(m;Y_\Lambda \mid k)$$
$$= -\int_R f(t_1)\log f(t_1)\,\mathrm{d}t - \log(2\pi e\sigma_n^2 \parallel k \parallel^2). \tag{18}$$

其中 $f(t_1)$ 为变量 T_1 的概率密度函数，属于混合高斯分布. 其均值与变量 T 相同，但方差更小.

由于混合高斯分布熵难以求得闭合解，但可采用数值积分的方法求解. 如载体系数方差 $\sigma_s^2 = 300$，内容水印比 DWR = 20 dB 时，随意选取几个 N 值，从 −25—20 dB 变化水印噪声比 WNR，可以计算出对应的鲁棒性度量值，如图 1 所示.

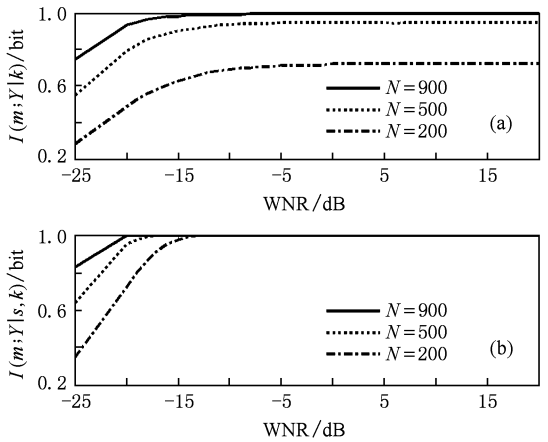

图 1　扩频水印鲁棒性分析不同扩频向量长度下互信息与水印噪声比 WNR 之间关系　(a)盲检测；(b)非盲检测

由(13)，(16)式和图 1 的仿真计算结果可知，互信息函数 $I_R = I(m;Y \mid k)$ 与载体系数方差，嵌入失真，攻击失真，扩频向量长度有关. 对于给定的载体和水印算法，载体系数方差确定，评测鲁棒性的互信息函数由内容水印比 DWR，水印噪声比 WNR，和扩频向量长度 N 决定. 由于内容水印比主要刻画与嵌入操作相关的透明性指标，而鲁棒性主要考虑适当的透明性要求下，水印和噪声的相对大小对鲁棒性的影响. 图 1 所示同样的水印噪声比 WNR 下，扩频向量长度越大，互信息函数值越大；同样的扩频向量长度，水印噪声比 WNR 越大，互信息函数值越大. 另外从图 1(a)和(b)比较中可见，同等条件下，非盲检测的互信息函数值大于盲检测互信息函数值. 根据本文互信息度量模型，更大的互信息，将体现更强的鲁棒性，即有更低的检测误码率. 后面第 4 节通过实验统计检测误码率来验证这个预测结论.

3.2. 量化水印鲁棒性分析

量化水印是根据要嵌入的水印信息，选用特定结构的量化器来量化载体系数的水印算法. Chen 和 Wornell 提出了一类用量化技术嵌入水印信息的量化索引调制(quantization Index Modulation, QIM)水印算法[7]，该类方法具有实现简单，消除载体干扰，以及低噪声环境下完全抗噪声的特点，并且可以做到无原始载体参与的完全盲提取.

一个简单的 QIM 水印方案可以由一个步长为 Δ 的标量均匀量化器 $Q(s) = \Delta[s/\Delta]$，符号 $[\cdot]$ 表示下取整，产生两个抖动量化器 $Q_m(\cdot)$ 来实现水印信息的嵌入.

$$Q_m(s) = Q(s - d_m) + d_m, \quad m = 0,1, \tag{19}$$

其中 $d_0 = -\frac{\Delta}{4}, d_1 = \frac{\Delta}{4}$，$m$ 表示二元水印信息位，s 表示载体系数.

水印嵌入函数为(x 表示嵌入水印后的信号)

$$x = \begin{cases} Q_0(s), & m = 0, \\ Q_1(s), & m = 1. \end{cases} \tag{20}$$

当水印信号 x 在被检测前受噪声 n 污染时，接收端收到的信号为

$$y = x + n. \tag{21}$$

水印译码函数为

$$\hat{m} = \underset{m \in \{0,1\}}{\operatorname{argmin}} \operatorname{dist}(y, \Lambda_m), \tag{22}$$

其中 $\operatorname{dist}(y,\Lambda_m) = \min_{s \in \Lambda_m} \mid y - s \mid$.

从译码规则可看出，检测时无需原始载体和水印信息参与，并且当 $\mid n \mid < \Delta/4$，水印信息可以无差错被检测，即 $\hat{m} = m$. 但当 $\mid n \mid > \Delta/4$ 时，水印检测可能出现错误. 由于 $Y = x + n = \Delta\mathscr{Z} + d_m + n$，令 $T = Y - \Delta\mathscr{Z} = d_m + n$，检测时作为分析变量，在高分辨率的假设前提下，载体分布可以看成是平滑的[15]，这样的模减操作对译码不会减少信息[16]，即有 $I(M;Y) = I(M;T)$.

当 $m = 0$ 时，T 可以看成一个冲激函数 $\delta\left(t + \frac{\Delta}{4}\right)$ 和噪声的叠加，其概率密度函数可以看成一个冲激函数和噪声概率密度函数的卷积，即

$$f_T(t \mid m = 0) = \delta\left(t + \frac{\Delta}{4}\right) * f_n(t) = f_n\left(t + \frac{\Delta}{4}\right). \tag{23}$$

类似地，当 $m = 1$ 时，有

$$f_T(t \mid m=1)=\delta\left(t-\frac{\Delta}{4}\right)*f_n(t)=f_n\left(t-\frac{\Delta}{4}\right). \tag{24}$$

由(23),(24)式可求得

$$f(t)=p(m=0)f_T(t \mid m=0)+p(m=1)f_T(t \mid m=1). \tag{25}$$

由信息论基础[14],对 m 等概率取值 0 或 1,可求得

$$I(M;T)=D(f_{MT}(m,t) \,\|\, f_T(t)p_M(m))=\frac{1}{2}\sum_{m=0}^{1}D(f_{T\mid M}(t \mid M=m) \,\|\, f_T(t)), \tag{26}$$

其中符号 D 表示相对熵.

当 n 为高斯噪声时,不妨设其均值为 0,方差为 σ_n^2,即 $n \propto N(0,\sigma_n^2)$. 此时,由(23),(24)式求得 $f_T(t \mid m=0) \propto N\left(-\frac{\Delta}{4},\sigma_n^2\right)$,$f_T(t \mid m=1) \propto N\left(\frac{\Delta}{4},\sigma_n^2\right)$,即它们分别服从取均值 $-\frac{\Delta}{4}$,$\frac{\Delta}{4}$,方差为 σ_n^2 的高斯分布. 由(25)式可求得 T 的概率密度函数

$$f(t)=p(m=0)f_T(t \mid m=0)+p(m=1)f_T(t \mid m=1)=\frac{1}{2}N\left(-\frac{\Delta}{4},\sigma_n^2\right)+\frac{1}{2}N\left(\frac{\Delta}{4},\sigma_n^2\right). \tag{27}$$

由 $f(t)$ 对称性和 m 值的均匀性[14,17],$D(f_{T\mid M}(t \mid M=1) \,\|\, f_T(t)) = D(f_{T\mid M}(t \mid M=0) \,\|\, f_T(t))$,所以有

$$I(M;T)=D(f_{T\mid M}(t \mid M=0) \,\|\, f_T(t))=\int f_{T\mid M}(t \mid M=0)\log_2\left(\frac{f_{T\mid M}(t \mid M=0)}{f_T(t)}\right)\mathrm{d}t=-\int \frac{1}{\sqrt{2\pi}\sigma_n}\mathrm{e}^{\frac{-\left(t+\frac{\Delta}{4}\right)^2}{2\sigma_n^2}}\log_2[(1+\mathrm{e}^{\frac{t\Delta}{2\sigma_n^2}})/2]\mathrm{d}t. \tag{28}$$

由(28)式,当信道条件为高斯噪声时,只要知道量化步长和噪声方差,则可以计算该信道条件下水印和接收内容之间的互信息量,据此评测水印通过该信道的鲁棒性.

4. 互信息度量模型的有效性验证

为了验证鲁棒性的互信息度量模型的有效性,本文在高斯噪声的背景下,在扩频水印中以图像为载体嵌入水印,采用相关检测统计译码时的误码率,在量化水印中以一段音频(某人的说话录音)为原始载体,在最小距离译码规则下统计误码率,以此来检验互信息度量方法. 在考虑其他攻击策略时,需要对具体的攻击建模,然后计算相应操作下接收内容和水印之间的互信息函数来评测鲁棒性. 后文以 JPEG 压缩为代表,基于 JPEG 压缩的一个近似模型,分析相应算法的鲁棒性.

4.1. 相关检测下扩频水印的误码率

图像压缩与编码领域的研究表明,图像 DCT 变换之后其交流系数服从广义高斯分布,概率密度函数具有如下形式[18,19]:

$$p_x(x)=A\mathrm{e}^{-|\beta x|^c}, \tag{29}$$

其中 $\beta=\frac{1}{\sigma}\left(\frac{\Gamma(3/c)}{\Gamma(1/c)}\right)^{1/2}$,$A=\frac{\beta c}{2\Gamma(1/c)}$,$\sigma$ 为系数的标准差,c 为形状参数,$\Gamma(\cdot)$ 为 Gamma 函数. 文献[18]采用 Buccigrossi 等人[20]提出的最小化相对熵的估计方法对参数 σ 和 c 进行估计,以标准图像库中的灰度图像 Fishing boat 和 Camera man 进行 8×8 分块 DCT 变换为例,采用最小化相对熵方法对 DCT 交流系数进行广义高斯概率密度函数估计,得到的结果如图 2 所示.

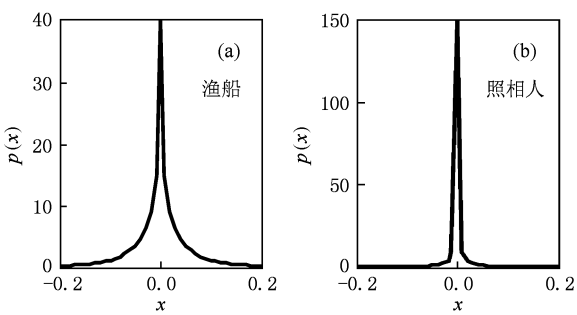

图 2 Fish boat 和 Camera man 的广义高斯分布拟合曲线

从图示结果看,与空域图像像素直方图分布相比,DCT 域的 AC 系数和高斯分布更接近,由于真正高斯分布的图像载体很难找到,实验中以标准图像库中 woman 的 DCT 中低频交流系数为载体,进行扩频水印嵌入和检测,并统计误码率. 图 3 为保持一定内容水印比 DWR = 20 dB,固定扩频向量长度 N = 500,水印噪声比 WNR 从 −20 dB 变化到 20 dB 时相关检测下实验测量误码率的变化曲线,为了比较和验证互信息度量模型,也将同样背景条件下的互信息函数绘于同一图中.

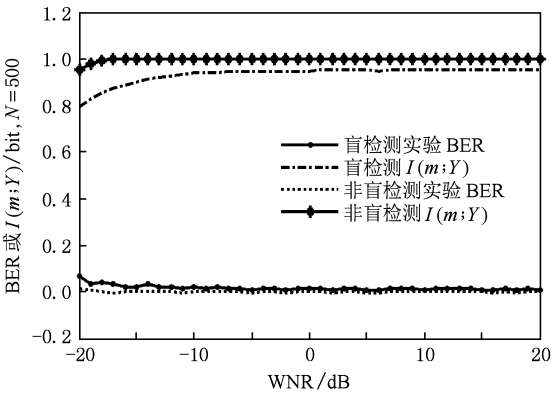

图 3　DCT 域扩频水印误码率与水印噪声比 WNR 间关系

从图 3 中可看出，水印噪声比 WNR 从 −20 dB 到 20 dB，随着 WNR 增大，互信息函数增大，误码率减小，且盲检测误码率大于与非盲检测误码率，盲检测互信息小于非盲检测互信息. 图 3 中 WNR 大于 −10 dB 时，互信息和误码率的曲线分别趋近于稳定最大值和最小值，但由于盲检测时原始载体的干扰，互信息略低于 1，误码率略大于 0. 对于非盲检测，WNR 大于 −15 dB 时，互信息的值较大，几乎为 1，误码率的值非常小，几乎为 0. 从互信息与误码率之间关系看，互信息越大，误码率越低. 所有这些结果表明，互信息函数对水印系统鲁棒性的描述和评测是有效的. 当 WNR 较大，攻击噪声较小时，趋于稳定的最大互信息说明这种情况下系统性能是最鲁棒的.

4.2. 最小距离译码时 QIM 水印的误码率

对标量 QIM 水印，在高分辨率量化的假设前提下，量化步长不能太大，相应的量化误差在量化区间近似服从均匀分布. 由于信噪比要求，相应的应用场景噪声强度也不能太大. 实验中以一段音频（某人的说话录音）为原始载体，选用三级 Daubechies 小波分解后的细节系数作为嵌入水印的量化载体. 仿真中，量化步长取 $\Delta = \frac{\sqrt{12}}{100}$，均方量化失真为 $D_w = \frac{\Delta^2}{12}$，相应的嵌入水印信噪比 SNR（即嵌入水印后音频内容信噪比）为 42 dB. 当噪声标准差取 $\sigma_n = 0.01$ 时，将该强度的噪声叠加到音频载体上，相应的攻击水印信噪比 SNR（受攻击后音频内容信噪比）为 20 dB 左右. 由于不出现明显的感知失真一般要求信噪比 SNR 不小于 20 dB[11]. 所以仿真中选取噪声方差 $\sigma_n^2 \leqslant \frac{\Delta^2}{12}$. 图 4 为高斯噪声条件下噪声标准差在 0.001—0.01 区间，QIM 水印实验的检测误码率随噪声标准差的变化曲线，和预测鲁棒性的互信息函数随噪声标准差变化曲线.

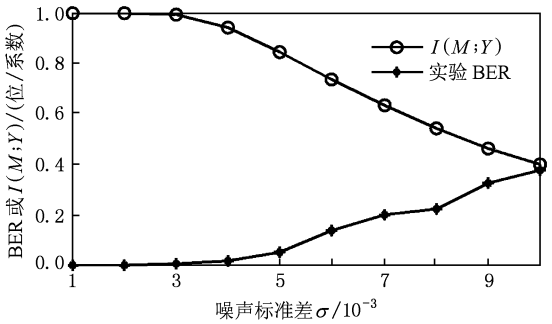

图 4　高斯噪声下 QIM 水印鲁棒性分析：互信息函数、误码率与噪声标准差之间关系

从图 4 中可知，当噪声方差很小时，度量鲁棒性的互信息函数很大，几乎保持最大值 1，而图中相应的实验误码率都很小，几乎为 0. 随着噪声强度的增强，互信息函数很快下降，根据互信息对鲁棒性的预测，误码率将很快增加，而图中实验统计的误码率曲线也验证了这个预测结论.

4.3. JPEG 压缩下水印鲁棒性估计

由于 JPEG 攻击是不可逆的非线性过程，不能像加性噪声那样简单地处理. 但是，根据 Chen[7,21]，Fei 等[22] 和肖俊等[23] 的研究，可以用一个线性模型来近似 JPEG 压缩过程. 下面先简单介绍 JPEG 攻击信道的近似模型，然后在这个近似模型的基础上计算度量鲁棒性的互信息函数.

假设嵌入水印之后的数据 x 是零均值的白高斯信源信号，有损压缩所允许的最大失真为 σ_z^2，即

$$\frac{1}{N}\| y - x \|^2 \leqslant \sigma_z^2, \tag{30}$$

其中 y 表示 JPEG 压缩后的数据，N 表示数据 x 长度. 根据率失真理论，此有损压缩所对应的限失真编码问题可以通过以下测试信道来表示[14,21]：

$$x = y' + z, \tag{31}$$

式中 x 是高斯信源 $N(0,\sigma_x^2)$，y' 是 x 对应的信源码字，服从高斯分布 $N(0,\sigma_x^2 - \sigma_z^2)$，$z \propto N(0,\sigma_z^2)$ 是信源编码误差，z 与 y' 独立，因此 x 与 y' 是联合高斯分布的，并且在给定 x 的条件下，y' 的条件均值和条件方差为

$$E[y' \mid x] = \frac{\sigma_x^2 - \sigma_z^2}{\sigma_x^2} x, \tag{32}$$

$$\mathrm{var}[y' \mid x] = (\sigma_x^2 - \sigma_z^2)\frac{\sigma_z^2}{\sigma_x^2}, \tag{33}$$

因此可用

$$y' = \beta x + z' \tag{34}$$

来表示(31)式所描述的测试信道,其中 $\beta = \frac{\sigma_x^2 - \sigma_z^2}{\sigma_x^2}$, $z' \propto N\left(0, (\sigma_x^2 - \sigma_z^2)\frac{\sigma_z^2}{\sigma_x^2}\right)$,并且 z' 与 βx 统计独立.

因此,与此类型的有损压缩近似的线性模型是:嵌入水印之后的数据首先与一个尺度因子相乘,然后加上一个独立的高斯噪声.在压缩所带来的失真可以预知的情况下,尺度 β 是已知的(假设:σ_x^2 是知道的,因为 x 信号方差仅仅依赖于嵌入函数,不依赖于压缩算法).并且,如果在解码器端知道尺度因子,在提取水印信息之前可以做一个预处理,即

$$y = \frac{1}{\beta} y' = x + n, \tag{35}$$

其中 $n = z'/\beta$ 是加性的零均值高斯噪声,并且与 x 独立,其方差为

$$\sigma_n^2 = \frac{\mathrm{var}(z')}{\beta^2} = \frac{\sigma_z^2}{1 - \sigma_z^2/\sigma_x^2}. \tag{36}$$

由(36)式可知,在 σ_x^2/σ_z^2 很大的情况下,此噪声方差趋向于压缩失真 σ_z^2.从这个意义上来说,可以用加性高斯噪声信道来模拟有损压缩信道,但解码器端要有一个前置的缩放.

将 JPEG 线性近似模型中等价的高斯噪声 $n \propto N\left(0, \frac{\sigma_z^2}{1 - \sigma_z^2/\sigma_x^2}\right)$ 分别代入扩频水印和 QIM 水印的互信息计算(16)和(28)式,可以求得扩频水印和 QIM 水印在 JPEG 压缩下的鲁棒性估计.仿真和实验中扩频水印以 N = 64,lena 图像 DCT 域低中频系数为载体,QIM 水印基于分块 8 × 8DCT 变换,选择每块中频系数 AC(3,3)作为量化载体嵌入水印信息.当压缩因子从 20 变化到 90 时,可得到相应的压缩失真以均方误差度量在[2.34—12.67]范围内.图 5 为 JPEG 压缩下盲检测扩频水印和 QIM 水印的鲁棒性估计结果.由图可知压缩因子越小,互信息函数越小,说明正确提取水印的可能性越小,误码率也将越大,而图 5 中检测误码率随压缩因子减小误码率增加的变化趋势也体现了这点.对比扩频水印和 QIM 水印,在压缩因子较小时,QIM 水印误码率较大,甚至达到 0.5,而当压缩因子较大时,QIM 水印误码率较小,甚至可以无差错检测.这是由于选用固定量化步长时,压缩因子越小,压缩失真越大,不仅图像块高频部分被压缩掉,甚至一些能量较低的图像块的中频部分包括我们选中的水印嵌置值C(3,3)也被压缩了,这部分压缩后的 AC(3,3)系数被置为 0,由于 0 离两个量化器的距离相等,都为 $\Delta/4$,根据译码规则,提取的水印等可能取 0 或 1,误码率将达到 0.5,此时检测器将无法正确提取水印.至于压缩因子大于 80 的情形,因为压缩率小,丢失的信息也少,互信息较大,此时 QIM 水印表现出较小的误码率,甚至 0 误码率的优势.而扩频水印虽然也随压缩因子减小,互信息函数减小,误码率增加,但通过选取合适的阈值,总体性能波动比 QIM 水印小.

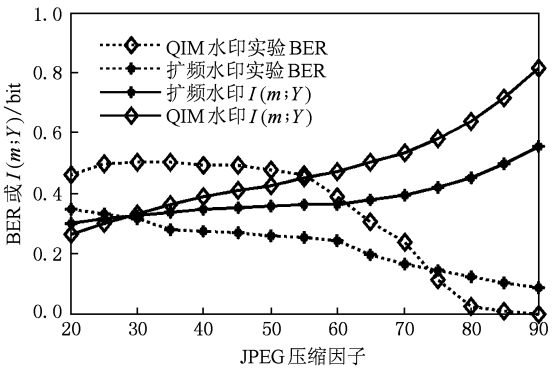

图 5 JPEG 压缩下水印鲁棒性估计

4.4. 水印系统经滤波处理时鲁棒性分析

滤波是图像或音频信号较常采用的一种信号处理操作.接收的内容是嵌入水印后的载体和滤波器脉冲响应的卷积,即

$$y(k) = h(k) * x(k) = \sum_{n=0}^{k} h(n)x(k-n), \tag{37}$$

其中 $h(k)$ 为脉冲响应, $x(k)$ 为嵌入水印后的载体.

可以看出滤波操作的输出是加水印后信号采样点 $x(k)$ 的加权和.设载体信号的采样点都是独立同分布的信号, $x(k)$ 的概率分布模型可以根据水印嵌入算法,类似第 3 节扩频水印和 QIM 水印分析方法求解,在此基础上可以给出 $y(k)$ 的概率分布模型

$$f_Y(y) = \frac{1}{|h(0)|} f_X\left(\frac{x}{|h(0)|}\right) * \frac{1}{|h(1)|} f_X\left(\frac{x}{|h(1)|}\right) * \cdots * \frac{1}{|h(k)|} f_X\left(\frac{x}{|h(k)|}\right). \quad (38)$$

在给定滤波器结构和参数的情况下，可以计算出 $f_Y(y)$，类似地可以计算 $f_{Y|M}(y|m)$，进而求得 $h_Y(y)$ 和 $h_{Y|M}(y|m)$，由互信息公式 $I(M;Y) = h_Y(y) - h_{Y|M}(y|m)$ 可以得到滤波处理后水印内容包含的水印信息量，然后据此评估水印系统经受滤波操作时的鲁棒性. 当滤波器参数未知时，需要先估计滤波器结构参数，再分析鲁棒性，这涉及到另一个研究主题，滤波器的参数估计.

4.5. 水印系统经受其他攻击处理时鲁棒性分析

对于其他的攻击类型，如时间和几何失真. 时间失真影响音频和视频信号，包括延迟和时间缩放. 几何失真影响图像和视频数据，包括旋转、平移、歪斜或错切、比例缩放、对称和投影变形等[1]. 对于一维的音频信号，作品经受时间缩放 k 和延迟 δ，可记为

$$y(t) = x(kt + \delta). \quad (39)$$

对于二维的图像信号，所有这些几何失真可由一个统一的变换形式表示

$$\begin{bmatrix} i'_1 & i'_2 & \cdots & i'_N \\ j'_1 & j'_2 & \cdots & j'_N \\ 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} a & b & k \\ c & d & l \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} i_1 & i_2 & \cdots & i_N \\ j_1 & j_2 & \cdots & j_N \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad (40)$$

其中 $(i_1,j_1),(i_2,j_2),\cdots,(i_N,j_N)$ 表示未失真像素的位置坐标，$(i'_1,j'_1),(i'_2,j'_2),\cdots,(i'_N,j'_N)$ 表示几何失真后对应像素的位置坐标. 当 $b = c = k = l = 0$，且 a,d 不为 0 可得到比例缩放失真. 当 $b = c = 0$，$a = d = 1$ 且 k,l 不为 0 可得到平移失真. 当 $k = l = 0$，且满足 $a = \cos\theta, b = -\sin\theta, c = \sin\theta, d = \cos\theta$，则可得到逆时针旋转角度 θ 的旋转失真. 当 $k = l = 0, a = d = 1, b, c$ 不同时为 0 可得到歪斜或错切失真. 当 $b = c = k = l = 0, a = 1, d = -1$ 可得到 x 轴为对称轴的镜像对称，当 $b = c = k = l = 0, a = -1, d = 1$ 可得到 y 轴为对称轴的镜像对称. 当 $b = c = k = l = d = 0, a = 1$ 和 $b = c = k = l = a = 0, d = 1$ 可分别得到 x 轴和 y 轴上的投影.

当变换矩阵中的参数确定时，可由几何失真后的图像数据完全可逆地恢复出失真前的图像数据，从而检测出水印信息. 当单纯只是几何失真时，接收内容中包含的水印信息量完全由像素的位置信息决定，即

$$I(m;Y) = I(a,b,c,d,k,l;Y). \quad (41)$$

根据互信息的链式法则[14]，$I(X_1,X_2,\cdots,X_n;Y) = \sum_{i=1}^{n} I(X_i;Y|X_{i-1},X_{i-2},\cdots,X_1)$，可以由基本几何失真来求解组合失真. 因此问题的难点就转化为已知基本几何失真后的数据，估计相应的变换参数，这可借助信号检测与估计理论进一步分析，也是目前抗几何攻击鲁棒水印研究的热点和难点，需要进一步研究和另文撰述.

对于各项信号处理操作，假设由 $h_1,h_2,\cdots,h_n$ 代表，则 x 经组合攻击后变为

$$y = h_n(h_{n-1}(\cdots h_1(x))). \quad (42)$$

当各项信号处理操作相互独立并叠加组合时，

$$f(y) = f(h_1(\cdot)) * f(h_2(\cdot)) * \cdots * f(h_n(\cdot)). \quad (43)$$

即各项操作中间结果的分布是卷积的关系. 如滤波处理后经加性噪声信道传输，

$$y_1(k) = h_1(x) = h(k) * x(k) = \sum_{n=0}^{k} h(n)x(k-n), \quad (44)$$

$$y = h_2(y_1) = y_1 + N \quad (45)$$

$$f(y) = f(h_1(\cdot)) * f(h_2(\cdot)) = f(y_1) * f_N(n). \quad (46)$$

当能够建立信号处理操作的概率模型时，可以利用互信息理论求解水印信息和接收内容之间的互信息函数，进而评估水印系统的鲁棒性. 理论上讲只要能对遭遇到的信号处理操作建模，则可以利用互信息度量方式对水印鲁棒性进行评测. 但由于涉及的信号处理多种多样，有些信号处理模型比较复杂，能满足所有信号处理的水印算法往往很难实现. 具体的数学模型和基于此的水印鲁棒性分析有待进一步研究.

5. 结　　论

避开以实验平台测试水印系统的评估方式，通过分析互信息和水印系统鲁棒性之间的关系，以互

信息为代价函数描述和评价一般水印系统的鲁棒性,并指出鲁棒性求解中两种极限状态对应水印信道容量和水印编码率失真函数.以两类经典水印算法:扩频水印和量化水印为范例,给出了高斯噪声JPEG压缩条件下评测鲁棒性的互信息函数计算公式.为了验证互信息度量模型的有效性,通过实验测量和统计水印比特错误率作为误码率.实验结果表明互信息函数评估的鲁棒性与实验误码率结果具有较好的符合度,即互信息函数越大,水印系统越鲁棒,相应的误码率越小.对于其他类型的水印算法以及涉及的各种攻击操作,可以预期,只要能对各种处理精确建模,互信息度量模型作为一种评测工具可评估水印的可能鲁棒性.

[1] Cox I J, Miller M, Bloom J 2007 *Digital Watermarking and Steganography* (2nd ed) (Morgan Kaufmann) p7

[2] Petitcolas F A, Anderson R J, Kuhn M G 1998 *Proceedings of 2nd International Workshop on Information Hiding* Portland, Oregon, USA, April 14—17 p218

[3] http://www. watermarkingworld. org/checkmark/checkmark. html

[4] http://www. watermarkingworld. org/optimark/index. html

[5] Fan X H, Xiao J, Wang Y 2008 *Journal of Image and Graphics* **13** 1979 (in Chinese) [樊晓华、肖 俊、王 颖 2008 中国图形图象学报 **13** 1979]

[6] Xiao J, Wang Y 2009 *Journal of Electronics and Information Technology* **31** 552 (in Chinese) [肖 俊、王 颖 2009 电子与信息学报 **31** 552]

[7] Chen B, Wornell G 2001 *IEEE Trans. on Information Theory* **47** 1423

[8] He H J, Zhang J S 2007 *Acta Phys. Sin.* **56** 3092 (in Chinese) [和红杰、张家树 2007 物理学报 **56** 3092]

[9] Zou L J, Wang B, Feng J C 2008 *Acta Phys. Sin.* **57** 2750 (in Chinese) [邹露娟、汪 波、冯久超 2008 物理学报 **57** 2750]

[10] Song W, Hou J J, Li Z H, Huang L 2009 *Acta Phys. Sin.* **58** 4449 (in Chinese) [宋 伟、侯建军、李赵红、黄 亮 2009 物理学报 **58** 4449]

[11] Cox I J, Killian J, Leighton T 1997 *IEEE Trans. on Image Processing* **6** 1673

[12] Feng J C, Yu Z B 2008 *Acta Phys. Sin.* **57** 1409 (in Chinese) [冯久超、余振标 2008 物理学报 **57** 1409]

[13] Poor H V 1998 *An Introduction to Signal Detection and Estimation* (2nd ed) (New York: Springer)

[14] Cover T M, Thomas J A 2006 *Elements of Information Theory* (2nd ed) (New York: Wiley) p11 [阮吉寿、张 华 2007 (译) 北京: 机械工业出版社 第 11 页]

[15] Gray R M, Neuhoff D L 1998 *IEEE Trans. on Information Theory* **44** 2325

[16] Pérez-Freire L, Pérez-González F 2008 *IEEE Trans. Information Forensics and Security* **3** 593

[17] Nie Z P, Xiao H L 2007 *Acta Phys. Sin.* **56** 1948 (in Chinese) [聂在平、肖海林 2007 物理学报 **56** 1948]

[18] Clarke R J 1985 *Transform Coding of Images* (New York: Academic)

[19] Sun Z W, Feng D G 2005 *Journal of Software* **16** 1798 (in Chinese) [孙中伟、冯登国 2005 软件学报 **16** 1798]

[20] Buccigrossi R W, Simoncelli P 1999 *IEEE Trans. on Image Processing* **8** 1688

[21] Chen B 2000 *Ph. D. Dissertation* (Cambridge, USA: Massachusetts Institute of Technology)

[22] Fei C, Kundur D, Kwong R 2001 *Proceedings of International Conference on Information Technology* Los Alamitos Las Vegas, NV, USA, April 2—4 p79—84

[23] Xiao J, Wang Y, Li X L 2007 *Acta Electronica Sin.* **35** 786 (in Chinese) [肖 俊、王 颖、李象霖 2007 电子学报 **35** 786]

Evaluation model for robustness of digital watermarking*

Zeng Gao-Rong† Qiu Zheng-Ding

(*Institute of Information Science*, *Beijing Jiaotong University*, *Beijing* 100044, *China*)

(Received 31 October 2009; revised manuscript received 23 November 2009)

Abstract

Robustness is one of the most important requirements when digital watermarking is applied. Different from the StirMark test and various simulation tests, a mutual information function is defined as a criterion to measure the robustness of watermarking algorithm. Taking the additive spread spectrum watermarking scheme and quantization index modulation (QIM) watermarking scheme as two examples, the calculation formulas of mutual information function are derived to evaluate the robustness of the algorithms. Numerical computation of mutual information is performed with change of watermark noise rate (WNR). In the experiment, spread spectrum watermarking is implemented in discrete cosine transform (DCT) and QIM watermarking is implemented in discrete wavelet transform (DWT). The statistic bit error rate (BER) is derived against Gaussian distribution noise and JPEG compression. Experiment results show that the evaluation conclusion of mutual information method is in accordance with the empirical BER. Mutual information can be selected as a cost function to predict the BER.

Keywords: digital watermarking, robustness, mutual information, bit error rate (BER)

PACC: 8710, 0250

* Project supported by the National Key Technology Research and Development Program of the Ministry of Science and Technology of China (Grant No. 2008BAH33B01), the National Natural Science Foundation of China (Grant No. 60773015), the National High Technology Research and Development Program of China (Grant No. 2007AA01Z460).

† E-mail: loch. zeng@ gmail. com