基于自然图像模型的扩频水印安全性研究

专　　业：通信与信息系统

博 士 生：张　东

指导教师：倪江群 教授

摘　　要

随着信息技术的发展和生活水平的提高，人们对于信息的安全问题越发重视。作为信息安全领域的主要技术之一，数字水印技术受到了科研人员的广泛关注。传统上数字水印的性能受到三个方面的制约：鲁棒性、不可见性和水印容量。近年来，数字水印的安全性逐步成为水印技术研究的新热点。

数字水印的安全性是衡量在水印算法细节公开的条件下水印密钥抵抗非授权攻击（估计）的能力。由于任何水印算法都不能永久保密，水印算法的细节必将公开，只能依靠水印密钥来保证水印系统的安全。密钥在水印系统中具有重要的作用。密钥的安全意味着水印系统的安全。对水印安全性攻击的目的是为了获取关于水印密钥的知识并最终实现对水印系统的完全破解，因此安全性对于水印系统尤为重要。水印安全性已经和水印不可见性、水印容量及水印鲁棒性一起成为水印性能的四大制约因素。

基于扩频调整的水印是最为广泛应用的技术之一，其安全性受到了学术界的格外重视。由于扩频水印安全性与水印载体信号的统计分布有密切联系，精确描述载体信号统计分布对于扩频水印安全性分析至关重要。原有的扩频水印安全性分析使用高斯模型来描述图像载体小波系数的统计分布，并不能准确刻画载体信号的实际分布特点，得出的分析结论与实际情况具有很大距离。本文利用自然图像模型实现对于图像载体小波系数统计分布的精确描述，并据此实现了对加性扩频水印和改进的扩频水印安全性的准确分析。

加性扩频水印是最为传统的扩频水印算法。对加性扩频水印的安全性分析中，本文利用高斯尺度混合模型描述自然图像载体小波系数的统计分布，通过Fisher信息来衡量水印通信过程中关于水印密钥的信息泄露。分析过程中使用攻击者对水印密钥无偏估计可达的最小误差来表示水印的安全程度。本文实现了对KMA和WOA条件下加性扩频水印安全性的理论分析和实验仿真，并证明原有的基于高斯模型的分析方法高估了加性扩频水印的安全性能。

改进的扩频水印算法的嵌入过程与载体信号的特点相关，是一种“有信的”水印嵌入方法。传统的加性扩频水印算法可以看作是改进的扩频水印算法的一种特例。在对改进的扩频水印算法安全性的分析中，本文利用高斯尺度混合模型描述自然图像小波系数的统计分布，通过Shannon互信息来表示水印通信过程中关于水印密钥的信息泄露并以此作为水印安全程度的衡量标准，实现了在KMA和WOA条件下对改进的扩频水印安全性的定量分析和实验仿真。

由于高斯模型描述的随机变量在平均功率相同条件下具有最大的不确定性，使用高斯尺度混合模型可以更加准确地刻画自然图像载体分布的特点，基于此模型的扩频水印安全性分析较以往工作更为准确和切合实际。本文的工作为进一步设计新一代鲁棒和安全的水印算法提供了依据。

关键词：

扩频水印，水印安全性，自然图像模型，Fisher信息，Shannon互信息

Spread-Spectrum Watermarking Security Incorporating Statistics of Natural Images

Major: Communications and Information Systems

Name: Dong Zhang

Supervisor: Professor Jiangun Ni

Abstract

The issue of information security has attracted people's great attention since two decades ago. Digital watermarking is one of the major areas of information security. Traditionally, robustness, imperceptibility and capacity have been considered as the three main constrains in the development of watermarking algorithms. In recent years, security has emerged as the domain of extensive research of watermarking.

Watermarking security measures the performance of watermarking scheme to resist intentional estimation (attacking) to secret key conditioned the algorithm details are publicly known. Since no algorithms could be kept in secret for ever, secret key is the only factor to ensure the security of watermarking. The security of secret key is actually equivalent to the security of watermarking system. Unlike the concept of robustness which deals with blind attacks, the security is more critical to watermarking as it deals with the intentional attacks which target to get knowledge of secret key, therefore offering complete break. The issue of security has become a fundamental constraint to be respected in order to guarantee the usability of a watermarking technology.

As a widely used watermarking technology, Spread-Spectrum (SS) based watermarking attracts great literature attention to its security. Since the distribution of host impacts the security of SS-based watermarking intimately, an accurate

description of host distribution is critical to achieve a perfect result of security analysis. Previous works acquired the performance of security for SS-based watermarking with the assumption of Gaussian host. Actually, the distribution of wavelet coefficients of natural images characterizes great non-Gaussian. With the help of natural images statistics, this dissertation analyzes the security of Additive Spread-Spectrum (Add-SS) watermarking as well as that of Improved Spread-Spectrum (ISS) watermarking.

Add-SS watermarking is one of the conventional schemes. This dissertation models the distribution of wavelet coefficients of natural images with Gaussian Scale Mixture (GSM) and measures information leakage of secret key with Fisher information. The minimum achievable error with an unbiased estimator is employed as the measurement of watermarking security. The security performances of Add-SS based watermarking under Known Message Attack (KMA) situation and Watermarked Only Attack (WOA) situation have been analyzed. With theoretical analysis and empirical simulation, this dissertation reveals the security of Add-SS watermarking was overestimated by previous works.

ISS watermarking implements an informed embedding as it adapts the embedding with hosts. Add-SS based watermarking can be regarded as a special case of ISS algorithm. This dissertation presents a theoretical analysis on the security of ISS watermarking from the viewpoint of Shannon Information theory with the help of GSM model to characterize the statistics of natural images. By using mutual information as a measurement of information leakage of secret key, this dissertation implements security analysis of ISS watermarking under KMA and WOA situations.

Because Gaussian random variable expresses the maximum uncertainty when average power is constrained, GSM model characterizes the statistics of natural images more accurately than Gaussian model does due to great non-Gaussian of natural images. Consequently, the security analysis based on GSM is more precise than that based on Gaussian model for Spread-Spectrum watermarking. This dissertation is believed to provide help to design the new generation of secure and robust watermarking algorithm.

Keywords

Spread-spectrum watermarking, watermarking security, natural images statistics, Fisher information, Shannon mutual information

论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：张东

日期：2009年12月5日

学位论文使用授权声明

本人完全了解中山大学有关保留、使用学位论文的规定，即：学校有权保留学位论文并向国家主管部门或其指定机构送交论文的电子版和纸质版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆、院系资料室被查阅，有权将学位论文的内容编入有关数据库进行检索，可以采用复印、缩印或其他方法保存学位论文。

学位论文作者签名：张杰　　导师签名：

日期：2009年12月5日　　日期：2009年 12月 5 日

第一章 绪论

1.1 数字水印简介

随着信息技术的飞速发展，越来越多的家庭拥有计算机、数码相机、摄像机等多媒体设备，人们能够非常容易地获取和制作各种多媒体作品。由于计算机和计算机应用技术的逐渐普及，特别是互联网技术已经走向千家万户，对于文本、图像、音频、视频等多媒体文件的制作、复制、编辑、传输等处理手段已经变得越来越简便。在今天，甚至非专业人士都可以利用数码设备和一些多媒体处理软件编辑制作出精美的多媒体作品。

在利用多媒体技术丰富生活的同时，人们对于多媒体作品的安全意识也逐步提高。因此，人们迫切需要相关的技术来保护多媒体信息的安全，防止多媒体信息在非授权的情况下被传输、复制、编辑或破坏。

传统上，人们可以使用密码技术来保护多媒体信息的安全。多媒体作品的所有者可以利用密码技术对多媒体信息进行加密，并且将加密信息的密钥分发给被授权的用户。被授权的用户可以对接收到的多媒体信息进行正确解密，而未被授权的接收者尽管也能接收到加了密的多媒体文件，但是因为没有密钥，不能进行正确解密，从而不能得到正确的多媒体信息。但是，利用密码学的方法只能保护多媒体信息的内容在解密前不被侵犯，而不能保护解密后的多媒体信息。例如已经购买了多媒体作品副本的合法用户，将多媒体作品解密之后，在未经许可的情况下可能继续分发该多媒体作品。对于这样的行为，密码学技术则无能为力。

因此，需要一种技术，能够作为对密码学技术的补充，对解密后的多媒体作品进行保护。这种保护作用不会因多媒体作品的解密而失效，而且其有效性能够伴随多媒体作品而存在。数字水印技术就是一种能够起到这种作用的技术。数字水印技术可以在不为人们感知的条件下将版权信息隐藏或嵌入多媒体作品内部。除非多媒体作品被破坏至失去意义，即使经过对作品的再次加密/解密、压缩、复制、数-模转换以及文件格式变换，隐藏的版权信息都能够随多媒体作品永远存在[1]。

数字水印技术的特点主要表现在三个重要方面[1]：首先，数字水印是不可感

知的，通过数字水印方法嵌入的信息将不会影响人们对于多媒体作品的感觉；第二，嵌入的数字水印与多媒体作品密不可分，即使改变了多媒体作品的存储格式，嵌入的数字水印依然存在；第三，数字水印将与多媒体载体作品经历相同的变化和处理过程。

由于数字水印的以上特点，数字水印技术被广泛地应用于媒体所有权的认定和保护、非法拷贝防护、媒体的真伪鉴别，保密通信、多语言电影系统和电影分级以及数字媒体附件描述和参考信息的携带等领域[2]。

有关数字水印的研究始于20世纪90年代，现在已经成为学术界研究的热点之一。数字水印涉及到信号处理、信息论、统计学、计算机科学、密码学和人类视/听系统研究等领域，是多学科交叉的新兴研究方向。

1.2 数字水印的应用

数字水印技术最早是作为版权保护的手段而产生的。随着研究的不断深入，数字水印技术的应用也越来越广泛。这些应用主要包括：版权保护、内容认证、广播监视、设备控制、拷贝控制、隐秘通信、数字指纹等方面。

（1）版权保护

传统的利用文本进行版权声明和保护的方法具有很多局限性，例如容易被去除、篡改和伪造[3]。利用数字水印技术实现版权保护，可以使版权信息以不可见的形式嵌入到原始多媒体作品里。在多媒体作品经历多种操作处理之后，只要多媒体作品的内容不被破坏至失去意义，嵌入的版权信息将仍然存在。这种应用的条件要求数字水印能够抵抗加/解密、文件拷贝、几何变换、模/数和数/模转换等信号处理操作。具有这样的性能的水印被称为鲁棒水印[4]。

（2）内容认证

利用数字水印技术进行内容认证可以判别多媒体作品是否曾经被修改[5]。随着计算机技术的迅速提高，人们利用计算机对多媒体作品进行修改变得越来越容易。例如，人们可以利用Adobe Photoshop®软件对拍摄的数码照片进行修改，人眼却很难发现作品被修改的痕迹。与用于版权保护的数字水印不同，在原始的多

媒体作品中嵌入的认证水印信息在作品发生改变时便会丢失。人们可以通过检测作品中是否仍然存在水印来判断该作品是否曾经被修改过。具有这样性能的水印被称为“脆弱水印”[6]。近年来在新闻界曾发生了多次记者修改照片报道不实新闻的事件[7]，因此利用脆弱水印实现内容认证的技术受到广泛关注。与此相似的是被称为“篡改取证”[8]的技术，即通过分析作品中的水印来判断原始多媒体作品曾经经历了哪些操作，甚至可以发现该作品的哪些地方曾被修改，并为检察机关的调查提供证据。

（3）广播监视

用于广播监视的数字水印技术将水印信息嵌入到将要播出的节目中，在节目的收视端通过检测水印来判断该节目是否已按照协议完整播出[1]。例如广告公司在广告中嵌入水印，付费给电视台用于播出一定次数的广告。广告公司可以通过统计在电视信号中检测到水印的次数来核实电视台是否正常履行协议。

（4）设备控制和拷贝控制

多媒体作品的播放设备可以根据在作品中检测到水印的信息作出相应的动作。例如使用设备控制技术的 DVD 播放机可以根据检测 DVD 影碟中的水印信息内容，来播放作品的某一部分。在文献[1]中，还描述了伴随电视节目运动的玩具，即玩具通过扫描电视节目中藏有的水印信号，来获得玩具运动的同步信号，这样玩具可以实现与电视信号协同运动。

用于拷贝控制的水印包含着允许或禁止多媒体作品被拷贝的信息。这种技术需要相关硬件设备的支持。当设备读取到多媒体作品中的水印信息为“禁止拷贝”时，便阻止拷贝行为的发生。当然，通过水印技术实现拷贝控制不仅仅是技术层面的问题，还涉及到许多商业和法律方面的问题。

（5）隐秘通信

数字水印技术在不被人感知的情况下将信息嵌入到多媒体作品中，因此可以将从水印嵌入到检测的过程看作是以作品为载体来传递嵌入信息的隐秘通信过程。隐秘通信要求能够在多媒体作品中嵌入尽可能多的信息（即要求大容量），

而且不能被对手发现曾经实施了信息嵌入。隐秘通信技术在军事上和安全领域具有很高的研究价值。

（6）数字指纹

数字指纹技术使用一定算法，在每个分发的作品副本中嵌入不同的水印信息，例如该副本的 ID 号。这样，如果某个合法副本的用户在未经授权的情况下进行再次拷贝致使作品被滥用，作品的所有者就可以根据再次拷贝中的水印信息，例如 ID 号，来追究非法拷贝的源头[9]。

1.3 数字水印的模型和特性

数字水印的应用可以被看作是一种保密通信的过程，并可由图 1-1 所示的模型[10]来描述。其中，多媒体作品（S^N）被称为载体信号，其所有者利用密钥（K^N）将秘密信息（M）经过一定嵌入算法隐藏入载体信号中，并生成嵌入了水印的信号（X^N）。嵌入了水印的信号在传输中经历攻击信道（A^N）后被接收者观测得到（Y^N）。接收者根据接收到的信号和密钥（K^N），利用译码器进行译码，可以得到传输的秘密信息。

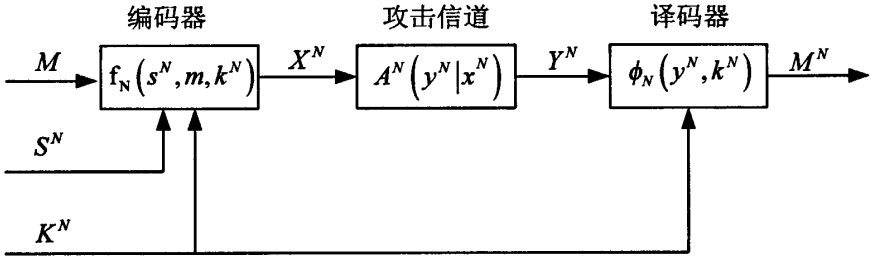


图 1-1 数字水印通信模型

Fig. 1-1 Model of digital watermarking communication

在以上数字水印的通信模型中，载体可以是任意的文本、图像、音频、视频等多媒体作品。秘密信息的发送方和接收方共同掌握相同的密钥。在通信过程中，嵌入了数字水印的信号受到的攻击可以为噪声、对载体的几何变换以及任何常规的信号处理过程。

由于自然图像作品易于获得，而且以自然图像为载体得到的研究成果可以方便地推广到音频、视频等领域，因此本文所述的研究都使用自然图像作为载体。为了叙述的方便，在不引起混淆的前提下，本文将嵌入了水印的图像简称为水印图像，对数字水印简称为“水印”并不做区分。

传统上，数字水印系统的性能由水印嵌入过程中的有效性、保真度（不可见性）、有效载荷以及检测过程中涉及的盲检测、含辅助信息的检测、检测器的虚警行为和鲁棒性来衡量[1]。嵌入有效性是指水印嵌入后直接进行检测时能检测到水印的概率。水印系统的保真度反映了嵌入水印后的作品和原始的载体作品在感觉上相似的程度。水印的有效载荷指的是在一个作品中能够嵌入秘密信息的比特数，有效载荷的性能也常用水印容量来代替。在水印的检测方，不需要原始的载体作品的检测器称为盲检测器；反之，需要原始载体作品的检测器称为含辅助信息检测器。检测器的虚警行为是指检测器在不含有水印信息的作品中检测出水印的行为。鲁棒性是指含水印的作品在经历了常规的信号处理操作后，检测器仍能够检测到水印的能力，其中对图像的常规信号处理操作包括空间滤波、有损压缩、扫描和打印以及进行几何变换[1]。传统上，在所有的水印性能中，最为受到研究人员关注的是水印的不可见性、水印容量和水印鲁棒性。任何成功的数字水印技术都需要兼顾这三种性能，它们构成了数字水印研究的基础。

随着对数字水印技术研究的不断深入，水印系统安全性成为人们更为关注的话题，并逐渐成为支持数字水印研究的第四个性能要求。

1.4 常用的数字水印嵌入方法

根据水印实现方法的不同，水印技术可以分为空域法和变换域法[2]。在空域法实现的水印算法一般通过改变图像像素的灰度值来实现水印的嵌入；利用变换域法实现的水印算法首先对图像进行变换（例如傅立叶变换、离散余弦变换、小波变换，等等），然后在变换域中嵌入水印信息，再实施逆变换得到嵌入了水印的图像。相比之下，通过变换域法实现的水印能够将嵌入信息的能量散布到图像载体的每个像素上，易于提高水印的不可见性和鲁棒性。而且由于变换域方法与很多国际通用的图像压缩标准相兼容，因此易于实现在压缩域内的水印嵌入[2]。

根据数字水印嵌入方法的不同，常用的数字水印主要可以分为利用载体信号

最低有效位（Least Significant Bit，LSB）进行嵌入的水印方法、基于扩频序列（Spread-Spectrum Sequence）的水印方法和基于量化索引调制（Quantization Index Modulation, QIM）的水印方法等。

1.4.1 利用载体 LSB 进行嵌入的水印方法

利用载体信号最低有效位进行水印嵌入的方法是将秘密信息的每一位嵌入到载体信号的 LSB 上。如果图像载体信号的每一个像素的灰度值或者图像 DCT 变换的每个系数可以用 8 bit 表示，即 $b=\{b_7b_6b_5b_4b_3b_2b_1b_0\}$，则其中 b_0 即为该载体信号的 LSB。嵌入时将秘密信息替换 b_0 的值；在检测时，只要直接提取接收到信号的 LSB 即可。这样，每个像素点将能够嵌入 1 bit 的秘密信息。如果需要增加嵌入容量，可以进一步利用到第二个 LSB 平面进行嵌入。但是这样会引入更大的嵌入失真。为了增强水印的安全性，水印的嵌入者和合法的检测者可以共同拥有水印的密钥，用于确定嵌入水印信息的位置。

基于 LSB 嵌入的水印方法非常简单，而且便于检测。但是这种方法的鲁棒性不高，例如只要攻击者对水印图像像素的 LSB 平面进行随机化，就可使水印失效[10]。基于 LSB 的嵌入技术常被用做隐写术的方法[11]，以及用于进行内容认证的脆弱水印。

1.4.2 基于扩频序列的水印方法

（1）加性扩频水印

传统的扩频水印嵌入方法可以由图 1-2 表示[10]。其中 $\mathbf{s}$ 表示图像载体信号，m 为将要嵌入的秘密信息，$\mathbf{k}$ 表示使用的密钥，γ 为嵌入强度，$\mathbf{x}=\mathbf{s}+\gamma p(m,\mathbf{k})$ 是生成的嵌入了水印的图像。在实现过程中，使用的密钥就是扩频序列（秘密载波），并由水印嵌入者和合法的水印接收者共同掌握。水印的嵌入者使用被传输的秘密信息对扩频载波进行调制，通过参数 γ 来调节嵌入水印强度后直接加至载体信号上生成水印图像，因此传统的扩频水印也被称为加性扩频水印。

在水印的检测方，接收到的是含有噪声干扰的水印图像，即 $\mathbf{y}=\mathbf{x}+\mathbf{n}$，其

中 $\mathbf{n}$ 为噪声。由于合法的接收者拥有水印密钥（即秘密载波），该接收者可利用可能的 $|M|$ 个模式和接收到的信号做线性相关，取其最大值就可以正确检测出传输的秘密信息[10]。检测的过程可表示为 $\hat{m}=\arg\max_{m\in M} t_m(\mathbf{y},\mathbf{k})$，其中，$t_m(\mathbf{y},\mathbf{k})=\sum_{n=1}^{N} y_n p_n^{(m,k)}, m\in M$。

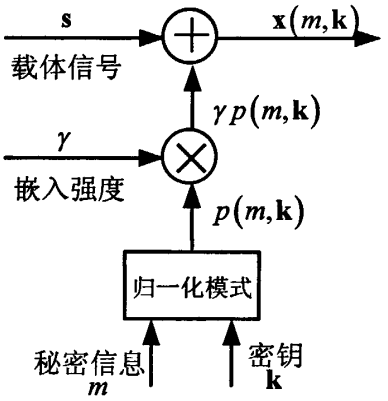


图 1-2 加性扩频水印的嵌入方法

Fig. 1-2 Embedding of Add-SS based watermarking scheme

加性扩频水印方法以通信领域中的扩频通信[12]为理论基础，其实质是在一个高宽带的信道（图像载体）中传输一个窄带信号（水印）[2]。由于水印信号分散到信道任意频率的能量很小，因此基于扩频序列的水印方法具有很好的不可见性和鲁棒性。同时，因为算法简单、实用，加性扩频水印在理论上和实践上都有广泛的研究和应用价值。

（2）改进的扩频水印

从加性扩频水印的嵌入方程可以看出，水印的嵌入与载体信号相独立。因此，载体信号相当于是通信的秘密信息的干扰，而且与通信中的其它噪声相比，载体信号造成的干扰最大。可以想象，如果在不考虑失真约束的条件下，水印图像中如果不存在图像载体的成分，携带秘密信息的扩频载波将仅受到其它通信噪声的影响，检测器的输出性能将显著提高。

由于水印算法受到不可感知性的约束，因此嵌入水印造成的失真不能太大，实用的水印在嵌入时不能完全除去载体信号。文献[13] 在水印嵌入时用取值在[0, 1]区间衰减因子引入对载体信号的衰减。文献[14]进一步提出改进的扩频水

印(Improved Spread Spectrum Watermarking，ISS 水印)，根据载体信号的特点，仅在扩频载波方向对载波进行衰减。由于 ISS 水印在鲁棒性和嵌入率等性能方面都有明显提升，而且可以看做是传统的加性扩频水印的推广，ISS 水印日益成为扩频水印的主流。

ISS 水印的嵌入方程可以表示为

$$\mathbf{x}=\mathbf{s}+\mu(\mathbf{s},m)\mathbf{k}$$

其中各变量的含义与加性扩频水印框图所述相同。可以看出 ISS 水印的嵌入过程与载体信号有关，因此属于“有信”的嵌入方法。实际中经常使用线性模型来实现 ISS 水印嵌入，即将嵌入过程表示为

$$\mathbf{x}=\mathbf{s}+(-1)^m v\mathbf{k}-\lambda\frac{\mathbf{s}_f^T\mathbf{s}}{\|\mathbf{s}\|^2}\mathbf{k}$$

ISS 水印的嵌入方程中，v和λ都是取值在[0, 1]区间的实数，分别用于控制嵌入秘密载波的强度和对载体信号的衰减。由于$\frac{\mathbf{s}_f^T\mathbf{s}}{\|\mathbf{s}\|^2}\mathbf{k}$表示载体信号$\mathbf{s}$在扩频载波$\mathbf{k}$方向的投影，ISS 水印只是在载波方向衰减载体信号。这样，一方面减小了载体信号对于扩频载波(被秘密信息调制过的)的干扰，有助于提高检测的性能；另一方面，由于对载体的衰减仅仅发生在扩频载波方向，并不改变其它方向的载体分量，有助于减小载体的失真。当λ的取值为 0 时，引入对载体的衰减为 0，此时 ISS 水印退化为传统的加性扩频水印。

ISS 水印实现了水印信号的“有信”嵌入，其实质是在失真相同条件下通过对载体的衰减加强了可嵌入的水印能量，因此提高了水印的鲁棒性。在强噪声攻击条件下，ISS 水印具有和 QIM 水印相当的性能。

1.4.3 基于 QIM 的水印方法

基于 QIM 的水印方法最早由 Chen 和 Wornell 于 1999 年提出[15, 16]。这种方法以 Costa 的“污纸编码”[17]理论为基础，根据秘密信息的不同，使用不同的量化器对载体信号进行量化。QIM 的基本原理如图 1-3 所示[10]。

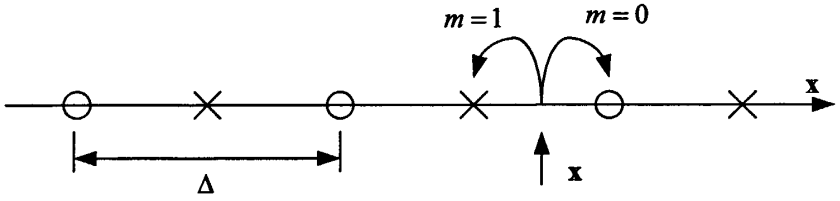


图 1-3 QIM 的基本原理

Fig. 1-3 Fundamental of QIM

图 1-3 中两个量化器的量化格点分别用“0”和“×”表示。每个量化器的量化步长都为Δ。x 为载体信号，m 为嵌入的秘密信息。如果嵌入的秘密信息为“1”，则对载体信号使用由“×”表示的量化器进行量化；若嵌入的秘密信息为“0”，则使用由“0”表示的量化器进行量化。记量化后的信号为$s(\mathrm{x},\mathrm{m})$。当嵌入信息相同，而载体信号不同时，量化后的信号值会在不同的“0”或不同的“×”之间变换，而不会由“0”变为“×”。对于相同的载体图像，如果嵌入不同的秘密信息，将会使用不同的量化器，量化后的信号将在距离载体“最近”的“0”节点或“×”节点之间选择。基于 QIM 的水印嵌入方法可以看作是将信源编码和信道编码同时进行的技术[15]。

水印图像在传输过程中会受到各种干扰或攻击，在 QIM 水印系统的解码端，可以利用“最小距离解码器”来实现对嵌入信息的解码检测[15]，即：

$$\hat{m}(y)=\arg\min_{m}\left\|y-s(y,\mathrm{m})\right\|$$

其中 y 为检测端接收到的受到噪声干扰的水印图像。

基于 QIM 的基本原理，文献[16]又提出了基于失真补偿的量化索引调制（Distortion-Compensated QIM）水印方法。同时，为了提高水印方法的实用性，在水印嵌入时利用秘密信息生成抖动向量，并根据抖动向量对基本量化器进行平移得到与嵌入信息相对应的量化器组，根据嵌入信息的不同采用不同的量化器实现对载体信号的量化。

基于 QIM 的水印方法具有大容量和方法简单等优点，而且量化器的性能直接反映了水印的性能。例如，QIM 水印系统中使用的量化器的数目决定了水印的嵌入率；量化器量化节点的分布形状和量化步长直接影响到水印的失真性能；相邻量化器重建节点之间的距离则决定了水印系统的鲁棒性。因此，基于 QIM

的水印系统能在率失真和鲁棒性之间取得更好的折中，并且被证明在加性高斯白噪声信道和均方误差约束的攻击信道中是容量最优的[16]。

1.5 数字水印密钥

密钥被广泛使用在数字水印算法中。数字水印的密钥常被用于产生嵌入函数的某些参数[18]，例如水印的嵌入区域，嵌入方向，量化抖动参数等等。

文献[19]提出在基于 LSB 的水印算法中利用 m 序列[20]作为水印密钥。水印嵌入方先将秘密信息用不同的m序列来表示，然后将m序列或其补(complement)嵌入到图像载体的 LSB 平面上。水印的检测方只需直接提取水印图像的 LSB 平面，并将其与所掌握的密钥（m 序列）进行比较即可得到嵌入的秘密信息。由于 m 序列具有平衡性、伪随机性和类似于高斯分布的自相关性，使得该方法生成的水印具有更好的安全性和视觉质量，并且易于解码和检测。

Cox 等人[21]在使用伪随机序列的基础上，提出利用高斯随机矢量来生成独立同分布的水印载波（密钥）能够更好地抵抗共谋攻击(Collusion Attack)。为了保证检测方在水印图像经历了有损压缩、一般的图像处理、几何变换和数/模-模/数转换后仍然能够检测出水印，文献[21]提出将水印嵌入到载体的感知重要区域，例如图像 DCT 系数的低频区域（除去直流分量），并使用感知模板来控制水印的嵌入强度以保证水印的不可见性。该方法显著提高了水印的鲁棒性和抵抗共谋攻击的能力。

为了提高水印系统的安全性，Fridrich[22]提出利用水印密钥来生成变换基函数，在变换域嵌入水印。该密钥由水印嵌入方和合法的检测方共同掌握。嵌入时由密钥作为种子生成在[0,1]区间均匀分布的伪随机序列，该序列经过低通滤波器平滑之后进行 Gram-Schmidt 正交化[23]，其结果与图像载体进行内积后用于嵌入水印。该算法的实质是由密钥来确定对于具体图像的变换基函数。由于不知道水印密钥，攻击者难以找到嵌入水印的变换域（或嵌入方向），从而难以展开对水印系统的攻击。该算法能够有效抵抗对水印的敏感性攻击[18]。

在基于量化索引调制的水印方法中，密钥被用于产生秘密的码本并对图像载体进行量化，水印就是对载体量化前后的变化量。为了提高基于量化的水印安全性，Eggers 等人[24]在 DC-DM QIM（失真补偿-抖动调制的量化索引调制）水印

算法的基础上，提出使用密钥来产生一个额外的随机抖动量，并附加在由嵌入信息决定的抖动量上，以此生成一系列量化器组。攻击者由于没有掌握密钥，因此不能得到水印嵌入时使用的码本。在提高水印安全性的同时，由密钥决定的随机抖动量的引入并不影响水印通信的其它重要性能。

在研究利用边信息（Side Information）的扩频水印时，Furon 提出了 JANIS（Just Another N-Order Side-Informed Scheme）[25]的方法。该方法中使用密钥来确定一系列检测函数，在固定检测函数的条件下对水印的嵌入进行优化。由于充分利用了水印图像的高阶统计特性，JANIS 方法取得了很好的检测性能。

总得来说，在利用密钥的数字水印方法中，密钥Θ被水印嵌入方和合法的检测方共同掌握。水印密钥经过一个映射函数$f(\cdot)$的作用生成秘密的嵌入参数$f(\Theta)$，并应用于水印的嵌入和检测。合法的检测者能够利用密钥实现对秘密信息的检测，而非法的检测者由于没有密钥的信息，只能够将攻击的能量平均地分散到水印图像上，针对于嵌入了水印区域的攻击能量就相对减小。所以，使用密钥的数字水印方法首先防止了秘密信息受到非法的嵌入和检测；同时也提高了被保护内容抵抗鲁棒攻击的能力[18]。因此，目前常用的水印系统都广泛使用密钥。

1.6 数字水印安全性

著名的“囚徒问题”[26]提出了许多与信息安全相关的问题，数字水印的安全性就是其中之一。数字水印技术面临着多种层次、多种目的的攻击。对于水印安全性的攻击是以获取水印密钥知识为目的的攻击。由于对水印密钥的正确估计意味着对水印系统的完全破解，水印密钥的安全就意味着水印系统的安全。水印安全性对于水印系统至关重要。

1.6.1 “囚徒问题”

在著名的“囚徒问题”[26]中，Alice 和 Bob 是监狱中的两个囚徒，Eve 是监狱的看守。Alice 和 Bob 被允许在 Eve 的监视下进行通信。Alice 和 Bob 希望通过表面正常的信件相互交换一些秘密的信息，而 Eve 则会对 Alice 和 Bob 之间的通信进行检查。Eve 对于正常的信件予以放行。这里假设 Alice 和 Bob 共同拥有

隐秘通信的密钥，而且都可以实施对秘密信息的嵌入和检测。假设 Eve 有足够的计算能力，能够对 Alice 和 Bob 之间所有通信进行检测。

“囚徒问题”可以由图 1-4 表示[27]。其中“x”表示原始的图像载体，m 是 Alice 将要传递给 Bob 的秘密信息，k 是 Alice 和 Bob 所共有的密钥。Alice 可以利用（或不利用）图像载体、密钥和秘密信息生成编码 c，然后将 c 嵌入到图像载体中，得到水印图像 y。Alice 可以通过公共信道选择发送水印图像或原始图像载体。Eve 对公共信道进行监视并检测 Alice 发送的信息。Bob 从公共信道中接收信息，利用密钥实施对编码的提取和对秘密信息的解码。

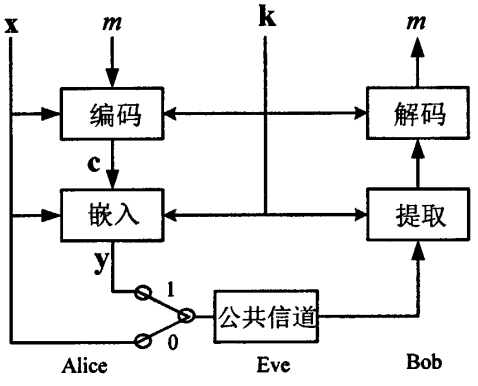


图 1-4 囚徒问题

Fig. 1-4 Prisoner's problem

根据看守者（Eve）发现信件中藏有隐秘信息后处理行为的不同，可以将看守者分为被动和主动两类：如果仅仅阻断该次通信，Eve 就被称为被动的看守者；如果得知信中有隐秘信息存在，会进一步估计隐秘信息并对其进行篡改，Eve 则被称为主动的看守者。对于被动的看守者，他（她）面临的是一个假设检验的问题。由于 Alice 可能发送水印图像或者原始图像载体，Eve 需要根据所掌握的知识实施判决：含有秘密信息或不含有秘密信息。这就是隐写分析[28]（Steganalysis）所主要研究的问题。对于主动的看守者，由于在发现隐秘通信存在之后还要进一步估计秘密信息的内容并对其进行篡改，因此 Eve 需要积累对于水印图像的知识，并对 Alice 和 Bob 之间使用的密钥进行估计，从而实现对该隐秘通信的最终破解。

从 Alice 和 Bob 的角度来看，以上“主动看守者”情况下的问题就是一个水印安全的问题，关系到通信中使用的水印算法能否抵抗 Eve 的攻击，能否在 Eve

的计算能力内被破解。

1.6.2 针对水印算法攻击的层次

数字水印算法可能受到多种层次的攻击。文献[18]将水印算法可能受到的攻击分为三个层次：盲的水印移除；基于密钥估计的攻击；篡改攻击。

盲的水印移除是指攻击者并不掌握水印密钥的信息，但试图去除或修改媒体中的水印信号。例如攻击者可以利用对媒体的压缩、加噪、滤波、进行几何变换等方法来降低合法检测者对水印信息的检测率，增加检测的误码率，从而实现对水印信息的移除。事实上，这类攻击就是对水印鲁棒性的攻击。

基于密钥估计的攻击是指攻击者根据对水印算法的所有知识，利用掌握的嵌入有水印的观测信号，利用统计的方法试图估计出水印算法中使用的密钥，并基于估计出的密钥来实现对水印系统的完全破解。根据密码学领域的 Kerckhoffs[29] 原理，任何水印算法都不会永久保密。因此密钥就成为保证数字水印安全的唯一因素。实际应用中，水印算法的密钥被分配给每个合法用户，每个合法用户使用自己的密钥进行水印的嵌入和检测，由每位合法用户生成的含水印图像将包含相同密钥的信息。正如 1.5 节中介绍，在水印嵌入方，密钥Θ经过一个函数映射为嵌入参数$f(\Theta)$，然后利用$f(\Theta)$实现嵌入。由于映射函数$f(\cdot)$通常为一个非可逆的函数，攻击者将只能得到对$f(\Theta)$的估计，而不能得到Θ本身。但是，对于攻击者而言，一旦获得某合法用户的所使用的$f(\Theta)$，攻击者就可以如同合法用户一样利用$f(\Theta)$任意地嵌入、移除和修改水印信息。因此，对$f(\Theta)$的估计等价于对水印密钥的攻击。

篡改攻击是指攻击者使用“逆向工程”的方法，试图由$f(\Theta)$估计出水印算法的密钥Θ。例如通过“逆向工程”对 DVD 硬件播放器密钥的破解。

由于对水印鲁棒性攻击的目的是增加水印检测的误检测率，而对于水印密钥的攻击则是为了掌握水印嵌入和检测的密钥，对水印密钥的攻击成功意味着对水印算法的完全破解。因此，水印算法的安全性就是水印密钥Θ的安全性，等价地，就是水印密钥经映射后的秘密参数$f(\Theta)$的安全性。

1.6.3 水印安全等级

基于 Kerckhoffs[29]原理，文献[27]对于水印的安全性进行了分类，分为不安全（Insecurity）、密钥安全（Key security）、子空间安全（Subspace security）和隐写安全（Stego security）四个等级，如图 1-5 所示[27]。

为了描述的方便，记 $p(\mathbf{X})$ 为 N_o 个载体信号的联合概率密度函数；$\mathbf{K}$ 为可能使用的密钥集合；$p(\mathbf{Y})$ 为 N_o 次观察水印图像的联合概率密度函数，其中每次观察的水印图像对应于不同的密钥；$p(\mathbf{Y}_\mathbf{K})$ 为 N_o 次观察水印图像的联合概率密度函数，其中每次观察的水印图像对应于相同且未知的密钥；$p(\mathbf{Y}|\mathbf{K}_i)$ 为已知密钥 $\mathbf{K}_i$ 条件下，对应于 N_o 次观察水印图像的联合概率密度函数，其中每次观察的水印图像对应的密钥相同。

不安全的水印算法被定义为：密钥集合中存在密钥 $\mathbf{K}_1$，有 $p(\mathbf{Y}|\mathbf{K}_1)=p(\mathbf{Y}_\mathbf{K})$，并且密钥集合中的其它密钥 $\mathbf{K}_i\,(i\neq 1)$ 生成的水印图像的联合密度模型都有 $p(\mathbf{Y}|\mathbf{K}_i)\neq p(\mathbf{Y}_\mathbf{K})$。以上定义意味着使用某种密钥生成的水印图像具有独特的统计分布，而使用其它密钥得到的水印图像均具有与此不同统计分布。因此对于这类水印算法，可以利用最大似然算法来估计可能使用的密钥。

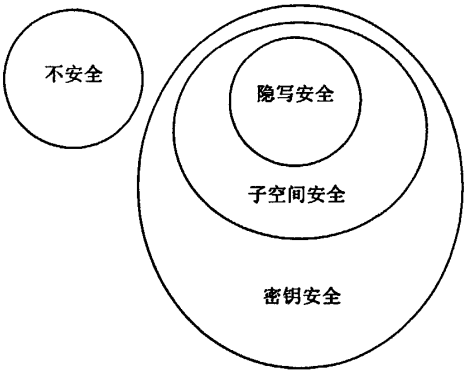


图 1-5 水印安全等级

Fig. 1-5 Embedding security classes

密钥安全的水印算法被定义为：存在密钥集中的一个子集，由该子集中的密钥元素生成的水印图像具有相同的概率密度函数，即 $p(\mathbf{Y}|\mathbf{K}_i)=p(\mathbf{Y}_\mathbf{K})$，$\forall\mathbf{K}_i\in S_\mathbf{K}$，

$S_{\mathrm{K}} \subset \mathbf{K}$。该定义表明，由$S_{\mathrm{K}}$子集内和子集外的元素（密钥）生成的水印图像会在统计特性上有所区别，而对于S_{K}子集内的密钥则不能根据水印图像的统计特性进行区分。因此，这种情况下，对密钥的估计只能是对密钥子集S_{K}的估计，并不能确定具体是S_{K}中的哪一个密钥。

子空间安全的水印是指由水印密钥集合中的任意密钥元素生成的水印图像都具有相同的概率密度函数，即有$\forall \mathrm{K}_i \in \mathrm{K}, p(\mathrm{Y}|\mathrm{K}_i) = p(\mathrm{Y}_{\mathrm{K}})$。子空间安全的概念可以看作是对密钥安全定义的延伸，此时不能利用水印图像统计特性予以区分的密钥集合扩展到整个密钥集合。因此对密钥的估计只能是对于密钥子空间的估计，而不能确定具体是密钥集合中的哪一个密钥。

隐写安全的水印算法是指，不论使用密钥集合中的哪一个密钥来实现水印嵌入，得到的水印图像都与原来的图像载体具有相同的联合概率密度函数，即$\forall \mathrm{K}_i \in \mathrm{K}, p(\mathrm{Y}|\mathrm{K}_i) = p(\mathrm{X})$。在这种情况下，由于水印图像和原始载体图像在统计特征上没有区别，将不能区分图像中是否含有水印信息。因此，隐写安全是水印安全的最高等级。

1.6.4 数字水印安全性研究的目的

由于对水印密钥的正确估计意味着对水印系统的完全破解，水印密钥的安全就成为保证水印算法安全的最终条件。水印密钥的安全就意味着水印的安全。

有关数字水印安全性的研究，就是基于 Kerckhoffs 原理，利用统计和信息论的方法，研究数字水印算法中密钥的安全程度。通过分析水印密钥安全与嵌入方法、嵌入参数、密钥长度、密钥分布特点和观测次数等因素之间的关系，对水印算法的安全性进行定性和定量的度量，为进一步设计高安全、大容量的新一代鲁棒水印提供理论依据。

1.6.5 水印安全性与隐写术的安全

隐写和隐写分析是信息隐藏技术研究的一大分支。由于隐写术也涉及到是否安全的问题，因此需要将隐写术的安全性进行介绍并将其与水印安全进行比较。

“囚徒问题”中，如果 Eva 作为被动看守者实施检测，即只是根据 Alice 发给 Bob 的图像来判断该图像是否含有秘密信息，仅阻拦含有隐秘信息的图像通信，而不试图破解隐秘通信系统，这就是一个隐写分析的问题。与隐写分析相对应的技术被称为隐写术（Stegnography）。隐写术在表面无害的载体信号中嵌入秘密信息，并努力不使检测者发现秘密通信的迹象[30]。因此，隐写术被称为“隐蔽通信的科学”[31]。隐写分析可以被看作是对隐写术的攻击，如果能够通过隐写分析算法正确区分图像是否经过隐写，该隐写分析就是成功的，而对应的隐写术就是不安全的；反之则隐写分析失败，而对应的隐写术是安全的。隐写术和隐写分析是信息安全领域的重要研究方向。

总得来说，一切数字水印的嵌入方法都可以用于隐写。根据应用的需要，隐写术具有不同的要求。早期隐写术的研究主要追求算法具有大容量和不可见性，而较少考虑隐写术在攻击下的性能。近年来，有关隐写术的研究更加侧重考虑算法是否安全。与此同时，隐写分析技术也从只能检测高嵌入率的隐写图像发展到能够检测很小嵌入率的隐写图像，从针对特定隐写方法的分析技术发展到通用的隐写分析。

常用的隐写方法包括基于 LSB 嵌入的隐写方法[32, 33, 34]，基于矩阵编码的隐写方法[35]，针对 JPEG 图像的隐写方法[36]，等等。隐写分析技术可以根据是否是针对特定的隐写技术被分为专用的隐写分析和通用的隐写分析。从具体的分析技术上来看，Westfeld 等人[37]利用 LSB 嵌入前后像素对出现频率的变化，通过 χ^2 检验的方法实现了对基于 LSB 嵌入的隐写分析；Harmsen 等人[38]利用加性高斯噪声模型，通过比较隐写前后图像的累计分布函数的质心位置的变化，实现了一种通用的隐写分析；Sullivan 等人[39]根据自然图像载体之间的相关性，提出利用图像共生矩阵主对角元素中概率最大的点作为特征，利用支持向量机实现对隐写和非隐写图像的分类；文献[40, 41, 42]通过提取图像高阶统计矩和图像预测值的特征，利用支持向量机实现了通用的隐写分析；Xuan 等人[43]从图像共生矩阵中提取高阶统计量，利用非主分量分析的方法完成了通用的隐写分析；Wang[44]和 Moulin 等人[45]则对基于学习的图像隐写分析进行了信息论分析和理论指导。

隐写术的安全与水印的安全是不同的概念。隐写分析与隐写术是一对矛盾。提供给隐写术攻击者（隐写分析检测者）的图像可能是隐写后的，也可能是没有

经过隐写的。安全的隐写术是指已有的隐写分析算法都不能有效地区分图像是否经过隐写；成功的隐写分析可以掌握（专用的隐写分析）或不掌握（通用的隐写分析）隐写算法的具体细节，只需要判断是否图像经过隐写，不需要对隐写算法实施破解。与此不同的是，提供给水印安全攻击者的图像都是嵌入了水印的图像，水印安全的攻击者则掌握了有关水印算法的具体细节，其攻击的目的是估计出水印算法使用的密钥，从而实现对水印算法的完全破解。

1.7 对数字水印安全的攻击

针对水印安全的攻击具体体现为对水印密钥的估计。由于对水印安全的攻击容易与对水印鲁棒性的攻击容易引起混淆，本小节首先对二者间的主要区别予以比较，然后对攻击水印安全的条件进行分类。

第一，对水印鲁棒性攻击的目的是增加信息隐藏信道的误码率；而对水印安全性的攻击则是为了获取水印算法中密钥的信息，并进而实现对于整个水印系统的破解。因此，相比于对水印鲁棒性的攻击，对水印安全性的攻击更为危险。

第二，对于水印鲁棒性的攻击，可以是盲的或非盲的，而对于水印安全的攻击一定是盲的。对水印鲁棒性的攻击者可以不掌握水印算法的有关知识，也可以针对水印算法的一些特点来实施攻击。根据 Kerckhoffs 原理，水印算法不能永久保密，对水印安全的攻击一定是在掌握水印方法的基础上实施的，因此一定是非盲的。

第三，从攻击的有意性方面来说，对水印鲁棒性的攻击可以是无意的或有意的，而对水印安全的攻击一定是有意的。例如，对于水印图像进行常规的压缩和图像处理属于对图像鲁棒性的攻击，这些处理可以为有意或无意。但是对水印安全的攻击是为了获取水印系统的密钥，这种攻击一定是有意的。

第四，水印算法的安全性和鲁棒性之间不存在必然的关系。一个安全程度高的水印算法并不意味着该算法具有高的鲁棒性[46]。

第五，对于水印安全性和水印鲁棒性的衡量指标不同。水印鲁棒性一般可以由检测器输出的信噪比来衡量，而水印安全则需要利用水印密钥和水印图像之间的互信息或者对水印密钥估计时可达的最小误差来衡量。

根据攻击者所掌握信息的不同，对水印安全性的攻击可以分为以下几类。

1.7.1. Known Message Attack (KMA)

KMA 条件下，攻击者掌握 N_o 次观测的水印图像以及对应的秘密信息，并以此估计嵌入水印过程中使用的密钥。这种攻击的条件类似于“共谋攻击”（Collusion Attack）[47]的条件，不过攻击的目的不是估计出嵌入的水印信号，而是水印密钥。KMA 攻击的条件可以具体化为攻击者首先通过合法手段购买到受保护的水印图像，因此能够掌握嵌入的秘密信息。在假设使用相同水印密钥的条件下，该攻击者可以通过这样的方法积累 N_o 次水印图像和嵌入信息对，并以此对使用的密钥予以估计。

1.7.2. Known Original Attach (KOA)

KOA 条件下，攻击者掌握对水印图像的 N_o 次观测以及对应的载体图像（未嵌入水印），并以此估计水印密钥。如果水印嵌入者利用一些常见的或容易获取的图像作为载体，攻击者将能够掌握与水印图像对应的原始载体图像，并在 KOA 条件下实施对水印安全的攻击。

1.7.3. Watermarked Only Attack (WOA)

WOA 是对水印安全性最为一般的攻击条件。此时攻击者仅拥有对水印图像的 N_o 次观测，并据此来估计水印密钥。WOA 条件就是“囚徒问题”中所描述的条件，因为看守者不能获取囚徒使用的原始图像载体或嵌入的秘密信息，只能得到隐藏了秘密信息的水印图像。相比之下，WOA 攻击是水印系统面临的最为普遍的攻击条件，但由于攻击者掌握的信息最少，这种攻击条件对于水印系统的威胁最小；而 KMA 攻击中，攻击者拥有更加充足的信息。尽管具有更强的约束条件，KMA 攻击对于水印安全的威胁最为严重。

1.8 数字水印安全性研究的回顾

在数字水印技术发展的早期，研究的重点主要集中在水印的鲁棒性、水印容

量和视觉质量等性能上。有关水印安全性研究的出现则得益于对水印鲁棒性的讨论[48]。Cox 等人[49]在研究对水印的统计平均攻击时提出了有关水印安全的问题，认为数字水印不仅需要抵抗常规的信号处理的攻击，还需要能够抵抗水印移除的攻击。

Cachin[50]在对隐写术的理论模型进行研究时，将隐写术的安全归结为假设检验的问题，并提出利用 K-L 散度（相对熵）来定义隐写术的安全性。如果记 P_C 为载体图像的分布函数，记 P_S 为隐写图像的分布函数，那么图像在隐写前后分布规律的差别可以由相对熵 $D(P_C \| P_S)$ 来衡量。如果对于 $\varepsilon \geq 0$，有 $D(P_C \| P_S) \leq \varepsilon$，则称该隐写系统是 ε 安全的。当 $\varepsilon = 0$ 时，该隐写系统称为绝对安全。由于实际中 P_C 和 P_S 难以准确计算，基于相对熵的安全性定义的价值更多体现在理论上。

Mitthelholzer 在文献[51]中从信息论的角度研究了隐写术和水印的模型，并利用互信息的概念定义了水印的安全性和鲁棒性。如果记 V 为需要传输的秘密信息，K 为水印密钥，记 X 为嵌入了水印的图像（水印图像），Y 为检测端接收到的受到噪声干扰的水印图像，考虑到攻击者不知道水印密钥，而合法的接收者拥有密钥，Mitthelholzer 利用秘密信息和水印图像之间的互信息 $I(V;X)$ 来衡量水印系统的安全性，用已知密钥 K 条件下秘密信息 V 和经过噪声干扰的水印图像 Y 之间的互信息 $I(V;Y|K)$ 衡量水印系统的鲁棒性。Mitthelholzer 提出水印设计的目的就是最小化 $I(V;X)$ 同时最大化 $I(V;Y|K)$。以上定义中认为对水印安全攻击的目的是获取秘密信息而非水印密钥，当 $I(V;X)=0$ 时攻击者不能从水印图像中获取秘密信息，此时水印系统完全安全。实际上即使攻击者获取了某水印图像中的秘密信息并不意味着破解了该水印系统，以上对于水印安全的定义并不正确。

Kalker 在文献[52]中对比了密码学领域中和水印研究中的安全问题，并试图对水印的鲁棒性和水印安全性概念进行区分，将鲁棒水印定义为“复用到原始内容上的通信信道”，其容量随水印图像的衰减而呈平滑衰减，同时将水印安全定义为“防止非授权用户移除、检测、估计和修改水印信息的能力”。但是，以上的工作未能对水印的鲁棒性和安全性给出足够明确的区分。

Furon 在文献[53]中对水印鲁棒性和安全性做了进一步区别，认为鲁棒性针

对的是盲攻击。对鲁棒性的攻击者并不知道水印的具体算法，只是利用常规的处理手段（例如滤波、压缩、几何变换等）试图去除水印。而且对鲁棒性的攻击必须通过检测器来检验。衡量鲁棒性的手段是检测器对水印的误检率[29]。Furon 在文献[53]中引入了密码学领域中的 Kerckhoffs 原理，认为水印安全性面临的是恶意的攻击。攻击者首先积累并掌握对于水印算法的知识，然后利用获取的知识对水印进行非法移除或非法嵌入。不同于对水印鲁棒性的衡量手段，水印安全性需要用水印密钥在水印图像中的信息泄漏来衡量。相比之下，对鲁棒性的成功攻击只是对水印的部分破解，而对安全性的成功攻击则意味着对水印算法的完全破解。因此，水印安全性比水印鲁棒性具有更广泛的研究内容。

Cayre 等人在文献[48]中的研究是水印安全领域中具有里程碑式意义的工作。在文献[53]的基础上 Cayre 等人明确指出攻击的有意或无意并不是区分鲁棒性攻击和安全性攻击的必要条件，同时进一步完善了 Kalker 关于水印鲁棒性的定义，将造成水印图像衰弱的原因定义为“典型的内容处理方法”。文献[48]还依据攻击者所掌握信息的不同，将对水印安全的攻击分为 Watermarked Only Attack （WOA），Known Message Attack （KMA）和 Known Original Attack （KOA）三类。这种分类后来一直被水印安全的研究者遵循。除了对水印安全性概念的明确定义之外，Cayre 等人还根据信号估计理论，对加性扩频水印的安全性能进行了研究。分析中假定图像载体为独立同分布（i.i.d.）的高斯矢量，根据加性扩频水印的嵌入模型，可以得到基于观察 N_o 次水印图像的似然函数，并构造出关于估计密钥的 Fisher 信息矩阵（Fisher Information Matrix，FIM）[54]。根据 Cramer-Rao 定理[55, 56]，通过求 FIM^{-1} 的迹，得到了无偏估计器对密钥估计可达的误差下界。分析表明，对于加性扩频水印，大的嵌入失真会导致更多的关于密钥的信息泄露，而有关密钥的信息泄露与观测次数呈线性关系。文献[48]还利用最大似然算法[57]和独立分量分析（Independent Component Analysis, ICA）[58, 59]的算法对加性扩频水印进行了攻击实验，并证明攻击者只能在 KMA 情况下实施对扩频序列的准确估计；在 KOA 和 WOA 情况下，由于扩频载波的符号不确定性，攻击者的估计不可能完全准确。文献[48]认为可以利用 Shannon 信息论[60]和基于 Fisher 信息的方法来研究水印密钥在通信过程中的信息泄露，但是由于连续随机变量的熵或条件熵实际上是相对熵的概念[60]，因此不能度量实际的信息量

[48]，所以基于 Fisher 信息的方法更加适用于水印安全性分析。实际上在引入微分熵和互信息的概念后，互信息反映了连续随机变量的微分熵之差，因此可以用来对信息量的变化即信息泄露进行度量[18,60]。基于 Fisher 信息的方法具有统计学和估计理论中的实际意义，与 Shannon 信息论的分析具有定性的关系。因为对水印密钥的无偏估计中可达的误差下界越高，就说明在掌握了 N_o 次水印图像之后，水印密钥的不确定性越大，因此水印密钥的剩余熵也越大，反之亦然。在实际的分析中，基于 Fisher 信息的分析方法也具有很多局限。主要有以下几方面：第一，基于 Fisher 信息的方法需要能够解析地得到基于观测的似然函数，但实际的分析中，由于嵌入算法的复杂性和多样性，多数水印算法难以得到解析的似然函数。第二，实际分析中得到的 Fisher 信息矩阵（FIM）的维数为 $(N_c \times N_v) \times (N_c \times N_v)$，其中 N_c 为水印中使用的秘密载波的数目，N_v 为每条秘密载波的维数。由于为了提高水印的安全性和鲁棒性，实际的扩频水印算法中常使用很长的密钥，例如 512 维或 1024 维，随着密钥长度的增加，在构造 FIM 进行分析时会遇到“维数灾难”。第三，为了求得对密钥无偏估计的误差下界，需要对 FIM 求逆，实际分析中构造的 FIM 很可能是奇异的，因此难以保证能够求出 FIM 的逆。由于以上限制，基于 Fisher 信息的分析方法难以对采用“有信”（Side-Informed）嵌入的水印算法的安全性进行分析。

Comesaña 等人在文献[46]中明确指出，对于水印安全的攻击就是为了获取水印系统的密钥，在评价水印系统安全性时必须假设水印信道是未受干扰的。文献[46]从攻击是否有意、攻击是否非盲、攻击的目的和衡量手段的角度对水印的鲁棒性和安全性进行了比较。同时，利用 Shannon 信息论[60]的方法，对加性扩频水印安全性进行了研究。文中引入微分熵来描述连续随机变量（矢量）的不确定性，利用已知水印图像条件下水印密钥的剩余熵作为衡量水印安全的指标，用水印密钥和 N_o 次观测得到的水印图像之间的互信息来度量水印密钥在通信中的信息泄漏，得出了加性扩频水印安全与其影响因素之间的关系。

由于扩频水印的安全性能与载体信号的统计特性有关[46,48]，而实际的水印应用中多使用自然图像作为载体信号，因此 Ni[61]等人的工作结合自然图像载体分布的非高斯性，利用 Shannon 信息论分析了加性扩频水印在自然图像载体条件下

的安全性，并且利用基于变分贝叶斯方法的独立分量分析(VB-ICA)[62]实现了对水印安全的攻击。

Pérez-Freire等人在文献[63]和[64]中利用基于Shannon信息论和集合估计[66]的方法对DC-DM QIM[10,16]的安全性进行了讨论。DC-DM QIM水印方法中的密钥就是用于产生量化码本的随机抖动量。Pérez-Freire 等人的工作假定图像载体为独立同分布的高斯矢量，在低嵌入失真的条件下，可以假设量化间隔远小于载体分布方差，从而认为落入每个量化区间中的载体系数和自噪声近似呈均匀分布。由于量化器具有相同的量化间隔，在量化器嵌入时的“求模”运算下，水印信号相对于载体信号和抖动密钥具有周期性，其周期等于量化间隔。根据对观察量进行以量化间隔为模的“模简化”，DC-DM QIM 水印系统的安全性就决定于水印密钥在“模简化”之后的统计分布。为了在量化器死区（Voronoi Region）[66]的量化值区间$Z(\Lambda)$内获得关于水印密钥最大的剩余熵，水印密钥应在$Z(\Lambda)$内服从均匀分布。文献[39]经过讨论得出在以上假设条件下，DC-DM QIM 水印安全性和鲁棒性可以同时达到最优，但是安全性和可达的嵌入率之间存在折中。

在文献[67]中，Pérez-Freire 等人假设图像载体为 i.i.d.的高斯分布，用 Shannon 信息论的方法总结了加性扩频水印、衰减的扩频水印的安全性，并对改进的扩频水印安全性进行了分析。

Cayre 和 Bas[27]基于 Kerckhoffs 原理，将水印的安全性分类为：不安全、密钥安全、子空间安全和隐写安全四类。根据隐写安全的特点，Cayre 等人构造出了具有隐写安全等级的自然水印（Natural Watermarking）。自然水印基于扩频水印系统中的相关检测器的特点，利用 Householder 变换[68]实现秘密信息的调制。其核心思想是：根据中心极限定律[69]，具有对称分布的随机矢量与呈高斯分布的秘密载波之间的自相关函数为零均值的高斯分布。嵌入 1 比特秘密信息时，根据秘密信息的具体值来改变或保持载体与秘密载波自相关的符号而保持其绝对值不变。由于零均值高斯分布函数曲线上的点或保持不变或仅改变符号，水印图像矢量与秘密载波之间的自相关函数仍然是零均值的高斯分布，而且方差与嵌入秘密信息之前完全相同。因此，图像在嵌入前后与秘密载波之间的自相关函数分布完全相同，从而得到具有隐写安全等级的水印系统。由于过于追求安全，自然水印在鲁棒性方面逊于传统的扩频水印，而且在嵌入强度因子不为 1 时将失去隐写

安全的特性，所以自然水印更多地具备理论上的意义[27]。作为对自然水印的改进和补充，文献[27]利用对随机矢量归一化的方法得到了更加实用的圆环水印（Circular Watermarking），可以达到密钥安全的等级。

以上的工作为水印安全性的研究给出了明确的定义和研究方向，并针对具体的水印算法安全性进行了分析。这些工作利用统计信号处理或 Shannon 信息论的方法给出了影响水印安全的因素，以及它们与水印安全性能之间的定量关系，为进一步设计更加安全的水印算法提供了理论依据。这些工作使得水印安全性逐步成为区别于水印容量、水印不可感知性和水印鲁棒性的第四个水印基本性能指标。

1.9 本文的工作

基于扩频序列的水印方法在扩频通信的理论框架内实现了对鲁棒性、不可感知性和水印容量的折中和优化，在信息隐藏领域中得到了广泛应用，其安全性也得到了广泛重视。基于扩频序列的水印根据水印的生成是否与载体信号有关可以分为传统的扩频水印（也称为加性扩频水印，Additive Spread-Spectrum Watermarking, Add-SS 水印）和改进的扩频水印（Improved Spread-Spectrum Watermarking，ISS 水印）。前者的水印信号由受嵌入信息调制的秘密载波构成，而后者水印信号的生成与载体有关，还包括对载体信号在秘密载波方向的衰减。由于水印的嵌入模型不同，导致了两种扩频水印在鲁棒性、安全性方面具有差异。

本文的工作是利用自然图像模型对水印图像载体统计分布进行描述，通过统计和信息论的方法对传统的扩频水印和改进的扩频水印安全性进行理论分析。在本文的工作中，依照 Kerckhoffs 原理和水印安全性分析的要求，假设对水印安全的攻击者掌握水印算法的所有细节，水印通信仅依靠密钥来保证通信安全。假设水印嵌入者在各次通信中使用相同的密钥。水印安全的攻击者观察到的水印图像不包含噪声。相比于以往的工作，本文的创新点主要在于：

（1） 利用自然图像模型描述水印载体系数的分布，得到了更加准确和贴近实际的扩频水印安全性能。

由于自然图像便于获取，人们往往采用自然图像作为水印载体。在扩频水印算法模型中，自然图像载体可以看作是对传输的秘密信息的强噪声干扰，因此对

自然图像载体分布的准确描述对于研究水印安全性具有重大意义。以往关于水印安全性的研究中，为了分析方便和直接利用信息论研究的结论，多是假设图像载体为高斯分布。众所周知，常用于作为图像水印载体的自然图像子带小波系数具有很强的非高斯性。图像分析领域中对自然图像的分布特征进行了深入研究，发现自然图像的小波系数呈现很强的高尖峰和重脱尾的现象，用高斯模型已经很难准确描述自然图像小波系数的分布特点。由于高斯分布在协方差相同条件下具有最高的不确定性，而利用自然图像模型来描述图像载体将能够在图像分布空间中减小描述的不确定度，从而更加准确地对自然图像分布的非高斯特点进行刻画，据此对水印系统的安全性分析也将更加准确和符合实际。

（2） 利用高斯尺度混合（Gaussian Scale Mixture，GSM）模型[70,71]描述自然图像载体分布，得到了基于扩频序列的水印安全的解析表达式。

有多种概率模型可以用于准确描述自然图像小波系数的统计分布，例如一般高斯模型、混合高斯模型、高斯尺度混合模型，等等。由于数学表示方法各异，使得多数模型难以达到在描述准确性和数学推导可操作性二方面的统一。GSM模型中，可以通过学习来获得模型尺度因子的知识，在尺度因子已知条件下自然图像小波系数是服从i.i.d.的高斯随机变量。因此，利用GSM模型既可以实现对自然图像载体系数分布特性的准确刻画，又具有数学上的可推导性，使得在对基于扩频序列的水印安全性分析中可以得到封闭的表达形式。

（3） 得到了扩频水印安全性与影响因素之间的关系。

本文利用统计和信息论的方法，结合自然图像模型，对传统的加性扩频水印和改进的扩频水印安全性进行了研究，找到了影响水印安全性的相关因素。由于这些因素也同时对水印的其它性能（鲁棒性、不可见性）产生影响，可以进一步研究扩频水印的安全性与其它性能之间的优化问题，为设计大容量、强鲁棒、高安全的新一代水印提供了依据。

1.10 文章结构安排

本文在第一章对数字水印及其安全性的相关概念进行了综述。水印密钥是保证水印安全的最终手段，水印安全性衡量了水印系统抵抗针对密钥攻击的性能。水印安全性与水印的鲁棒性、不可见性和水印容量是水印技术的四个重要的约束

条件。针对水印鲁棒性的攻击是为了增加水印检测的误码率，而针对水印安全性的攻击则是为了获得水印系统的密钥。与水印鲁棒性相比，水印的安全性能更为重要。

本文的第二章在回顾了自然图像模型的发展历程之后，介绍一般高斯模型、高斯混合模型、隐马尔可夫模型和高斯尺度混合模型等自然图像统计模型的特点，并对高斯尺度混合模型描述自然图像小波系数边际分布的性能进行了实际的仿真。由于高斯尺度混合模型既能有效描述自然图像子带小波系数的统计分布特征，在数学表示上又易于操作，因此适合应用于有关自然图像的信息论分析。

在本文的第三章，介绍了水印安全性分析的具体方法，包括基于 Fisher 信息的方法和基于 Shannon 信息论的方法。基于 Fisher 信息的方法从信号估计的理论出发，利用对水印密钥无偏估计可达的最小误差作为衡量水印安全性的指标。基于 Shannon 信息论的方法使用水印图像和水印密钥之间的互信息来衡量水印安全性。这两种方法都可以用于进行水印安全性分析。

本文的第四章使用 GSM 模型描述自然图像载体的统计分布特点，利用基于 Fisher 信息的方法对加性扩频水印的安全性进行了理论分析和仿真。加性扩频水印算法是最为传统的扩频水印技术，算法中水印的生成和嵌入独立于图像载体。图像载体可以视为水印通信过程中的强噪声。本章从 KMA 和 WOA 两种情况下对加性扩频水印安全性进行了讨论，并从理论和实验仿真的角度将分析结论与文献[48]中的结果进行了比较。

在第五章，本文利用基于 Shannon 信息论的方法对 ISS 水印的安全性进行了理论分析。与传统的加性扩频水印不同，ISS 水印算法中水印的嵌入与图像载体有关，因此是一种利用边信息的有信（Side-Informed）嵌入。传统的加性扩频水印可以看作为 ISS 水印的一种特例。在利用自然图像模型的基础上，第五章的分析从 KMA 和 WOA 两种情况展开，并将本章结论与文献[67]的分析结论进行了比较。在章节的最后，以自然图像为载体进行了仿真并得出结论。

本文的第六章对基于自然图像模型的扩频水印安全性分析的相关工作进行了总结，并对今后的研究方向进行了展望。

第二章 自然图像统计模型

随着互联网的普及和数码相机、数码摄像机等多媒体设备逐渐进入千家万户，人们可以更加方便地拍摄、获取和传输自然图像。自然图像也广泛地被用作图像数字水印的主要载体。这里的自然图像指的是使用照相机、摄像机等设备拍摄的图像，而不是人工的、利用计算机等技术生成的图像（例如绘画作品或计算机卡通等）。

人眼可以很容易地区分拍摄的自然图像和人工生成的图像。这是由于自然图像包含着独特的结构[72]。事实上，人们所能见到的所有图像构成了很大的图像空间，而自然图像只是包含于其中的一个子集。研究自然图像的统计模型，对于充分了解自然图像、准确获取自然图像的统计特性，进而实现对自然图像的处理和利用具有重要的意义。

由于自然图像千差万别，对自然图像统计模型的研究应该既能够反映出所有自然图像共有的最为基本的模型特征，又能够利用一些参数来反映具体自然图像的特殊性质[73]。

在对自然图像模型的研究中，有两个假设被广泛应用[74]。一是马尔可夫性，即自然图像中的每个象素，在邻域中其它象素确定的条件下，其象素值只与邻域中的象素有关，而与该邻域以外的象素无关。二是空间齐次性，即自然图像的某个象素邻域中的概率密度函数与其在图像中的具体位置无关，对自然图像的尺寸进行放大和缩小不改变图像的概率结构。

2.1 自然图像模型研究的回顾

关于自然图像模型的研究开始于二十世纪五十年代，当时研究的主要目的是为了在电视系统中更好地表示和传输光学信号[73]。研究发现自然图像中相邻的象素灰度之间具有很强的相关性，而且这种相关性随着象素间的距离增加而迅速减小。由于多种概率密度函数都能够体现出这种相关性。为了在图像平均功率受限的条件下体现出最大的不确定性[75]，研究人员采用多维高斯分布来描述自然图像象素值的分布，即假设自然图像象素矢量 $\vec{x}$ 服从零均值的高斯分布

$P(\vec{x}) \propto \exp\left(-\vec{x}^T C_x^{-1}\vec{x}/2\right)$，其中$C_x = E\left(\vec{x}\vec{x}^T\right)$。由于$C_x$为对称阵，可以进行特征值分解，得到$C_x = UDU^T$，以及$U^T C_x U = D$，其中$U$为包含$C_x$的特征向量的正交矩阵，$D$为对角阵（其主对角元素为特征向量对应的特征值）。当$\vec{x}$的分布平稳时，傅立叶变换可以作为对角化矩阵U，此时构造描述自然图像的高斯模型，只需要确定D，即自然图像在傅立叶域频率成分的方差[73]。可以通过估计功率谱的方法得到D的值[76]。这种基于象素高斯分布的模型具有描述简单，数学处理方便的特点，但是对于自然图像描述的效果却不好。其原因是傅立叶变换很难捕捉到图像的相位信息[73]。

二十世纪八十年代开始，人们开始使用小波对图像进行多分辨率分析。对于图像进行小波分析的实质是利用一系列不同尺度的带通滤波器对图像进行滤波。人们发现，在对自然图像进行小波变换之后，得到的小波系数呈现出很强的非高斯性。与高斯分布特点不同，自然图像小波分解后的子带系数的边际分布在零点处具有很高的尖峰，而且比高斯分布有更重的脱尾现象。小波系数分布中零点处的尖峰来自于自然图像中象素值变化不大的区域，即“平坦”的区域；而重脱尾现象来源于自然图像中的一些“边缘”特征。文献[77]和[78]使用一般高斯（Generalized Gaussian）模型来描述图像小波系数的边缘分布，即：

$$p_c(c;s,p) = \frac{1}{Z(s,p)}\exp\left(-\left|\frac{c-\mu}{s}\right|^p\right) \tag{2-1}$$

其中$Z(s,p) = 2\frac{s}{p}\Gamma\left(\frac{1}{p}\right)$，$\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}dt$（Gamma 函数）。可以通过图像来估计具体的分布参数(s,p,μ)。与空间域的高斯模型相比，利用小波域的一般高斯模型可以更有效地描述自然图像的统计分布，在图像压缩和图像去噪等领域具有更好的应用效果。

小波域的一般高斯模型是对小波系数的边际分布进行建模，这种方法是建立在图像的子带小波系数之间相互独立的假设之上。如果该假设成立，子带内小波系数的联合概率密度函数可以简化为各个系数边际概率密度函数的乘积。尽管小波变换具有很强的去相关性，但是这种“去相关”只是在协方差意义上的二阶相关性，小波变换并不能去除更高阶的相关，因此小波系数之间并不独立。实际上，

考察对自然图像进行小波分解的结果，可以发现子带内大幅值的小波系数往往呈“簇”出现，而且一个“大”的小波系数的“父节点”和“子节点”也都会为“大”系数。因此，自然图像的小波系数的方差具有局部性，而不是传统的方法中那样认为图像小波系数的方差都相同。

近年来，文献[74]提出使用小波域高斯尺度混合（Gaussian Scale Mixture, GSM）模型作为自然图像的统计模型。GSM 模型用一个高斯随机场和一个尺度随机场的乘积来描述自然图像小波系数的统计特性。与以往的其它模型相比，GSM 模型既可以有效地刻画自然图像小波系数的统计分布，又具有计算简单、便于数学处理的特点。因此，GSM 模型在图像处理、图像分析和计算机视觉等领域得到了广泛的应用[79, 80]。

2.2 一般高斯模型

一般高斯模型（Generalized Gaussian Model，GGM）描述的随机变量的概率密度函数可以表示为式(2-1)。一般高斯模型的概率密度函数中，参数μ是分布的均值；参数$s>0$称为尺度参数，用于描述随机变量分布的方差；参数$p>0$称为形状参数，用于描述随机变量分布的尖锐程度，其值与分布的尖峰下降速率呈反比例[81]。通过参数(s,p,μ)的取值不同，可以由(2-1)式表示多种著名分布的概率密度函数。例如，当$p=2$时，式(2-1)就是以μ为均值，以$\frac{s^2}{2}$为方差的高斯分布；当$p=1$时，式(2-1)则为拉普拉斯分布（如图 2-1 所示）。

在一般高斯模型中，可以用最大似然算法或矩方法来估计参数(s,p)。估计时一般需要使用数值的方法，而不能用封闭的形式表示。文献[82]通过数值计算实验发现，大约有 98%的自然图像小波子带系数的分布可以用 GGM 模型来描述。对于自然图像，可以认为μ的值为零，典型的p值位于[0.5,1]区间内[83]。

一般高斯模型可以用于描述自然图像的离散余弦变换（Discrete Cosine Transform, DCT）系数或小波变换系数的分布，并被广泛用于视频、图像压缩、图像检索、去噪和数字水印等领域[84]。

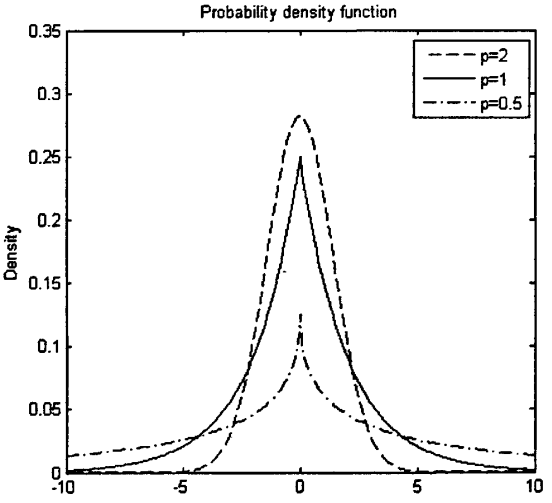


图 2-1 一般高斯模型的概率密度函数

Fig. 2-1 Probability density of Generalized Gaussian Model

2.3 高斯混合模型

高斯混合模型(Gaussian Mixture Model)是一种有效描述自然图像小波域子带系数分布的图像模型。

自然图像经二维小波分解后，在各个子带的小波系数都呈现出高尖峰、重脱尾的分布特性，而且分布的均值近似为 0 。这样的分布特点可以由 GMM 模型予以准确描述。记 x_i 表示在某一子带中处于 i 位置的小波系数，$i=1,2,\cdots,K$，K 为该子带内所有小波系数的数目。则 x_i 的概率密度函数可以由一个多状态的高斯混合模型[85]表示为：

$$f(x_i)=\sum_{m=1}^{M} P_m \cdot g\left(x_i,0,\sigma_m^2\right) \tag{2-2}$$

其中，m 表示该小波系数的状态，假设共有 M 个状态。P_m 表示小波系数处于 m 状态的概率，因此有 $\sum_{m=1}^{M} P_m=1$。在 m 状态下，x_i 将服从零均值的正态分布 $N\left(0,\sigma_m^2\right)$，其中 σ_m^2 为方差，即有

$$g\left(x_i,0,\sigma_m^2\right)=\frac{1}{\sqrt{2\pi\sigma_m^2}}\exp\left(-\frac{x_i^2}{2\sigma_m^2}\right) \tag{2-3}$$

由于在自然图像小波域子带系数的分布中，绝对值大的系数对应于图像中的能量比较大奇异区域，例如边缘，绝对值小的系数反映的是图像能量较小即相对平缓的区域。而且绝对值大的系数较少，而绝对值小的系数很多，因此在描述自然图像小波系数分布时，以上的 GMM 模型常常可以简化为含有两个状态，即系数绝对值为“大”和系数绝对值为“小”的状态。这样，自然图像的 GMM 模型可以简化为

$$f\left(x_i\right)=P_s\bullet g\left(x_i,0,\sigma_s^2\right)+P_l\bullet g\left(x_i,0,\sigma_l^2\right) \tag{2-4}$$

其中 P_s 和 P_l 分别为小波系数状态为“小”和“大”的概率，有 $P_s+P_l=1$；在“小”和“大”状态下，x_i 的概率密度函数仍然服从零均值的高斯分布

$$g\left(x_i,0,\sigma_s^2\right)=\frac{1}{\sqrt{2\pi\sigma_s^2}}\exp\left(-\frac{x_i^2}{2\sigma_s^2}\right) \tag{2-5}$$

$$g\left(x_i,0,\sigma_l^2\right)=\frac{1}{\sqrt{2\pi\sigma_l^2}}\exp\left(-\frac{x_i^2}{2\sigma_l^2}\right) \tag{2-6}$$

σ_s^2 和 σ_l^2 分别为小波系数在“小”和“大”状态下的方差。

因此，在使用 GMM 模型描述自然图像小波系数分布时，需要得到一组模型参数 $\theta=\left[P_s,P_l,\sigma_s^2,\sigma_l^2\right]$。该模型参数可以由自然图像利用 EM（Expectation Maximization）[86,87]算法估计得到。EM 算法由以下几步构成。

第一步：参数初始化，同时令迭代次数为 $n=0$。为模型参数设置一组初始值，该初始值可以任意设定。即

$$\theta(0)=\left[P_s(0),P_l(0),\sigma_s^2(0),\sigma_l^2(0)\right] \tag{2-7}$$

第二步：E-Step。由 Bayes 公式，根据自然图像样本和初始参数计算子带内每个小波系数的状态概率。可由公式(2-8)和公式(2-9)分别计算每个小波系数状态为“小”和“大”的概率。由于 $g\left(x_i,0,\sigma_s^2(n)\right)$ 和 $g\left(x_i,0,\sigma_s^2(n)\right)$ 的计算需要利用到自然图像样本的信息，E-Step 计算的结果是在已知样本和当前参数集条件下小波系数为“大”和“小”状态的后验概率。

$$P_{s,i}=\frac{P_s(n)\cdot g\left(x_i,0,\sigma_s^2(n)\right)}{P_s(n)\cdot g\left(x_i,0,\sigma_s^2(n)\right)+P_l(n)\cdot g\left(x_i,0,\sigma_l^2(n)\right)} \tag{2-8}$$

$$P_{l,i}=\frac{P_l(n)\cdot g\left(x_i,0,\sigma_l^2(n)\right)}{P_s(n)\cdot g\left(x_i,0,\sigma_s^2(n)\right)+P_l(n)\cdot g\left(x_i,0,\sigma_l^2(n)\right)} \tag{2-9}$$

$$i=1,2,\cdots,K$$

第三步：M-Step。根据第二步的结果，对模型参数进行更新，得到参数集 $\theta(n+1)=\left[P_s(n+1),P_l(n+1),\sigma_s^2(n+1),\sigma_l^2(n+1)\right]$ 中的各参数为

$$P_s(n+1)=\frac{1}{K}\sum_{i=1}^{K}P_{s,i} \tag{2-10}$$

$$P_l(n+1)=\frac{1}{K}\sum_{i=1}^{K}P_{l,i} \tag{2-11}$$

$$\sigma_s^2(n+1)=\frac{1}{K\cdot P_s(n+1)}\sum_{i=1}^{K}x_i^2\cdot P_{s,i} \tag{2-12}$$

$$\sigma_l^2(n+1)=\frac{1}{K\cdot P_l(n+1)}\sum_{i=1}^{K}x_i^2\cdot P_{l,i} \tag{2-13}$$

第四步：令迭代次数加 1，即 $n=n+1$。判断以上计算的模型参数是否收敛。若收敛，结束训练；否则返回 E-Step 进行下一次迭代。

GMM 模型能够很好地描述自然图像小波系数的统计分布，在图像去噪、图像增强、脆弱水印等领域得到广泛应用[82, 85]。

2.4 隐马尔可夫模型

由于小波变换对于实信号不能实现完全的去相关[89]，自然图像小波变换系数在尺度之间仍然存在一定的相关性。主要表现在：

聚集性（Clustering）：如果某处的小波系数大（或小），则它临近的小波系数值很可能为大（或小）；

持续性（Persistence）：小波系数为大（或小）的状态很可能会在不同尺度之间传递。

根据小波变换的以上性质，在 GMM 模型的基础上，文献[90]提出使用小波

域隐马尔可夫树（DWT-HMT）模型来描述图像小波系数的概率分布特点。在DWT-HMT模型中，不同尺度下相关位置处小波系数的隐状态之间具有马尔可夫关系，由隐状态决定小波分布的方差，即决定小波系数为“大”或“小”，每个小波系数分布为混合高斯。

记 $x_{j,i}$ 为 j 尺度下、位置为 i 的小波系数，j 越大表示越精细的尺度；记 $S_{j,i}$ 为 $x_{j,i}$ 对应的隐状态。每个小波系数可取 M 种隐状态，每种隐状态出现的概率为 $P\left(S_{j,i}=m\right)=p_{j,i}^{(m)}$，其中 $m=1,2,\cdots,M$。在给定隐状态的条件下，小波系数呈现高斯分布，即 $g\left(x_j,\mu_j^{(m)},\sigma_j^{(m)}\right)$。在描述自然图像小波系数分布时，常采用两种隐状态用 $p_j^{(l)}$ 和 $p_j^{(s)}$，即小波系数方差为“大”和方差为“小”，并假设分布均值为0。因此，j 尺度下小波系数的分布为[91]

$$f\left(x_j\right)=p_j^{(l)}\cdot g\left(x_j,0,\sigma_j^{(l)}\right)+p_j^{(s)}\cdot g\left(x_j,0,\sigma_j^{(s)}\right) \tag{2-14}$$

其中 $p_j^{(l)}+p_j^{(s)}=1$，$g\left(x_j,0,\sigma_j\right)$ 为0均值、方差为 σ_j^2 的高斯分布。

由于相邻尺度对应位置小波系数隐状态之间具有马尔可夫性，记相邻尺度间小波系数父节点和四个子节点隐状态之间的转移概率矩阵为 A_j，则有

$$A_j=\begin{bmatrix} p_j^{s\to s} & p_j^{s\to l} \\ p_j^{l\to s} & p_j^{l\to l} \end{bmatrix} \tag{2-15}$$

其中 $p_j^{m\to m'}$ 表示 j 尺度下小波系数的父节点到该小波系数节点隐状态之间的转移概率。如果假设在 j 尺度下小波系数隐状态的概率为 $\mathbf{p}_j=\left(p_j^{(s)},p_j^{(l)}\right)$，可以得到各个尺度下小波系数隐状态之间的关系为：

$$\mathbf{p}_j=\mathbf{p}_1A_2A_3\cdots A_j,\quad j=2,3,\cdots,J \tag{2-16}$$

其中 J 为小波分解的最精细尺度级。

因此，在DWT-HMM模型中，自然图像小波系数的参数集可以表示为：

$$\theta=\left[\mathbf{p}_1,A_2,\cdots,A_J,\sigma_1^s,\cdots,\sigma_J^s,\sigma_1^l,\cdots,\sigma_J^l\right] \tag{2-17}$$

模型参数 θ 可以根据图像样本通过EM算法估计得到[92]。

2.5 自然图像的高斯尺度混合模型

高斯尺度混合（Gaussian Scale Mixture, GSM）模型是近年来被广泛用于描述自然图像小波系数统计分布的有效模型。与其它的图像统计模型相比，GSM 模型既能准确刻画自然图像小波系数的统计分布特性，同时又具有数学上易于推导的特点。

2.5.1 GSM 模型

在数学表达上，一个随机矢量 $\mathbf{X}$ 如果能够被表示为一个零均值的高斯随机矢量 $\mathbf{U}$ 和一个正的、独立的尺度随机变量 S 的乘积，该随机矢量 $\mathbf{X}$ 就被称为是一个高斯尺度混合[79]，并且可以被表示为

$$\mathbf{X} \overset{d}{=} S \cdot \mathbf{U} \tag{2-18}$$

其中 $\mathbf{U} \sim N(0,\mathbf{Q})$，$\mathbf{Q}$ 为 $\mathbf{U}$ 的协方差矩阵；“$\overset{d}{=}$”表示分布意义上的相等。因此，令 $z = s^2$，一个 GSM 矢量的概率密度函数可以表示为以下积分。

$$p_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{N}{2}} |z\mathbf{Q}|^{\frac{1}{2}}} \exp\left(-\frac{\mathbf{x}^T \mathbf{Q}^{-1} \mathbf{x}}{2z}\right) p(z)\, dz \tag{2-19}$$

其中 N 为随机矢量 $\mathbf{X}$ 的维度。

文献[71]研究了一个随机矢量可以表示为 GSM 的充分必要条件，并且发现 GSM 可以用于描述多种分布类型的随机矢量，例如 Cauchy 分布，一般高斯分布等等。由于 GSM 模型中，在尺度变量 S 确定的条件下，随机矢量 $\mathbf{X}$ 的分布为高斯，因此从公式（2-18）和（2-19）可以看出，GSM 模型相当于是无数个零均值的高斯模型的混合，每个高斯模型的方差仅由尺度随机变量决定。混合后得到随机矢量 $\mathbf{X}$ 的分布则具有关于零点对称和重托尾的现象[74]。

2.5.2 用 GSM 随机场描述自然图像小波系数的分布

自然图像小波系数的分布可以由 GSM 模型进行准确描述[71]。在实际中，对自然图像进行小波分解并将子带小波系数划分为不相重叠的块，每一个块内小波系数的分布可以用 GSM 模型予以描述，而该子带内的所有块的系数则构成 GSM

随机场。若记自然图像在某一子带内的小波系数为 $\mathbf{X}=\{\mathbf{X}_i : i\in I\}$，其中 i 为矢量的位置索引，I 为所有位置索引的集合，则 $\mathbf{X}_i$ 为每一个块内的小波系数。结合 GSM 模型的特点，子带小波系数的分布为

$$\mathbf{X}\overset{d}{=}S\cdot\mathbf{U}=\{S_i\cdot\mathbf{U}_i, i\in I\} \tag{2-20}$$

其中 $\mathbf{U}$ 为一个高斯随机场；$\mathbf{U}_i$ 为高斯随机矢量，其均值为 0，协方差阵为 $\mathbf{Q}$；S 是一个值为正的尺度随机变量，用于控制图像小波系数的方差。S 和 $\mathbf{U}$ 相互独立. 给定 S，$\mathbf{X}$ 的概率密度函数表示为：

$$P_{\mathrm{x}|s}\left(\mathrm{x}\middle|s\right)=\frac{1}{\left(2\pi\right)^{\frac{N}{2}}\left|s^2\mathrm{Q}\right|^{\frac{1}{2}}}\exp\left(-\frac{\mathrm{x}^T\mathrm{Q}^{-1}\mathrm{x}}{2s^2}\right) \tag{2-21}$$

文献[71]对尺度随机变量 S^2 的分布进行了研究，这是因为与尺度随机变量相乘的是零均值的高斯随机矢量，直接影响 $\mathbf{X}$ 分布的参数实际上是 S^2。利用到 Jeffrey's prior[73, 74]的方法，文献[71]得出 S^2 的分布正比于 $\frac{1}{S^2}$。

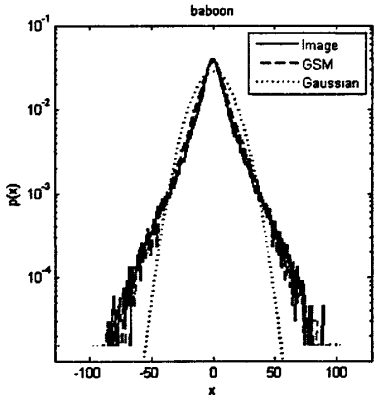
实际中，在掌握自然图像样本的条件下，S 的值可以由最大似然算法估计得到。

$$\begin{aligned}\hat{s} &= \arg\max_{\mathrm{s}}\left\{\log p\left(\mathrm{x}\middle|s\right)\right\}\\ &= \arg\max_{\mathrm{s}}\left\{N\log\left(\mathrm{s}\right)+\frac{\mathrm{x}^TQ^{-1}\mathrm{x}}{2s^2}\right\}\\ &= \sqrt{\frac{\mathrm{x}^TQ^{-1}\mathrm{x}}{N}}\end{aligned} \tag{2-22}$$

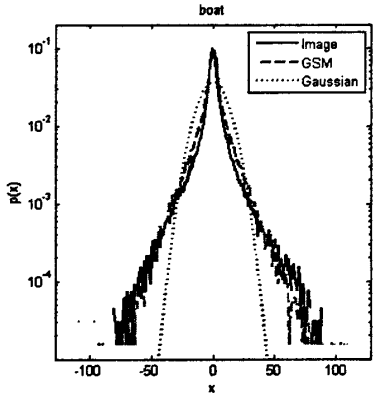
如果假定 $\mathbf{X}_i$ 为标量，模型（2-20）退化为标量 GSM 模型。此时，对于不同的 i 和 j，X_i 与 X_j 相互独立；给定 $S=s_i$，X_i 的分布为高斯，即有 $p_{X_i|s_i}\left(x_i\middle|s_i\right)\sim N\left(0, s_i^2\sigma_u^2\right)$，其中 u 为全局高斯随机场，不失一般性，可以假定 σ_u^2 为单位值。

图 2-2 (a) 和 (b) 是分别用 GSM 模型和高斯模型拟合自然图像 baboon 和 boat 小波系数统计分布的性能比较。其中实线为图像 HL1 子带系数的实际边缘分布，

虚划线和虚点线分别为 GSM 模型和高斯模型对图像实际分布的拟合，可以发现 GSM 模型可以很好地拟合图像小波系数的分布，而传统的高斯模型则难以对自然图像小波系数分布的高尖峰和重脱尾特征进行准确的描述。



(a)



(b)

图 2-2 利用 GSM 模型和高斯模型描述自然图像小波系数边际分布的性能比较

Fig. 2-2 Performance of GSM and Gaussian when modeling the marginal distribution of wavelet coefficients for natural images

由 GSM 模型的分析和实际仿真结果可以看出，GSM 既具备对自然图像子带小波系数分布进行准确描述的能力，又具有数学上的易推导性。这是因为可以通过对图像的训练得到 GSM 模型的尺度参数，在已知尺度参数的条件下，自然图像小波系数呈现高斯分布，因而便于在数学上对其进行分析，特别适合在信息论框架下进行讨论。本文将利用 GSM 模型对基于扩频序列的水印安全性进行讨论。

第三章 水印安全的分析方法

对水印安全进行攻击的目的是获取水印系统中使用的密钥，因此水印密钥的安全等价于水印系统的安全。对于攻击者而言，不安全的水印系统意味着攻击者可以通过对水印图像进行估计来获取水印密钥的知识；而安全的水印系统意味着攻击者很难或者几乎不能通过对水印图像的观察来获取水印密钥的知识。从水印系统本身来说，安全的水印系统在水印图像中泄漏给攻击者关于水印密钥的知识很少；而不安全的水印系统中，有较多的关于水印密钥的知识泄漏给攻击者。

分别对应于以上分析问题的角度，有两种方法可以用于对水印安全的定量分析：基于 Fisher 信息的方法和基于 Shannon 信息的方法。

3.1 基于 Fisher 信息的方法

Fisher 信息是来自于数学统计和信息理论的概念，具体定义为分数（score）的方差[60,90]。Fisher 信息主要用于衡量观察到的随机变量携带的关于未知参数信息的多少。

在统计估计中理论，标准的问题是根据抽自某一分布的样本函数来确定该分布的参数[94, 95]。在实际的参数估计中，由于参数分布情况的不同，有时样本（随机变量）中携带了较多的关于未知参数的信息，对随机变量进行少量观测就可以较为准确地估计出未知参数的值；而有时样本携带的关于未知参数的信息较少，需要大量的样本数据才能实现对未知参数进行较准确估计。因此需要研究观测到的随机变量所携带的关于未知参数的信息有多少。估计理论中定义 $f(X|\theta)$ 为随机变量 X 关于未知参数 θ 的似然函数。似然函数表示当未知参数取某一值时，观测到 X 的可能性的大小。由于对函数取对数的运算具有单调性，因此分析中经常使用对数似然函数 $\ln f(X|\theta)$，其中 $\ln(\cdot)$ 表示取自然对数运算。

假设 $(X_1, X_2, \cdots, X_n)$ 为对随机试验的 n 次观测，θ 是随机变量分布的参数，X_i 则为观测到的随机变量。根据对随机变量的观测，得到对参数 θ 估计值为 $\hat{\theta} = T(X_1, X_2, \cdots, X_n)$。可以有如下定义。

定义 3-1[93]：对参数θ估计的偏定义为估计量误差的期望值，即偏等于$E_\theta\left(\hat{\theta}-\theta\right)$，其中$E_\theta$表示相对于密度函数$f\left(\cdot|\theta\right)$取期望。

定义 3-2[93]：若对估计量误差的期望值为零，则称该估计为无偏估计。即无偏估计量的期望值等于实际的参数值。

定义 3-3[93]：若对所有的θ都有$\left[E_\theta\left(\hat{\theta}_1-\theta\right)\right]^2 \geq \left[E_\theta\left(\hat{\theta}_2-\theta\right)\right]^2$，则称估计量$\hat{\theta}_2$优于估计量$\hat{\theta}_1$。

可见，对于未知参数估计的准确程度可以由估计的方差来衡量。为了研究对未知参数无偏估计的最小方差，可以定义分布$f\left(X|\theta\right)$的分数（score）。

定义 3-4[60]：　分布$f\left(X|\theta\right)$的分数（score）定义为

$$V=\frac{\partial}{\partial\theta}\ln f\left(X|\theta\right)=\frac{\frac{\partial}{\partial\theta}f\left(X|\theta\right)}{f\left(X|\theta\right)}$$

其中$X\sim f\left(X|\theta\right)$。

由定义（3-4）可知，分数是一个随机变量，表示了分布$f\left(X|\theta\right)$随参数θ的变化情况。对分数V取均值，可得

$$\begin{aligned}E(V)&=\int\frac{\frac{\partial}{\partial\theta}f\left(x|\theta\right)}{f\left(x|\theta\right)}f\left(x|\theta\right)dx\\&=\int\frac{\partial}{\partial\theta}f\left(x|\theta\right)dx\\&=\frac{\partial}{\partial\theta}\int f\left(x|\theta\right)dx\\&=\frac{\partial}{\partial\theta}1\\&=0\end{aligned}\qquad(3\text{-}1)$$

即分数的均值为 0 。求分数V的方差，可以得到 Fisher 信息的定义。

定义 3-5[60]：Fisher 信息$J\left(\theta\right)$定义为分数的方差，即

$$J\left(\theta\right)=E\left(\frac{\partial}{\partial\theta}\ln f\left(X|\theta\right)\right)^2$$

在 $\ln f(X|\theta)$ 满足关于 θ 二次可偏导和 $\int\frac{\partial^2}{\partial^2\theta}f(X|\theta)dx=0$ 条件下，Fisher 信息可以等价地表示为

$$J(\theta)=-E\left(\frac{\partial^2}{\partial^2\theta}\ln f(X|\theta)\right) \tag{3-2}$$

根据定义(3-5)，可知 Fisher 信息 $J(\theta)$ 的值非负。一个随机变量携带有高的 Fisher 信息意味着对应似然函数（分布）具有高的分数。从(3-2)式可以看出，Fisher 信息也可以看作是对数似然函数的二阶导数期望的相反数，它衡量了关于未知参数 θ 的对数似然函数曲线的尖锐程度。如果对数似然函数曲线比较尖锐，其二阶导数将有较大的值，对应的 Fisher 信息就高；反之，若对数似然函数曲线较平坦，对应的二阶导数较小，其分布的 Fisher 信息就低。对于 Fisher 信息高的对数似然函数（分布），观测到的随机变量 X 携带的关于未知参数 θ 的信息就多，根据较少的观测就可以得到对未知参数 θ 的较准确估计；对于 Fisher 信息低的对数似然函数（分布），观测到的随机变量 X 携带的关于未知参数 θ 的信息则少，要得到关于未知参数 θ 的较准确估计就需要对随机变量 X 进行更多的观测。

Fisher 信息可以用于衡量对未知参数无偏估计可达的最小误差。考虑一个无偏的估计器 $\hat{\theta}(X)$，根据定义（3-2）有

$$E\left[\hat{\theta}(X)-\theta\right]=\int_{-\infty}^{\infty}\left[\hat{\theta}(X)-\theta\right]\cdot f(X|\theta)dx=0 \tag{3-3}$$

由以上无偏条件对 θ 求偏导可得

$$\begin{aligned}&\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\left[\hat{\theta}(X)-\theta\right]\cdot f(X|\theta)dx\\&=-\int_{-\infty}^{\infty}f(X|\theta)dx+\int_{-\infty}^{\infty}\left[\hat{\theta}(X)-\theta\right]\frac{\partial}{\partial\theta}f(X|\theta)dx\\&=0\end{aligned} \tag{3-4}$$

由于 $\int_{-\infty}^{\infty}f(X|\theta)dx=1$，可知 $\int_{-\infty}^{\infty}\left[\hat{\theta}(X)-\theta\right]\frac{\partial}{\partial\theta}f(X|\theta)dx=1$。考虑到 $\frac{\partial}{\partial\theta}f(X|\theta)=\frac{\partial\ln f(X|\theta)}{\partial\theta}f(X|\theta)$，因此可得[96]

$$\int_{-\infty}^{\infty}\left[\hat{\theta}(X)-\theta\right]\frac{\partial}{\partial\theta}f(X|\theta)dx$$

$$= \int_{-\infty}^{\infty}\left[\hat{\theta}(X)-\theta\right]\frac{\partial \ln f(X|\theta)}{\partial\theta} f(X|\theta)dx$$

$$= \int_{-\infty}^{\infty}\left[\hat{\theta}(X)-\theta\right]\sqrt{f(X|\theta)}\cdot\sqrt{f(X|\theta)}\frac{\partial \ln f(X|\theta)}{\partial\theta}dx$$

$$=1 \tag{3-5}$$

根据 Cauchy-Schwarz 不等式[97]，由（3-5）式可得

$$\int_{-\infty}^{\infty}\left[\hat{\theta}(X)-\theta\right]^2 f(X|\theta)dx\cdot\int_{-\infty}^{\infty} f(X|\theta)\left[\frac{\partial \ln f(X|\theta)}{\partial\theta}\right]^2 dx \geq 1 \tag{3-6}$$

当且仅当$\frac{\partial \ln f(X|\theta)}{\partial\theta}=k(\theta)\left[\hat{\theta}(X)-\theta\right]$时以上不等式取等号。由于 Fisher 信息被定义为$J(\theta)=E\left(\frac{\partial}{\partial\theta}\ln f(X|\theta)\right)^2=\int_{-\infty}^{\infty} f(X|\theta)\left[\frac{\partial \ln f(X|\theta)}{\partial\theta}\right]^2 dx$，而对未知参数$\theta$进行无偏估计的方差可由$\int_{-\infty}^{\infty}\left[\hat{\theta}(X)-\theta\right]^2 f(X|\theta)dx$表示，因此可得 Cramer-Rao 不等式[96]如下：

$$\int_{-\infty}^{\infty}\left[\hat{\theta}(X)-\theta\right]^2 f(X|\theta)dx \geq \frac{1}{\int_{-\infty}^{\infty} f(X|\theta)\left[\frac{\partial \ln f(X|\theta)}{\partial\theta}\right]^2 dx} \tag{3-7}$$

即$E\left[\left(\hat{\theta}(X)-\theta\right)^2\right]\geq\frac{1}{J(\theta)}$。

Cramer-Rao 定理反映了对未知参数进行无偏估计可达的最小方差的界（Cramer-Rao Bound，CRB）。CRB 的值越大说明对未知参数进行无偏估计可达的最小误差越高，因此得到的对参数的无偏估计的准确性就越低；反之，低的 CRB 意味着对未知参数进行无偏估计可达的最小误差低，因此得到的估计值的准确性高。

如果有多个未知参数需要估计，可以记为未知参数矢量$\Theta=(\theta_1,\theta_2,\cdots,\theta_N)$。此时 Fisher 信息则扩展为$N\times N$维的 Fisher 信息矩阵[96]（Fisher Information Matrix）$J(\Theta)$。

$$J(\Theta)=E\left(\left\{\nabla_\Theta \ln p(Y|\Theta)\right\}\left\{\nabla_\Theta \ln p(Y|\Theta)\right\}^T\right) \tag{3-8}$$

其中，$\nabla_{\Theta}(\cdot)$表示对函数求Θ的梯度。Fisher信息矩阵的每个元素为

$$J_{i,j}(\Theta)=E\left[\frac{\partial}{\partial\theta_i}\ln f(X|\Theta)\frac{\partial}{\partial\theta_j}\ln f(X|\Theta)\right] \tag{3-9}$$

此时对每个未知参数的无偏估计的最小误差由Cramer-Rao不等式确定，为

$$Var\left[\hat{\Theta}-\Theta\right]\geq CRB(\Theta)=tr\left(J^{-1}(\Theta)\right) \tag{3-10}$$

其中“$J^{-1}(\Theta)$”表示Fisher信息矩阵$J(\Theta)$的逆，“$tr(\cdot)$”表示求矩阵的迹。

跟据Fisher信息的定义和Cramer-Rao定理，可以利用Fisher信息和CRB从信号估计的理论出发来分析攻击者对水印密钥进行无偏估计所能达到的准确程度，并进而对水印的安全性能予以衡量。根据对含水印图像的观察，攻击者可以对其中包含的水印密钥进行估计。估计值与真实值的误差越小，意味着水印的安全性越低，反之水印安全性则越高。

3.2 基于Shannon信息的方法

基于Shannon信息的方法对扩频水印安全性进行分析，需要利用Shannon信息论的定义来衡量随机变量（矢量）的不确定性以及随机变量在通信中的信息泄露。以下首先介绍Shannon方法中的相关定义，然后介绍利用Shannon信息论研究扩频水印安全的具体方法。

3.2.1 微分熵

微分熵是连续随机变量的信息论特征。如果随机变量X的累积分布函数$F(X)=\Pr(X\leq x)$是连续的，则称该随机变量为连续的。

以$f(x)$为密度函数的连续随机变量X的微分熵定义为[98]

$$h(X)=-E[\log f(x)]=-\int_S f(x)\log f(x)\,dx \tag{3-11}$$

其中S为随机变量X的支撑集，$E(\cdot)$表示求数学期望。由定义可知，微分熵仅决定于连续随机变量的概率密度函数。

与离散随机变量的熵[60]相比，连续随机变量微分熵的定义具有相似的形式，

但是两者的物理意义不同。离散信源的熵表示离散信源输出的信息量；连续随机变量的微分熵是一种“相对熵”，并不是信源的实际熵，在概念上不能作为信息熵来理解[98]。实际上连续随机变量的绝对熵是在微分熵基础上附加了$-\lim_{\Delta\to 0}(\log\Delta)$，其中$\Delta$表示对连续随机变量进行量化的步长。当$\Delta\to 0$时，该附加项趋于无穷大，其物理意义是连续信源可取值的数目是无穷多的，因此其不确定性是无限大，连续随机变量的绝对熵也为无穷大。但是，实际运用中常常讨论的是连续随机变量熵的差值问题，例如计算平均互信系、信道容量等。如果在讨论中使用相同的量化步长，以上附加项可以被抵消掉。因此，尽管连续随机变量的微分熵不能作为信息熵来理解，它并不影响对于平均互信息、信道容量等性能的计算。

根据连续随机变量微分熵的定义，可以进一步得到连续随机变量的联合微分熵和条件微分熵。

定义 3-6[98]：一组连续随机变量$X_1,X_2,\cdots,X_n$的联合微分熵定义为

$$h(X_1,X_2,\cdots,X_n)=-\int f(x_1,x_2,\cdots,x_n)\log f(x_1,x_2,\cdots,x_n)dx_1dx_2\cdots dx_n$$

定义 3-7[98]：如果连续随机变量X、Y的联合密度函数为$f(x,y)$，条件微分熵$h(X|Y)$可以定义为$h(X|Y)=-\int f(x,y)\log f(x|y)dxdy$。当$h(X,Y)$和$h(Y)$不为无穷大时，有$h(X|Y)=h(X,Y)-h(Y)$。

连续随机变量的微分熵具有以下性质[98]：

（1）可加性（微分熵的链式法则）

$$h(X_1,X_2,\cdots,X_n)=\sum_{i=1}^{n}h(X_i|X_1,X_2,\cdots,X_{i-1}) \tag{3-12}$$

（2）上凸性和极值性。即连续随机变量的微分熵是概率密度函数$p(x)$的上凸函数，可以求得微分上的极值。

（3）连续随机变量的微分熵不具有非负性。例如在$[0,0.5]$区间均匀分布的连续随机变量的微分熵为$-\log 2$比特。

由于正态分布的微分熵被广泛用于有关连续随机变量（矢量）的信息论分析

中。根据对微分熵的定义，可以求得多元正态分布的微分熵。

若 $X_1,X_2,\cdots,X_n$ 为服从均值为 μ，协方差矩阵为 Σ 的多元正态分布 $N(\mu,\Sigma)$，则

$$h(X_1,X_2,\cdots,X_n)=h[N(\mu,\Sigma)]=\frac{1}{2}\log\left[(2\pi e)^n|\Sigma|\right](\text{bit}) \tag{3-13}$$

其中 $|\Sigma|$ 为 Σ 的行列式[60]。

3.2.2 连续随机变量的互信息

定义 3-8[60]：两个连续随机变量 X、Y 的互信息 $I(X;Y)$ 定义为

$$I(X;Y)=\int f(x,y)\log\frac{f(x,y)}{f(x)f(y)}dxdy$$

根据互信息、微分熵和条件微分熵的定义可得：

$$I(X;Y)=h(X)-h(X|Y)=h(Y)-h(Y|X) \tag{3-14}$$

两个连续随机变量的互信息具有和离散情况下相似的性质[98]，即

（1）非负性，即 $I(X;Y)\geq 0$

（2）对称性，即 $I(X;Y)=I(Y;X)$

（3）凸状性，即 $I(X;Y)$ 是输入连续变量 X 的概率密度函数 $p(x)$ 的上凸函数；$I(X;Y)$ 是连续信道转移概率密度函数 $p(y|x)$ 的下凸性函数。

（4）信息不增性（数据处理定理），即 $I(X;Z)\leq I(X;Y)$，其中 Z 是 Y 的函数，即 $z=f(y)$，当且仅当 z 与 y 一一对应时等号成立。

连续随机变量 X、Y 的互信息表示出通过观察 Y（或 X）能够获取关于 X（或 Y）的知识的多少，即经过 X 到 Y（或 Y 到 X）的通信所传输的信息量。

3.2.3 基于 Shannon 信息的水印安全性分析

基于 Shannon 信息的分析方法中，利用微分熵 $h(\cdot)$ 来表示随机量的不确定程

度，用互信息$I(\cdot)$来表示水印通信中的信息泄漏。如果记Θ为需要估计的水印密钥（或密钥映射结果），$\mathbf{Y}^{N_o}$为对水印图像的N_o次观测，$\mathbf{M}^{N_o}$为嵌入的信息，$\mathbf{X}^{N_o}$为相应的载体图像信号，则水印密钥在水印通信前的不确定程度为$h(\Theta)$；在掌握水印图像的条件下，水印密钥的不确定程度表示为观测量的条件熵$h\left(\Theta\middle|\mathbf{Y}^{N_o}\right)$，也叫做水印密钥的剩余熵；水印密钥在水印通信过程中所泄漏的信息则表示为水印密钥和观测量之间的互信息，即$I\left(\Theta,\mathbf{Y}^{N_o}\right)=h(\Theta)-h\left(\Theta\middle|\mathbf{Y}^{N_o}\right)$。

结合对水印安全的攻击条件，水印安全性在KMA、KOA和WOA条件下可以具体表示为以下情形。

（1）KMA条件下：

水印密钥的信息泄漏为：

$$I\left(\Theta,\mathbf{Y}^{N_o}\middle|\mathbf{M}^{N_o}\right)=h\left(\mathbf{Y}^{N_o}\middle|\mathbf{M}\right)-h\left(\mathbf{Y}^{N_o}\middle|\Theta,\mathbf{M}^{N_o}\right) \tag{3-15}$$

水印密钥的剩余熵为：

$$h\left(\Theta\middle|\mathbf{Y}^{N_o},\mathbf{M}^{N_o}\right)=h(\Theta)-h\left(\mathbf{Y}^{N_o}\middle|\mathbf{M}\right)+h\left(\mathbf{Y}^{N_o}\middle|\Theta,\mathbf{M}^{N_o}\right) \tag{3-16}$$

（2）KOA条件

水印密钥的信息泄漏为：

$$I\left(\Theta,\mathbf{Y}^{N_o}\middle|\mathbf{X}^{N_o}\right)=h\left(\mathbf{Y}^{N_o}\middle|\mathbf{X}^{N_o}\right)-h\left(\mathbf{Y}^{N_o}\middle|\Theta,\mathbf{X}^{N_o}\right) \tag{3-17}$$

水印密钥的剩余熵为：

$$h\left(\Theta\middle|\mathbf{Y}^{N_o},\mathbf{X}^{N_o}\right)=h(\Theta)-h\left(\mathbf{Y}^{N_o}\middle|\mathbf{X}^{N_o}\right)+h\left(\mathbf{Y}^{N_o}\middle|\Theta,\mathbf{X}^{N_o}\right) \tag{3-18}$$

（3）WOA条件

水印密钥的信息泄漏：

$$I\left(\Theta,\mathbf{Y}^{N_o}\right)=h\left(\mathbf{Y}^{N_o}\right)-h\left(\mathbf{Y}^{N_o}\middle|\Theta\right) \tag{3-19}$$

水印密钥的剩余熵：

$$\begin{aligned}&h\left(\Theta\middle|\mathbf{Y}^{N_o}\right)=h(\Theta)-h\left(\mathbf{Y}^{N_o}\right)+h\left(\mathbf{Y}^{N_o}\middle|\Theta\right)\\&=h(\Theta)-h\left(\mathbf{Y}^{N_o}\right)+I\left(\mathbf{Y}^{N_o};\mathbf{M}^{N_o}\middle|\Theta\right)+h\left(\mathbf{Y}^{N_o}\middle|\Theta,\mathbf{M}^{N_o}\right)\end{aligned} \tag{3-20}$$

因此，水印密钥和观测量之间的互信息反映了水印通信中泄漏的关于水印密钥的知识的多少，而水印密钥的剩余熵则反映了在掌握水印图像之后水印密钥的不确定性。水印密钥的剩余熵越小表明攻击者掌握的有关水印密钥的知识越多，反之则表示掌握水印密钥的知识越少，水印密钥越不确定。

水印安全问题的理论分析方法表明，水印安全问题可以被视为一个通信的问题：被传输的信息是水印密钥，而图像载体和嵌入的信息构成了通信的干扰信道。作为水印算法的设计者，在对水印视觉质量予以约束条件下，即固定失真条件下，应努力使通信中密钥和观测量的互信息最小（以保证安全性），而同时使在已知密钥条件下观测量和秘密信息之间的互信息最大（以保证秘密信息的通信容量）。因此水印安全性与水印鲁棒性、水印不可见性之间存在优化问题。

在实际的分析过程中，基于 Fisher 信息的方法需要能够将观测量对于待估计参数的对数似然函数表达为封闭的解析形式，而且要求该对数似然函数可导，因此从技术上限制了这种方法的应用。例如在对 DC-QIM 的安全性分析和对一些利用边信息有信（Side-Informed）嵌入的水印方法的安全性分析中，基于 Fisher 信息的方法将非常困难。

不论是基于 Fisher 信息的方法还是基于 Shannon 信息的方法，水印安全性都是对攻击者获取水印密钥的困难程度进行衡量。从物理意义上分析，对水印密钥进行估计的可达最小误差反映了在掌握水印图像之后水印密钥的不确定性；关于水印密钥的 Fisher 信息衡量了水印密钥在水印通信中泄漏的信息量。实际上，Fisher 信息和 Shannon 信息都是描述不确定性的量，Fisher 信息描述的是不确定空间的表面积，而 Shannon 信息描述的是不确定空间所占的体积[60]。

水印安全性的理论分析获得的结论是攻击者对水印密钥估计时可达到的准确程度的上限，即攻击者最多能了解到的水印密钥信息。因此关于水印安全的理论分析与实际的攻击不同。实际的攻击算法涉及到具体的估计器的设计。攻击者使用的估计器不一定能够达到理论分析的性能，理论分析给出的安全性是实际攻击算法可达性能的上界。

第四章 加性扩频水印安全性分析

随着数字水印技术的不断发展，水印安全性的研究日益受到重视。传统的水印系统设计主要考虑三个方面的性能，即水印鲁棒性、不可见性和水印的容量。近年来，Cayre 等人在数字水印安全性方面的开创性工作[48] 使得安全性正成为新一代水印系统设计中需考虑的第四个性能指标。

根据密码学领域的 Kerckhoffs[29]原理，没有任何一种加密的算法能够得到永久的保密。水印算法也不例外，即有关水印的嵌入和提取算法终会公开。要保证水印通信的安全，只能依赖水印的密钥。水印通信中密钥的安全性决定了水印通信的安全。对于扩频水印算法[48]，密钥就是产生扩频载波的随机数；而对于基于量化索引调制（Quantization Index Modulation, QIM）的水印算法，密钥就是用于抖动量化网格的随机序列[63]。

水印安全性的概念与水印的不可见性和水印容量有明显的区别，但是其与水印鲁棒性的区别却值得进一步明确。对水印鲁棒性的攻击指的是对水印通信信道的攻击，其目的是增加水印通信的误码率[46]；而对水印安全性攻击的目的是获取有关水印密钥的知识。对水印鲁棒性的攻击可能是有意的或者无意的，而对水印安全性的攻击一定是有意的。对于水印安全性的攻击目前主要分为以下几种情况[48]：KMA（Known Message Attack），KOA（Known Original Attack）和 WOA（Watermarked Only Attack）。KMA 是指攻击者已知嵌入了水印的信号和水印信息本身，通过多次观察来估计水印密钥；KOA 是指攻击者除了掌握隐藏水印的信号，还拥有未嵌入水印的信号载体；而 WOA 是最为困难的一种情况，即攻击者只拥有嵌入了水印的信号。

近年来有关水印安全性的研究中，Cayre 等人的工作[48]被认为是具有开创性意义的。他们明确地区分了水印安全性与鲁棒性的定义，并通过求 Fisher 信息矩阵（Fisher Information Matrix, FIM）给出了扩频水印载波估计误差的 Cramer-Rao 边界（Cramer-Rao Bound, CRB）。在文献[46]中，Comesaña 等人根据 Shannon 信息论，利用求水印信号和扩频载波的互信息的方法分析了扩频水印的安全性。Péres-Freire [63]等人通过分析水印抖动量化调制的支撑集来研究基于 QIM 的水印安全性。Ni[61]等人的工作则充分考虑了自然图像载体的统计特征，利用 Shannon

互信息的方法对扩频水印安全性进行了理论分析和实际的攻击，得到了对秘密载波的更加准确的估计。

目前针对水印安全性的分析，大多假设图像载体呈高斯分布，所得到的分析结论都以高斯分布为基础。众所周知，自然图像小波系数的边缘分布呈现很强非高斯性，即高尖峰和重脱尾。因此，Cayre 等人基于载体高斯分布的水印安全性分析与实际的情况有很大差别。本文根据自然图像的统计特性，利用 GSM（Gaussian Scale Mixture）模型描述其小波系数的分布，并以 FIM 对加性扩频水印的安全性进行理论分析，得到了对秘密载波无偏估计的 CRB 和 MCRB（Modified Cramer-Rao Bound）边界。该结果对于设计更为安全的水印算法具有重要的意义。

本文第一节给出了水印安全性分析一般方法；第二节和第三节分别为基于 GSM 模型的扩频水印安全性理论分析和实验仿真结果；文章的第四节为本文的结论。

4.1 水印安全性的分析方法

对于扩频水印系统，扩频载波（扩频序列）由密钥产生，该载波经嵌入信息调制后嵌入到图像载体中从而生成水印图像。（为了叙述方便，以下将嵌入了水印的图像简称为水印图像）。对于攻击者而言只需要有效估计出扩频载波就可以达到攻击水印安全的目的[48]（例如：去除、替换和破坏水印信号），因此在对水印安全的攻击中对密钥的估计等价为对扩频载波的估计。扩频水印系统不是绝对安全的[48]，在水印通信中会泄漏出关于扩频载波的信息。对水印通信安全的攻击就是通过对水印图像的多次观察，估计出有关扩频载波的信息；而水印安全性分析则是评估水印系统的安全性能并研究影响信息泄漏的有关因素。

考察信息泄漏的方法主要有两大类。一种是利用 Shannon 的信息理论[60]。该方法中，令 $\mathbf{Z}$ 表示秘密的扩频载波；用 $h(\mathbf{Z})$ 表示在水印通信之前扩频载波的熵，它衡量了扩频载波的不确定度；令 $h(\mathbf{Z}|\mathbf{Y})$ 表示水印通信之后，在已知水印图像 $\mathbf{Y}$ 的条件下，扩频载波的条件熵。这里“$h(\cdot)$”表示随机变量的微分熵。扩频载波和水印图像的互信息可以表示为

$$I(\mathbf{Z},\mathbf{Y}) = h(\mathbf{Z}) - h(\mathbf{Z}|\mathbf{Y}) \tag{4-1}$$

即由于水印通信，对于秘密载波不确定度的减少量。对于水印通信安全的攻击者来说，$I(\mathbf{Z},\mathbf{Y})$表示了水印系统关于扩频载波的信息泄漏。另一种方法基于 FIM 来考察对于秘密扩频载波进行估计的准确程度。记$\mathbf{\Theta} = (\theta_1, \theta_2, \cdots \theta_K)^T$为需要估计的一组参数，$\mathbf{Y}$为观察值。对于一个无偏的估计器，记$\mathbf{\Theta}$的估计值为$\hat{\mathbf{\Theta}} = (\hat{\theta}_1, \hat{\theta}_2, \cdots \hat{\theta}_K)^T$，该估计对应的 FIM 为一个$K \times K$矩阵[96]：

$$J = E\left(\{\nabla_\Theta \ln p(Y|\Theta)\} \{\nabla_\Theta \ln p(Y|\Theta)\}^T \right) \tag{4-2}$$

FIM 的每个元素为：

$$J_{ij} = E\left[\frac{\partial \ln p(Y|\Theta)}{\partial \theta_i} \frac{\partial \ln p(Y|\Theta)}{\partial \theta_j} \right] \tag{4-3}$$

根据 Cramer-Rao 不等式[60]，在 FIM 可逆的条件下，估计量均方误差的下界即为 CRB，定义为：

$$Var\left[\hat{\Theta} - \Theta\right] \geq CRB(\Theta) = tr\left(J^{-1}\right) \tag{4-4}$$

其中“$tr(\bullet)$”表示求矩阵的迹，“J^{-1}”表示 FIM 的逆矩阵。FIM 是衡量信息泄漏的量；CRB 是描述根据观察信息估计参数准确程度的量。在本文中，需要估计的参数是扩频载波。信息泄漏得越多，攻击者掌握的关于载波的知识就越丰富，可以得到的估计误差就越小，对秘密载波的估计就更加准确，而相应的水印安全性则越低。

在实际分析中，有时J^{-1}并不存在，尤其是存在一些未知的、但是并不需要进行估计的干扰参数时。在计算传统的 CRB 时，FIM 的定义如（4-2）所示。在随机干扰参数u存在的情况下，有$p(Y|\Theta) = \int_{-\infty}^{\infty} p(Y|u, \Theta) p(u) du$。由于一般情况下该积分很难解析地表达，所以此时难以求出传统的 CRB[99]. 一种替代的办法是使用改进的 CRB（或 MCRB）来衡量对于参数估计的最小界，MCRB 定义为[100]：

$$MCRB(\theta) = \left\{ E_{Y,u}\left[\left(\frac{\partial \ln p(Y|\Theta, u)}{\partial \Theta} \right) \left(\frac{\partial \ln p(Y|\Theta, u)}{\partial \Theta} \right)^{T} \right] \right\}^{-1} \tag{4-5}$$

式（4-5）表明基于Y有关Θ和u的条件似然函数，可以通过对观察信号和干扰参数求统计平均得到待估计参数的MCRB。虽然MCRB比传统的CRB更“松”，但它更容易计算到[100, 101]；如果干扰信息为已知的确定量，MCRB 就是传统的CRB。

在本文中，分别利用CRB和MCRB作为对秘密载波在KMA和WOA情况下的安全性的度量指标。CRB或MCRB的值越小，表示对秘密载波的无偏估计误差越小，水印的安全性越低；反之则表示安全性越高。

4.2 加性扩频水印模型

加性扩频（Additive Spread-Spectrum）水印模型[46, 48]可由图4-1所示。在水印的嵌入部分，秘密载波（密钥）经秘密信息调制之后与图像载体信号相加，得到水印图像。水印图像在传输过程中受到信道噪声的干扰。在水印的检测端，水印检测器利用秘密载波（密钥）实现对秘密信息的检测。

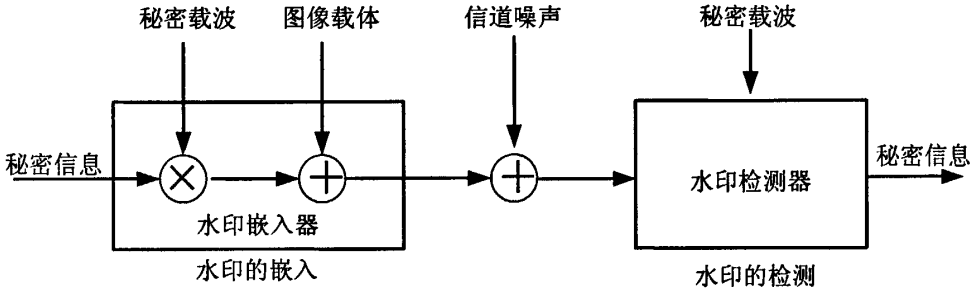


图4-1 加性扩频水印模型

Fig. 4-1 Add-SS based watermarking

在以上模型中，水印图像信号由图像载体与水印信号直接相加得到，水印信号为经过嵌入信息调制后的秘密载波，因此该嵌入过程可以表示为：

$$\mathrm{y}^{j} = \mathrm{x}^{j} + \frac{\gamma}{\sqrt{N_c}} \sum_{i=1}^{N_c} \mathrm{z}_i a_i^{j} \tag{4-6}$$

其中 $\mathbf{x}^j = \left(x_1^j, x_2^j, \cdots, x_{N_v}^j\right)^T$ 和 $\mathbf{y}^j = \left(y_1^j, y_2^j, \cdots, y_{N_v}^j\right)^T$ 分别表示第 j 次观察中的图像载体信号和水印图像信号；$\mathbf{z}_i$ 为第 i 个秘密载波列向量，并用 Z 表示由所有秘密载波列向量构成的矩阵；a_i^j 表示第 j 次观察中第 i 位嵌入信息；γ 为载波的嵌入强度；N_c 用于表示秘密载波的数目；N_o 表示观察的次数；N_v 表示每次观察时图像载体和水印图像的维数，这里假定秘密载波和图像载体维数相同。定义载体和水印的相对功率比为 DWR（Document to Watermark Ratio），有 $DWR = 10\log_{10}\left(\frac{\sigma_x^2}{\gamma^2\sigma_a^2}\right)$。其中 σ_x^2 为载体平均功率，而 $\gamma^2\sigma_a^2$ 为水印的嵌入功率。

以下的分析中，利用 GSM 模型对自然图像载体 $\mathbf{x}^j$ 的统计特性进行描述，假设载体系数相互独立，每次观察之间也相互独立，调制方法采用 BPSK（Binary Phase Shift Keying），并且假设信息嵌入之前进行了伪随机化，因此 a_i^j 的取值也相互独立。由于在 KOA 情况下攻击者已经掌握了载体图像，此时利用 GSM 模型不会带来额外的帮助。因此本文仅对 KMA 和 WOA 情况下的扩频水印安全性进行分析，有关 KOA 的安全性分析可以参见文献[48]。

4.3 KMA 情况

在 KMA 攻击的条件下，攻击者不仅拥有水印图像，而且知道嵌入图像的信息。攻击者的目标就是通过对水印信号的多次观察实现对秘密扩频载波的估计。

4.3.1 单载波估计

为了说明的方便，首先以单载波的情况进行分析。此时，$N_c = 1$，每次嵌入的信息只有 1 位，攻击者拥有对水印图像 N_o 次独立的观察和对应的嵌入信息。由于图像载体各维之间也相互独立，对于第 j 次观察中第 k 维，有 $y_k^j = x_k^j + \gamma a_1^j z_{1k}$，其中 z_{1k} 表示载波（此时只有一个载波）的第 k 位分量；a_1^j 表示对应第 j 次观察时的嵌入信息，即 $a_1^j = 1$ 或 -1。用标量 GSM 模型来描述水

印载体的分布，则 x_k^j 为零均值的高斯分布，其方差为 $s_k^{j2}\sigma_u^2$。由此可以得到在观测集 $\mathbf{Y}^{N_o}=\left(\mathrm{y}^1,\mathrm{y}^2,\cdots,\mathrm{y}^{N_o}\right)$ 下的似然函数和对数似然函数分别为：

$$f\left(\mathrm{Y}^{N_o}\middle|\,\mathrm{z}_1\right)=f\left(\mathrm{y}^1,\cdots,\mathrm{y}^{N_o}\middle|\,\mathrm{z}_1\right)$$
$$=\prod_{j=1}^{N_o}\prod_{k=1}^{N_v}\frac{1}{\sqrt{2\pi s_k^{j2}\sigma_u^2}}\exp\left(-\frac{(y_k^j-\gamma a_1^j z_{1k})^2}{2s_k^{j2}\sigma_u^2}\right) \tag{4-7}$$

$$\log f\left(\mathrm{Y}^{N_o}\middle|\,\mathrm{z}_1\right)=\log\left(\prod_{j=1}^{N_o}\prod_{k=1}^{N_v}\frac{1}{\sqrt{2\pi s_k^{j2}\sigma_u^2}}\exp\left(-\frac{(y_k^j-\gamma a_1^j z_{1k})^2}{2s_k^{j2}\sigma_u^2}\right)\right)$$
$$=\sum_{j=1}^{N_o}\sum_{k=1}^{N_v}\left[\log\frac{1}{\sqrt{2\pi s_k^{j2}\sigma_u^2}}-\frac{\left(y_k^j-\gamma a_1^j z_{1k}\right)^2}{2s_k^{j2}\sigma_u^2}\right] \tag{4-8}$$

求对数似然函数对被估计参数的偏导数，得到

$$\frac{\partial}{\partial z_{1i}}\log f\left(\mathrm{Y}^{N_o}\middle|\,\mathrm{z}_1\right)=\gamma\sum_{j=1}^{N_o}\frac{a_1^j x_i^j}{s_i^{j2}\sigma_u^2} \tag{4-9}$$

根据本章第 1 节的说明，FIM 的元素可由（4-10）和（4-11）确定.

对于 FIM 的对角元素，有：

$$J_{ii}\left(\mathrm{z}_1\right)=\int f\left(\mathrm{Y}^{N_o}\middle|\,\mathrm{z}_1\right)\left(\frac{\partial}{\partial z_{1i}}\log f\left(\mathrm{Y}^{N_o}\middle|\,\mathrm{z}_1\right)\right)^2 d\mathrm{y}^1\cdots d\mathrm{y}^{N_o} \tag{4-10-1}$$

$$=\int\left(\gamma\sum_{j=1}^{N_o}\frac{a_1^j x_i^j}{s_i^{j2}\sigma_u^2}\right)^2\prod_{j=1}^{N_o}\prod_{k=1}^{N_v}\frac{1}{\sqrt{2\pi s_k^{j2}\sigma_u^2}}\exp\left(-\frac{(y_k^j-\gamma a_1^j z_{1k})^2}{2s_k^{j2}\sigma_u^2}\right)d\mathrm{y}^1\cdots d\mathrm{y}^{N_o}$$

$$=\int\left(\gamma\sum_{j=1}^{N_o}\frac{a_1^j x_i^j}{s_i^{j2}\sigma_u^2}\right)^2\prod_{j=1}^{N_o}\prod_{k=1}^{N_v}\frac{1}{\sqrt{2\pi s_k^{j2}\sigma_U^2}}\exp\left(-\frac{(x_k^j)^2}{2s_k^{j2}\sigma_U^2}\right)dx_1^1dx_2^1\cdots dx_{N_v}^1$$
$$\cdots\cdots dx_1^{N_o}dx_2^{N_o}\cdots dx_{N_v}^{N_o} \tag{4-10-2}$$

$$=\int\left(\gamma\sum_{j=1}^{N_o}\frac{a_1^j x_i^j}{s_i^{j2}\sigma_u^2}\right)^2\prod_{j=1}^{N_o}\frac{1}{\sqrt{2\pi s_i^{j2}\sigma_U^2}}\exp\left(-\frac{(x_i^j)^2}{2s_i^{j2}\sigma_U^2}\right)dx_i^1dx_i^2\cdots dx_i^{N_o} \tag{4-10-3}$$

$$= \int\left[\left(\frac{\gamma a_1^1 x_i^1}{s_i^{12}\sigma_U^2}\right)^2 + \left(\frac{\gamma a_1^2 x_i^2}{s_i^{22}\sigma_U^2}\right)^2 + \cdots \left(\frac{\gamma a_1^{N_o} x_i^{N_o}}{s_i^{N_o 2}\sigma_U^2}\right)^2 + 2\frac{\gamma^2 a_1^1 a_1^2 x_i^1}{s_i^{12}\sigma_U^2}\frac{x_i^2}{s_i^{22}\sigma_U^2} + \cdots \right.$$

$$\left. \cdots + 2\frac{\gamma^2 a_1^{N_o-1} a_1^{N_o} x_i^{N_o-1}}{s_i^{(N_o-1)2}\sigma_U^2}\frac{x_i^{N_o}}{s_i^{N_o 2}\sigma_U^2}\right] \cdot \prod_{j=1}^{N_o}\frac{1}{\sqrt{2\pi s_i^{j2}\sigma_U^2}}\exp\left(-\frac{(x_i^j)^2}{2s_i^{j2}\sigma_U^2}\right)dx_i^1 dx_i^2 \cdots dx_i^{N_o}$$

$$= \frac{\gamma^2 s_i^{12}\sigma_U^2}{s_i^{14}\sigma_U^4} + \frac{\gamma^2 s_i^{22}\sigma_U^2}{s_i^{24}\sigma_U^4} + \cdots + \frac{\gamma^2 s_i^{N_o 2}\sigma_U^2}{s_i^{N_o 4}\sigma_U^4} \tag{4-10-4}$$

$$= \sum_{j=1}^{N_o}\frac{\gamma^2}{s_i^{j2}\sigma_U^2} \tag{4-10-5}$$

其中(4-10-1)式是根据 FIM 的定义得到。(4-10-2)式是考虑到在已知载体尺度因子的条件下，每次观察 $\mathbf{y}^j$ 的各维相互独立，其对应载体的各维也相互独立，有 $d\mathbf{y}^j = dy_1^j dy_2^j \cdots dy_{N_v}^j$，因此有以下等式成立：

$$d\mathbf{y}^j = d\mathbf{x}^j = dx_1^j dx_2^j \cdots dx_{N_v}^j$$

$$d\mathbf{y}^1 \cdots d\mathbf{y}^{N_o} = dx_1^1 dx_2^1 \cdots dx_{N_v}^1 \cdots\cdots dx_1^{N_o} dx_2^{N_o} \cdots dx_{N_v}^{N_o}$$

(4-10-3)是因为被积函数中只有各次观察对应载体的第 i 维，其它各维在积分后为值 1。(4-10-4)利用了求高斯分布的均值和方差的定义。

对于 FIM 中的非对角元素，可得：

$$J_{ik}(\mathbf{z}_1) = \int f\left(\mathrm{Y}^{N_o}|\mathbf{z}_1\right)\left(\frac{\partial}{\partial z_{1i}}\log f\left(\mathrm{Y}^{N_o}|\mathbf{z}_1\right)\right)\left(\frac{\partial}{\partial z_{1k}}\log f\left(\mathrm{Y}^{N_o}|\mathbf{z}_1\right)\right)d\mathbf{y}^1 \cdots d\mathbf{y}^{N_o} \tag{4-11-1}$$

$$= \int\left(\sum_{j=1}^{N_o}\frac{\gamma a_1^j x_i^j}{s_i^{j2}\sigma_U^2}\right)\left(\sum_{j=1}^{N_o}\frac{\gamma a_1^j x_k^j}{s_k^{j2}\sigma_U^2}\right)\prod_{j=1}^{N_o}\prod_{t=1}^{N_v}\frac{1}{\sqrt{2\pi s_t^{j2}\sigma_U^2}}\exp\left(-\frac{(x_t^j)^2}{2s_t^{j2}\sigma_U^2}\right)d\mathbf{x}^1 d\mathbf{x}^2 \cdots d\mathbf{x}^{N_o}$$

$$= \int\left(\sum_{j=1}^{N_o}\frac{\gamma a_1^j x_i^j}{s_i^{j2}\sigma_U^2}\right)\prod_{j=1}^{N_o}\frac{1}{\sqrt{2\pi s_i^{j2}\sigma_U^2}}\exp\left(-\frac{(x_i^j)^2}{2s_i^{j2}\sigma_U^2}\right)dx_i^1 dx_i^2 \cdots dx_i^{N_o}$$

$$\cdot \int\left(\sum_{j=1}^{N_o}\frac{\gamma a_1^j x_k^j}{s_k^{j2}\sigma_U^2}\right)\prod_{j=1}^{N_o}\frac{1}{\sqrt{2\pi s_k^{j2}\sigma_U^2}}\exp\left(-\frac{(x_k^j)^2}{2s_k^{j2}\sigma_U^2}\right)dx_k^1 dx_k^2 \cdots dx_k^{N_o} \tag{4-11-2}$$

$$= 0 \qquad \text{for all } i \neq k \tag{4-11-3}$$

其中，(4-11-1)式由 FIM 的定义得到。(4-11-2)利用了载体各维在已知尺度因子条件下相互独立的性质。(4-11-3)利用了求高斯分布的均值的定义。

所以，在单载波条件下，可以求出 FIM 为：

$$J(\mathbf{z}_1)=\frac{\gamma^2}{\sigma_u^2}\begin{bmatrix}\sum_{j=1}^{N_o}\frac{1}{s_1^{j2}} & 0 & \cdots & 0\\ 0 & \sum_{j=1}^{N_o}\frac{1}{s_2^{j2}} & \cdots & 0\\ \cdots & \cdots & \cdots & 0\\ 0 & 0 & \cdots & \sum_{j=1}^{N_o}\frac{1}{s_{N_v}^{j2}}\end{bmatrix} \tag{4-12}$$

根据 Cramer-Rao 定理，可以得到在 KMA 攻击下，对单载波的估计的界为：

$$CRB(\mathbf{z}_1)=tr\left(J(\mathbf{z}_1)^{-1}\right)=\frac{1}{\gamma^2}\sum_{i=1}^{N_v}\frac{\sigma_u^2}{\sum_{j=1}^{N_o}1/s_i^{j2}} \tag{4-13}$$

CRB 反映了对参数无偏估计的准确程度。由（4-13）式可知，在 KMA 条件下对单载波无偏估计的最小均方误差与秘密载波的长度和观察的次数有关。秘密载波越长，越难以对秘密载波进行准确估计；同时观察次数越多，获得的有关秘密载波的信息越多，对其估计也就越准确。考虑到 GSM 模型对图像小波系数非高斯分布的刻画，以上 CRB 与秘密载波的长度以及观察次数的关系也是非线性的。

在文献[48]中，Cayre 等人利用高斯模型描述图像载体，通过计算 FIM 得到了 KMA 条件下对秘密单载波的无偏估计界，记为 $tr\left(J_c(\mathbf{z}_1)^{-1}\right)=\frac{N_v\sigma_x^2}{\gamma^2 N_o}$，其中 σ_x^2 为载体图像小波系数的方差。以下比较（4-13）和 Cayre 的结果，为公平起见，假定 $\frac{1}{N_oN_v}\sum_{i=1}^{N_v}\sum_{j=1}^{N_o}s_i^{j2}\sigma_u^2=\sigma_x^2$，即 GSM 模型和高斯模型中载体图像小波系数的平均方差相等。由“算数平均－调和平均”不等式[102]

$$\frac{1}{N}\sum_{j=1}^{N}\frac{1}{m_j}\geq\frac{N}{m_1+m_2+\cdots+m_N} \tag{4-14}$$

（其中 m_j 均为正数），可知

$$tr(J(\mathbf{z}_1)^{-1})=\frac{1}{\gamma^2}\sum_{i=1}^{N_v}\frac{\sigma_u^2}{\sum_{j=1}^{N_o}1/s_i^{j2}}\leq\frac{1}{\gamma^2N_o^2}\sum_{i=1}^{N_v}\sum_{j=1}^{N_o}s_i^{j2}\sigma_u^2=tr\left(J_C(\mathbf{z}_1)^{-1}\right) \tag{4-15}$$

其中当 $s_i^1=s_i^2=\cdots=s_i^{N_o}$ 时取“＝”。（4-15）表明：文献[48]中基于高斯

模型的扩频水印系统安全性能评估是相对“放大”的。比较高斯模型，利用 GSM 模型可以更加准确地描述自然图像小波系数的统计分布，据此对秘密载波进行的无偏估计和对扩频水印安全性的评价也更为准确。

4.3.2 多载波估计

在 KMA 条件下，对多个秘密载波进行估计时，其对应的调制信息（嵌入信息）已知。此时在观测集 $\mathrm{Y}^{N_o}=\left(\mathrm{y}^1,\mathrm{y}^2,\cdots\mathrm{y}^{N_o}\right)$ 下的对数似然函数和 FIM 分别为：

$$\log f\left(\mathrm{Y}^{N_o}\middle|Z\right)=\sum_{j=1}^{N_o}\sum_{k=1}^{N_v}\left[\log\frac{1}{\sqrt{2\pi s_k^{j2}\sigma_u^2}}-\frac{\left(y_k^j-\frac{\gamma}{\sqrt{N_c}}\sum_{m=1}^{N_c}a_m^j z_{mk}\right)^2}{2s_k^{j2}\sigma_u^2}\right] \tag{4-16}$$

$$J(Z)=E\left\{\left[\frac{\partial\log f\left(\mathrm{Y}^{N_o}\middle|Z\right)}{\partial\left(\mathrm{z}_1^T,\cdots,\mathrm{z}_{N_c}^T\right)^T}\right]\left[\frac{\partial\log f\left(\mathrm{Y}^{N_o}\middle|Z\right)}{\partial\left(\mathrm{z}_1^T,\cdots,\mathrm{z}_{N_c}^T\right)^T}\right]^T\right\} \tag{4-17}$$

FIM 中的各元素可由以下方法求出：

$$\begin{aligned}
&E\left(\frac{\partial\log f\left(\mathrm{Y}^{N_o}\middle|\mathrm{z}\right)}{\partial z_{mn}}\right)\left(\frac{\partial\log f\left(\mathrm{Y}^{N_o}\middle|\mathrm{z}\right)}{\partial z_{pq}}\right)\\
&=\frac{\gamma^2}{\sigma_u^4 N_c}E\left(\sum_{j=1}^{N_o}\frac{a_m^j x_n^j}{s_n^{j2}}\sum_{k=1}^{N_o}\frac{a_p^k x_q^k}{s_q^{k2}}\right)\\
&=\frac{\gamma^2}{\sigma_u^2 N_c}\sum_{j=1}^{N_o}\frac{a_m^j a_p^j}{s_n^{j2}}\delta_{n,q}\\
&=\frac{\gamma^2}{\sigma_u^2 N_c}J_{(m,n)(p,q)}
\end{aligned} \tag{4-18}$$

其中 z_{mn} 表示第 m 个载波的第 n 维。$J_{(m,n)(p,q)}$ 对应与 FIM 中第 $(m-1)\times N_v+n$ 行，第 $(p-1)\times N_v+q$ 列的元素，即 FIM 按照以下规律构成：

$$J(Z) = \frac{\gamma^2}{\sigma_u^2 N_c}\begin{bmatrix} J_{(1,1)(1,1)} & J_{(1,1)(1,2)} & \cdots & J_{(1,1)(N_c,N_v)} \\ J_{(1,2)(1,1)} & J_{(1,2)(1,2)} & \cdots & J_{(1,2)(N_c,N_v)} \\ \vdots & \vdots & & \vdots \\ J_{(N_c,N_v)(1,1)} & J_{(N_c,N_v)(1,2)} & \cdots & J_{(N_c,N_v)(N_c,N_v)} \end{bmatrix} \tag{4-19}$$

以上 FIM 可以看作由 $N_c \times N_c$ 个分块矩阵构成，每个分块矩阵为 $N_v \times N_v$ 的对角矩阵。与文献[48]中得到的 KMA 条件下的 FIM 相似，（4-19）式是对于嵌入信息敏感的，即嵌入信息的取值影响到（4-19）式的逆存在与否。这里采用文献[48]中使用的方法，令 N_o 趋于无穷大，近似地分析观测次数无穷大时 FIM 的特征以及对应的 CRB。

如果每次观察中，各个图像载体对应的 s_i^j 相等，即 $s_i^j = s$ $\left(1 \le i \le N_v, 1 \le j \le N_o\right)$。则当 N_o 趋于无穷大时，（4-19）中主对角线元素为 $\frac{N_o}{s^2}\left(i = 1, \cdots, N_v\right)$，主对角线以外的非零元素 $\sum_{j=1}^{N_o} \frac{a_m^j a_n^j}{s_i^{j2}} = \frac{1}{s^2}\sum_{j=1}^{N_o} a_m^j a_n^j \to 0$。这样（4-19）近似成为一个对角阵，可以求出该对角阵的逆，进而得到对应的 CRB 为：

$$CRB\left(Z\right)_C = tr\left(J_C(Z)^{-1}\right) = \frac{N_c^2 N_v s^2 \sigma_u^2}{N_o \gamma^2} \tag{4-20}$$

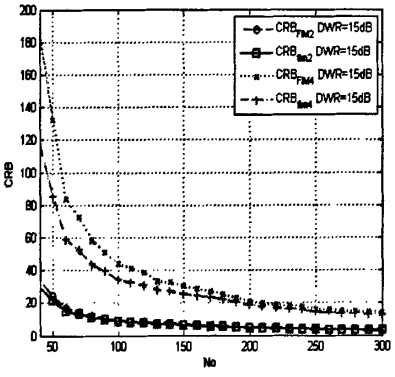
这便是[48]中的结果，在此记为 $CRB(Z)_C$。

如果每次观察中，各个图像载体对应的 s_i^j 不相等，这是自然图像载体的一般情况。当观测次数无限大时，（4-19）中的主对角线元素为 $\sum_{j=1}^{N_0} 1/s_i^{j2}\left(i = 1, \cdots, N_v\right)$。由于 $a_m^j a_n^j$ 随机取值（1 或−1），且 s_i^{j2} 均为正值，所以主对角线以外元素远小于主对角线上的元素。这种现象随着观测次数增大而趋于明显。文献[45]中对于 GSM 模型的研究表明，自然图像中尺度随机变量 s^2 的概率密度函数满足 $p\left(s^2\right) \propto 1/s^2$。由于 s^2 恒为正值，不失一般性，可以假设 s^2 同分布。根据嵌入信息和尺度变量的独立性以及大数定律可知

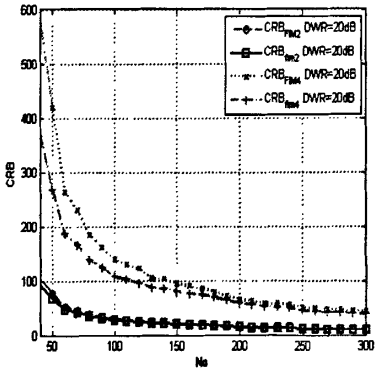
$$\sum_{j=1}^{N_o} \frac{a_m^j a_n^j}{s_i^{j2}} \xrightarrow{N_o \to \infty} N_o E\left(\frac{a_m^j a_n^j}{s_i^{j2}}\right) = 0 \tag{4-21}$$

此时（4-19）退化成为仅保留其主对角线元素的对角阵，于是可以用（4-19）中的主对角线元素构成的对角阵来简化处理，通过求其逆矩阵的迹得到对应的CRB，记为CRB_{fim}：

$$CRB_{fim}(Z) = \frac{N_c^2 \sigma_u^2}{\gamma^2} \sum_{i=1}^{N_v} \frac{1}{\sum_{j=1}^{N_o} 1/s_i^{j2}} \tag{4-22}$$



（a）DWR=15dB



（b）DWR=20dB

图 4-2　CRB_{FIM} 和 CRB_{fim} 性能比较

Fig. 4-2 Performance of CRB_{FIM} and CRB_{fim}

实验仿真结果也进一步验证了（4-22）的正确性。记（4-19）式逆矩阵的迹

为CRB_{FIM}，图 4-2 给出了对于自然图像，简化的CRB_{fim}和CRB_{FIM}的关系。图 4-2（a)和图 4-2(b)分别为载波长度为 512 维，DWR 为 15dB 和 20dB 条件下CRB_{FIM}和CRB_{fim}的曲线。图中载体随机选自 8 幅自然图像小波分解的 HL2，LH2 和 HH2 的子带系数，曲线名称下标中的“2”和“4”分别表示所使用的扩频载波数。实验结果表明，随着观测次数N_0的增大，CRB_{fim}和CRB_{FIM}趋于接近。

由（4-22）同时可知，在 KMA 条件下，由于需要估计的各条载波地位平等，即当嵌入功率$\gamma^2 = 1$时，$N_c = 1$，对多载波进行估计的CRB_{fim}与对单载波$(N_c = 1)$估计的 CRB 有相同的形式。此时，（4-22）式与（4-13）式相同。

4.4 WOA 情况

在 WOA 情况下，攻击者仅仅拥有对水印图像的观察值。需要根据这些观察值来估计秘密载波。对于攻击者来说，这是对水印安全性攻击中最为困难的一种。此时，只有秘密载波是攻击者的估计目标，但是未知的嵌入信息却会影响到对秘密载波的正确估计[48,96]。

由于秘密载波和嵌入信息都是未知参量，在 WOA 条件下不能保证 FIM 的逆矩阵存在[48]，从而难以直接求出传统的 CRB。 Cayre 等人在单载波估计的情况下[48]，通过引入对载波能量的约束条件，使用文献[99]中介绍的方法构建了载波约束条件的零空间H，在$\left(H^T \cdot FIM \cdot H\right)^{-1}$存在的条件下求得估计秘密载波的 CRB。而在多载波情况下，必须假设攻击者已知N_m个嵌入信息，从而得到$N_m \times N_c$个附加的约束关系，用与单载波情况下相同的方法得到了相应的 CRB。

文献[48]中为了求得估计秘密载波的 CRB 需要假定攻击者已知N_m个嵌入信息，这并不符合 WOA 攻击的要求。考虑到实际的 WOA 中，如果存在未知的干扰参数，特别是在非高斯噪声条件下，载波估计的 CRB 并不一定能用封闭的形式表达[103]，本文利用 MCRB 对 WOA 情况下扩频水印安全性进行评估。

在 WOA 情况下，根据 MCRB 对秘密载波矢量进行估计，可以定义对应的

FIM 为：$J_M(Z) = E_{Y,a}\left(\psi\psi^T\right)$，其中

$$\psi = \frac{\partial \log p(\mathrm{Y}^{N_o}\left| Z, \mathrm{a}\right.)}{\partial\left(\mathrm{z}_1^T, \cdots, \mathrm{z}_{N_c}^T\right)^T} \tag{4-23}$$

$$\log p(\mathrm{Y}^{N_o}\left| Z, \mathrm{a}\right.) = \sum_{j=1}^{N_o}\sum_{k=1}^{N_v}\left[\log \frac{1}{\sqrt{2\pi s_k^{j2}\sigma_u^2}} - \frac{\left(y_k^j - \frac{\gamma}{\sqrt{N_c}}\sum_{m=1}^{N_c} a_m^j z_{mk}\right)^2}{2s_k^{j2}\sigma_u^2}\right] \tag{4-24}$$

$$\left[J_{\mathrm{M}}\left(Z\right)\right]_{(m,n),(p,q)} = E_{\mathrm{Y,a}}\left[\frac{\partial \ln p\left(\mathrm{Y}^{N_o}\left| Z, \mathrm{a}\right.\right)}{\partial z_{mn}}\frac{\partial \ln p\left(\mathrm{Y}^{N_o}\left| Z, \mathrm{a}\right.\right)}{\partial z_{pq}}\right] \tag{4-25}$$

而且 $Var\left[\hat{Z} - Z\right] \geq tr\left(J_M(Z)^{-1}\right) = MCRB\left(Z\right)$。其中 $J_{\mathrm{M}}\left(Z\right)$和载波 Z 的下标含义与 4.2 节中所述相同。由于

$$\begin{aligned} &E_{\mathrm{Y,a}}\left[\frac{\partial \ln p\left(\mathrm{Y}^{N_o}\left| Z, \mathrm{a}\right.\right)}{\partial z_{mn}}\frac{\partial \ln p\left(\mathrm{Y}^{N_o}\left| Z, \mathrm{a}\right.\right)}{\partial z_{pq}}\right] \\ &= E_a\left\{E_{\mathrm{Y}|a}\left[\frac{\partial \ln p\left(\mathrm{Y}^{N_o}\left| Z, \mathrm{a}\right.\right)}{\partial z_{mn}}\frac{\partial \ln p\left(\mathrm{Y}^{N_o}\left| Z, \mathrm{a}\right.\right)}{\partial z_{pq}}\right]\right\} \end{aligned} \tag{4-26}$$

可以看出，计算 $J_{\mathrm{M}}\left(Z\right)$是先求出在假设秘密载波和嵌入信息已知情况下的 FIM，然后对嵌入信息求期望值。因此

$$E_{\mathrm{Y}|\mathrm{a}}\left(\frac{\partial \log p\left(\mathrm{Y}^{N_o}\left| Z, \mathrm{a}\right.\right)}{\partial z_{mn}}\right)\left(\frac{\partial \log p\left(\mathrm{Y}^{N_o}\left| Z, \mathrm{a}\right.\right)}{\partial z_{pq}}\right)^T = \frac{\gamma^2}{\sigma_u^2 N_c}\sum_{j=1}^{N_o}\frac{a_m^j a_p^j}{s_n^{j2}}\delta_{n,q} \tag{4-27}$$

$$\left[J_{\mathrm{M}}\left(Z\right)\right]_{(m,n)(p,q)} = E_{\mathrm{a}}\left[\frac{\gamma^2}{\sigma_U^2 N_c}\sum_{j=1}^{N_o}\frac{a_m^j a_p^j}{s_n^{j2}}\delta_{n,q}\right] = \frac{\gamma^2}{\sigma_u^2 N_c}\sum_{j=1}^{N_o}\frac{E_{\mathrm{a}}\left(a_m^j a_p^j\right)}{s_n^{j2}}\delta_{n,q} \tag{4-28}$$

由（4-28）式可见 $J_{\mathrm{M}}\left(Z\right)$与嵌入信息的分布有关。若嵌入信息经过伪随机化且均

值为零，则有

$$E_{\mathrm{a}}\left(a_m^j a_p^j\right)=\begin{cases}\sigma_a^2 & m=p\\ 0 & m\neq p\end{cases}$$

其中σ_a^2为嵌入序列的方差。因此式（4-28）可以表示为

$$\left[J_{\mathrm{M}}(Z)\right]_{(m,n)(p,q)}=\frac{\sigma_a^2\gamma^2}{\sigma_u^2 N_c}\sum_{j=1}^{N_o}\frac{1}{s_n^{j2}}\delta_{m,p}\delta_{n,q} \tag{4-29}$$

由此可得在 WOA 条件下估计秘密载波的 MCRB 为：

$$MCRB(Z)=tr\left(J_M(Z)^{-1}\right)=\frac{\sigma_u^2 N_c^2}{\sigma_a^2\gamma^2}\sum_{i=1}^{N_v}\frac{1}{\sum_{j=1}^{N_o}1/s_i^{j2}} \tag{4-30}$$

若嵌入信息的方差为 1，则 WOA 情况下的结论与 KMA 情况下的结论相同。在一般情况下，秘密载波在 WOA 情况下的安全性与嵌入序列的方差有线性关系。

求 MCRB 的过程中只需要假定干扰信息的分布，而不需要攻击者已知N_m条干扰信息，因此利用 MCRB 计算秘密载波的估计界更为接近 WOA 攻击的条件。

4.5 仿真结果及其分析

本节根据以上有关基于 GSM 模型扩频水印安全性的理论分析结果，对多幅具有不同纹理特征的自然图像（包括 aerial，baboon，barb，boat，f16，lena，peppers 和 sailboat），分别计算相应扩频水印系统在 KMA 和 WOA 条件下的 CRB 和 MCRB，并和 Cayre 的结果[48]进行比较和分析。实验中使用双正交 9/7 小波对自然图像进行 2 层分解，随机选取 HL2, LH2 和 HH2 的子带系数作为载体，载波维数为 512，载波的数目为 1，2 和 4。

4.5.1 KMA 条件下的 CRB

图 4-3 是载波长度为 512，DWR 分别为 20 *dB* 和 15 *dB* 条件下，当观测次数（N_o）增大时，基于 GSM 模型和高斯模型的 CRB 比较。由于 GSM 模型可以更准确地描述自然图像小波系数的统计分布，与高斯模型相比，具有更低的

CRB，即对于秘密载波的估计更加准确。CRB 在观察次数较小时下降地很快，这说明攻击者的前几次观察对于估计秘密载波的贡献很大，之后的观察中携带的载波信息与以前获得的信息会有重复，所以对于估计秘密载波的贡献逐步减小。

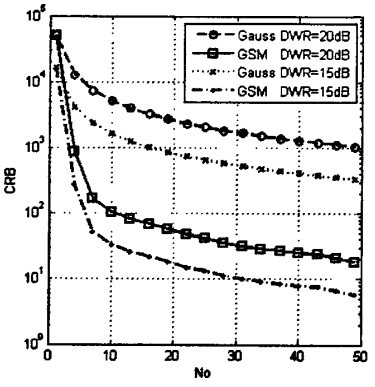


图 4-3 KMA 条件下 CRB 与观测次数的关系（载波长为 512）

Fig. 4-3 CRB against the number of observation under KMA ($N_v = 512$)

图 4-4 比较了 GSM 模型和高斯模型下，观察次数为 50 次，DWR 分别为 20 *dB* 和 15 *dB* 时，CRB 随载波长度（N_v）变化的关系。由于高斯模型认为各次观察中载体方差相同，CRB 与载波的维数成线性关系；而基于 GSM 模型的分析中，各次观察的载体方差由尺度因子控制，由于各次观测的图像载体系数方差不同，使得由秘密载波长度增加所带来的估计不确定度的增加量不相同，即有非线性关系。一般地，CRB 会随载波维数增长而增大，表明增加秘密载波长度可以加大估计的难度，从而提高了水印系统的安全性。

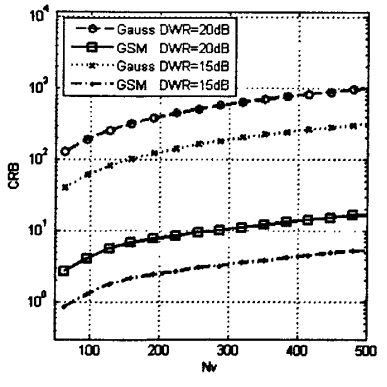


图 4-4 KMA 条件下 CRB 与载波长度的关系（观测次数为 50）

Fig. 4-4 CRB against the length of carrier under KMA ($N_o = 50$)

由图 4-3 和图 4-4 可以看出，DWR 越低，表示嵌入水印的功率相对越高；相应的 CRB 值越低，表示对秘密载波估计的准确性越高，与此 DWR 对应的水印安全性就越低。

图 4-5 和图 4-6 分别表示 KMA 条件下当 DWR 为 15 *dB* 和 20 *dB* 时，对多载波（载波数为 2 和 4）进行无偏估计的 CRB，其中载波的长度为 512。图中标记 GSM_{fim} 的曲线表示为了简化计算由（4-22）式得到的 CRB；而标记 Gauss 的曲线为根据文献[48]中的基于高斯模型方法计算的 CRB。图中曲线标记的数字表示估计的载波数。可以看出基于 GSM 的 CRB 具有更低的值。

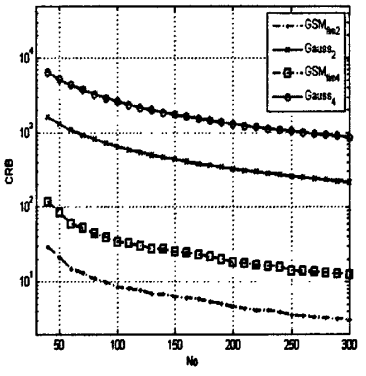


图 4-5 KMA 条件下对秘密载波估计的 CRB （DWR=15dB）

Fig. 4-5 CRBs for 2 and 4 carriers under KMA （DWR=15dB）

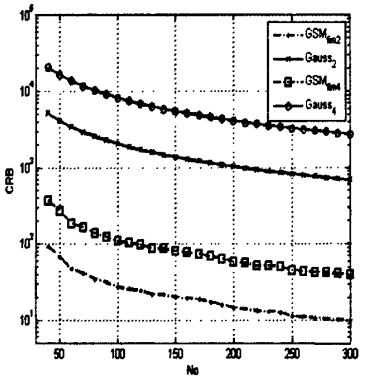


图 4-6 KMA 条件下对秘密载波估计的 CRB （DWR=20dB）

Fig. 4-6 CRBs for 2 and 4 carriers under KMA （DWR=20dB）

4.5.2 WOA 条件下的 MCRB

图 4-7 和图 4-8 分别是在 WOA 条件下，DWR 为 15 *dB* 和 20 *dB* 时，载波长

为 512 时的 MCRB。$MCRB_1$，$MCRB_2$ 和 $MCRB_4$ 分别为单载波，2 载波和 4 载波估计时的 MCRB。由于 MCRB 和 CRB 都是对于参数估计总的误差的度量，随着载波数目的增多，在相同条件下总的估计误差增加；而随着观察次数的增多，对于参数估计的误差逐渐降低。通过比较图 4-7 和图 4-8，可以看出在其他条件相同时，高的 DWR 值导致了高的 MCRB 值，即嵌入水印功率越低，水印的安全性越高。

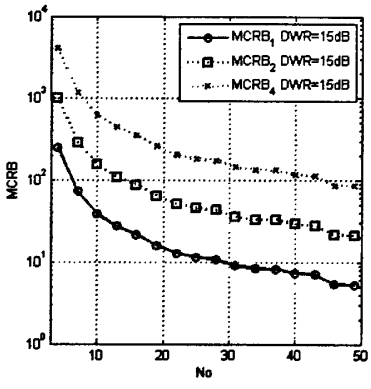


图 4-7 WOA 条件下的 MCRB (DWR=15dB)

Fig. 4-7 MCRBs for 1, 2 and 4 carriers under WOA (DWR=15dB)

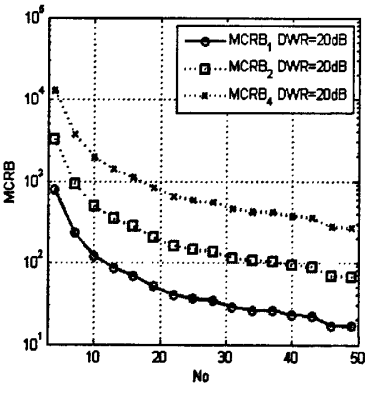


图 4-8 WOA 条件下的 MCRB (DWR=20dB)

Fig. 4-7 MCRBs for 1, 2 and 4 carriers under WOA (DWR=20dB)

4.5.3 不同图像载体的 CRB 比较

图 4-9 和图 4-10 比较了在 KMA 情况下，baboon 和 f16 分别在固定载波长度为 512 和固定观测次数为 30 次条件下，DWR＝15 *dB* 时的 CRB。同时，用带圆

圈的曲线表示了对应条件下基于高斯模型的 CRB。考虑到 baboon 与 f16 相比，图像细节变化更为丰富；采用 GSM 模型描述时，前者的小波系数方差总体上会大于后者。从图 4-9 和图 4-10 中可以看到，在不同的观察次数和载波长度下，由 baboon 得到的 CRB 都大于 f16 的 CRB，故以 baboon 作为载体的扩频水印图像将具有更高的安全性。一般地，对于扩频水印系统，选择具有丰富细节变化的图像作为载体将能得到更高的安全性[104]。

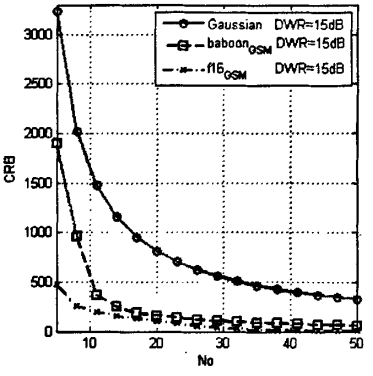


图 4-9 不同图像载体的 CRB 比较 $N_v = 512$

Fig. 4-9 Comparison of CRB from 'baboon' and 'f16' ($N_v = 512$)

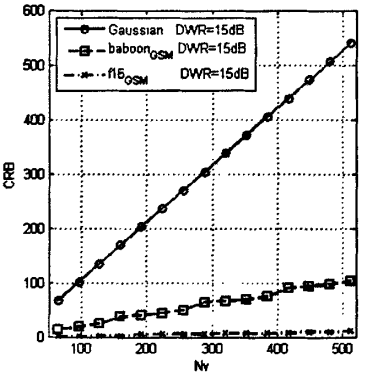


图 4-10 不同图像载体的 CRB 比较 $N_o = 30$

Fig. 4-10 Comparison of CRB from 'baboon' and 'f16' ($N_o = 30$)

4.6 本章小节

本文利用 GSM 模型刻画自然图像小波系数的统计分布，并对加性扩频水印

系统的安全性进行了理论分析，得到了 KMA 和 WOA 情况下对扩频载波估计的 CRB 和 MCRB。分析结果表明由于载体图像统计分布的非高斯性，加性扩频水印安全性与观测次数和扩频载波的长度具有非线性关系，而与扩频载波的嵌入能量及嵌入信息的方差有线性关系。与 Cayre 等人的工作[48]相比，由于本文对于自然图像载体统计分布特点进行了更为精确的刻画，获得了对加性扩频水印系统安全性的更准确评价。本文的工作同时对设计新一代的安全、鲁棒水印算法具有重要的意义。

第五章 基于自然图像模型的 ISS 水印安全性分析

有关数字水印安全性的研究，是基于 Kerckhoffs[29]原理，利用统计和信息论的方法，研究水印算法中密钥的安全程度。通过分析水印密钥的信息泄露与嵌入方法、嵌入参数、密钥长度、密钥分布特点和观测次数等因素之间的关系，对水印算法的安全性能予以度量，并为进一步设计高安全、大容量的新一代鲁棒水印提供依据。近年来，安全性研究已经成为数字水印领域中的重要方向。

基于扩频的调制方法已被广泛地用于数字图像水印领域。传统的扩频水印（或被称为加性扩频水印，Additive Spread –Spectrum Watermarking）算法中，作为载波的扩频序列经过秘密信息的调制，被直接加到图像载体上构成水印图像。在这样的嵌入方法中，载体信号可以被看作是一个干扰源，而且它造成的干扰远强于秘密信息通信过程中受到的其它干扰。为了减少载体信号对于秘密信息通信的干扰，文献[14]中提出了一种称为 ISS 水印(Improved Spread-spectrum Watermarking, 改进的扩频水印)的方法。该算法在秘密载波方向引入对载体信号的衰减，使得水印的嵌入与载体及控制参数相关。由于 ISS 水印具有更好的鲁棒性[14]，并且可以将传统的加性扩频水印视为一种特例，ISS 水印已日益成为扩频水印的主流。

近年来，关于传统的加性扩频水印安全性的研究已经多次见于文献。其中，Cayre 等人最早明确了水印鲁棒性和水印安全性的区别，并利用 Fisher 信息对传统的扩频水印安全进行了研究[48]。Comesaña 等人从信息论的角度出发，在假设图像载体分布为高斯的条件下分析了传统扩频水印的安全[46]。文献[61]和文献[105]利用自然图像模型来描述图像载体分布，分别通过信息论和统计的方法来讨论传统的扩频水印安全性。目前，对于 ISS 水印安全的研究仅见于文献[67]，Pérez-Freire 等人利用高斯模型来描述图像载体分布，通过 Shannon 信息论的方法对 ISS 的安全性进行了分析。

由于基于扩频的水印算法安全性不仅与秘密载波有关，还受到图像载体分布影响[46, 48, 61]，因此，准确描述图像载体分布特性对于扩频水印安全性研究至关重要。自然图像小波系数的分布具有很强的非高斯性，例如小波系数边际分布的高尖峰、重托尾现象，高斯模型已经不能很好地描述这样的特点。本文通过 GSM

（Gaussian Scale Mixture，高斯尺度混合）模型[70]来描述自然图像的统计分布，利用信息论的方法来分析 ISS 水印的安全性。相比于高斯模型，GSM 模型能够更准确地刻画自然图像载体小波系数的分布特点，据此得到的水印安全性能也更为准确。

按照数学量表示方法的惯例，在没有特别的说明时，本章使用大写字母表示随机量，用粗体字母表示矢量。本章的第一部分介绍 ISS 水印模型；第二部分介绍利用 Shannon 信息论分析水印安全的方法；在第三部分，分别在 KMA（Known Message Attack）和 WOA（Watermarked Only Attack）条件下分析了 ISS 水印的安全性；第四部分进行了仿真和讨论；第五部分给出了本文的结论。

5.1 ISS 水印模型

ISS 水印的嵌入方程[14]可以表示为公式（5-1）

$$\mathbf{Y}=\mathbf{X}+\mu(\mathbf{X},\mathbf{M})\mathbf{Z} \tag{5-1}$$

其中，$\mathbf{X}$ 表示载体信号矢量；$\mathbf{Z}$ 为秘密载波；$\mathbf{M}$ 为嵌入的秘密信息；$\mu(\mathbf{X},\mathbf{M})$ 表示嵌入的水印强度是载体信号和秘密载波的函数；$\mathbf{Y}$ 是含水印的信号，若不考虑水印传输过程中的噪声，$\mathbf{Y}$ 就是观测到的信号。与传统的加性扩频水印模型相比，最终嵌入的水印强度是与载体信号相关的，是一种“有信”（Informed）的水印嵌入。

为了分析和使用的方便，实际的 ISS 水印都使用线性模型[14, 67]，该模型可以由公式（5-2）定义。

$$\mathbf{Y}_j=\mathbf{X}_j+(-1)^{M_j}v\mathbf{Z}-\lambda\frac{\mathbf{X}_j^T\mathbf{Z}}{\|\mathbf{Z}\|^2}\mathbf{Z} \tag{5-2}$$

在这里，仅考虑每次嵌入 1 比特秘密信息。第 j 次水印通信中的载体信号、秘密载波和嵌入了水印的信号均被视为 N_v 维随机矢量，分别用 $\mathbf{X}_j$、$\mathbf{Z}$ 和 $\mathbf{Y}_j$ 表示；第 j 次嵌入的秘密信息为随机标量，记为 M_j；v 用于表示秘密扩频载波的嵌入强度；$\lambda\ (0\le\lambda\le1)$ 用于表示对载体信号的衰减参数。由公式（5-2）可以看出，ISS 的线性嵌入模型中引入了对于信号载体的衰减，而且衰减与秘密载波的方向

相同，衰减程度由载体在秘密载波方向上的投影以及衰减参数λ的乘积来确定。ISS 水印嵌入方法在失真约束条件下明显提高了水印的鲁棒性[14, 67]。

在利用 GSM 模型描述自然图像载体统计分布的方法中，用一个随机场$\mathbf{X}$来表示自然图像的小波系数矢量分布。该随机场由一个高斯随机场$\mathbf{U}$和一个尺度标量场S来控制，表示为$\mathbf{X}=S\cdot\mathbf{U}=\{s_i\cdot\mathbf{u}_i, i\in I\}$，其中$I$为矢量位置的索引，$\mathbf{U}\sim N(0,\mathbf{Q})$为零均值的高斯随机场，协方差矩阵为$\mathbf{Q}$。$S$与$\mathbf{U}$相互独立，是一个值为正的尺度随机变量，用于控制图像小波系数的方差。对于自然图像，s可根据最大似然算法由公式（5-3）估计得到[70]：

$$\hat{s}=\arg\max_{s}\left\{\log p\left(\mathbf{x}|s\right)\right\}=\sqrt{\frac{\mathbf{x}^T\mathbf{Q}^{-1}\mathbf{x}}{N}} \tag{5-3}$$

在给定s的条件下，$\mathbf{X}$服从高斯分布，其概率密度函数可以表示为公式（5-4）。

$$P_{\mathbf{x}|s}\left(\mathbf{x}|s\right)=\frac{1}{\left(2\pi\right)^{\frac{N}{2}}\left|s^2\mathbf{Q}\right|^{\frac{1}{2}}}\exp\left(-\frac{\mathbf{x}^T\mathbf{Q}^{-1}\mathbf{x}}{2s^2}\right) \tag{5-4}$$

如果假定$\mathbf{X}$为标量，以上模型退化为标量 GSM 模型。此时，在s_i已知的条件下，X_i的分布为高斯，即有$p_{X_i|s_i}\left(x_i|s_i\right)\sim N\left(0,s_i^2\sigma_u^2\right)$，其中$U$为全局高斯随机场，不失一般性，可以假定$\sigma_u^2$为单位值。对于不同的$i$和$j$，$X_i$与$X_j$相互独立；

本文利用标量 GSM 模型描述自然图像载体小波系数的分布。分析过程中，记$\mathbf{Y}^{N_o}=\left(\mathbf{Y}_1,\mathbf{Y}_2,\cdots,\mathbf{Y}_{\mathbf{N}_o}\right)$表示对水印图像的$N_o$次观测；$\mathbf{X}^{N_o}=\left(\mathbf{X}_1,\mathbf{X}_2,\cdots,\mathbf{X}_{N_o}\right)$表示$N_o$次观测对应的图像载体信号；$\mathbf{M}^{N_o}=\left(M_1,M_2,\cdots,M_{N_o}\right)$为$N_o$次观测对应的秘密信息，其中$M_i\in\{0,1\}$且取值等概率；记$\mathbf{Z}$为水印算法使用的秘密载波，秘密载波为独立同分布的高斯矢量，即服从$\mathbf{Z}\sim N\left(0,\sigma_z^2 I_{N_v}\right)$的分布。假设图像载体、秘密载波和观测信号长度均为$N_v$。根据（5-2）式容易得出，对于第$j$次观测，ISS 线性模型中水印的平均嵌入功率可由以下推导得出。

$$D_w=\frac{1}{N_v}E\left[\left\|\mathbf{Y}_j-\mathbf{X}_j\right\|^2\left|\mathbf{Z}=\mathbf{z}\right.\right]$$

$$
\begin{aligned}
&=\frac{1}{N_v}E\left[\left\|(-1)^{M_j}v\mathbf{z}-\lambda\frac{\mathbf{X}_j^T\mathbf{z}}{\|\mathbf{z}\|^2}\mathbf{z}\right\|^2\right]\\
&=\frac{1}{N_v}E\left[\left((-1)^{M_j}v\mathbf{z}-\lambda\frac{\mathbf{X}_j^T\mathbf{z}}{\|\mathbf{z}\|^2}\mathbf{z}\right)^T\left((-1)^{M_j}v\mathbf{z}-\lambda\frac{\mathbf{X}_j^T\mathbf{z}}{\|\mathbf{z}\|^2}\mathbf{z}\right)\right]\\
&=\frac{1}{N_v}E\left[\left((-1)^{M_j}v\mathbf{z}\right)^T(-1)^{M_j}v\mathbf{z}-\left((-1)^{M_j}v\mathbf{z}\right)^T\lambda\frac{\mathbf{X}_j^T\mathbf{z}}{\|\mathbf{z}\|^2}\mathbf{z}\right.\\
&\left.-\left(\lambda\frac{\mathbf{X}_j^T\mathbf{z}}{\|\mathbf{z}\|^2}\mathbf{z}\right)^T(-1)^{M_j}v\mathbf{z}+\left(\lambda\frac{\mathbf{X}_j^T\mathbf{z}}{\|\mathbf{z}\|^2}\mathbf{z}\right)^T\left(\lambda\frac{\mathbf{X}_j^T\mathbf{z}}{\|\mathbf{z}\|^2}\mathbf{z}\right)\right]\\
&=\frac{1}{N_v}\left(v^2\|\mathbf{z}\|^2-0-0+\frac{\lambda^2}{\|\mathbf{z}\|^2}E\left[\left(\mathbf{X}_j^T\mathbf{z}\right)^2\right]\right)\\
&=v^2\sigma_z^2-0-0+\frac{\lambda^2}{N_v\|\mathbf{z}\|^2}E\left[\left(\sum_{i=1}^{N_v}x_{j,i}z_i\right)^2\right]\\
&=v^2\sigma_z^2+\frac{\lambda^2}{N_v^2}\sum_{i=1}^{N_v}s_{j,i}^2\sigma_u^2
\end{aligned}
\tag{5-5}
$$

因此可以定义载体和水印的相对功率比为 DWR（Document to Watermark Ratio），为 $DWR=10\log_{10}\left(\frac{\sigma_x^2}{D_w}\right)$，其中 $\sigma_x^2=\frac{1}{N_v}\sum_{i=1}^{N_v}s_{j,i}^2\sigma_u^2$ 为载体平均功率。

5.2 ISS 水印安全的分析方法

由于所有的有关水印嵌入、检测等方法不能永久保密，水印的安全将取决于水印密钥的安全[18]。有关水印安全的讨论，就是研究密钥在水印通信中泄漏的信息量以及密钥在水印通信后的剩余不确定度。对于安全性高的水印方法，密钥在水印通信中信息泄露少，在水印通信后仍具有高的不确定度，即攻击者掌握了嵌入了水印的信号和相关信息后，仍很难确定使用的水印密钥。相反，对于安全性不高的水印方法，密钥在水印通信中信息泄露多，水印通信后密钥的不确定度低，攻击者将能够较准确地确定通信中使用的水印密钥。

在 ISS 水印安全性的分析中，密钥就是水印通信中的秘密载波[18, 67]。记 $h(\mathbf{Z})$

表示秘密载波的微分熵；使用$h\left(\mathbf{Z}\middle|\mathbf{Y}^{N_o}\right)$来表示在已知$N_o$次水印图像条件下密钥的剩余熵；将密钥和水印图像的互信息记为$I\left(\mathbf{Z},\mathbf{Y}^{N_o}\right)$。由互信息的定义可知

$$h\left(\mathbf{Z}\middle|\mathbf{Y}^{N_o}\right)=h(\mathbf{Z})-I\left(\mathbf{Z},\mathbf{Y}^{N_o}\right) \tag{5-6}$$

为了求出$I\left(\mathbf{Z},\mathbf{Y}^{N_o}\right)$，再利用一次互信息的定义，可得

$$I\left(\mathbf{Z},\mathbf{Y}^{N_o}\right)=h\left(\mathbf{Y}^{N_o}\right)-h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z}\right) \tag{5-7}$$

其中$h\left(\mathbf{Y}^{N_o}\right)$是$N_o$次观测水印图像的微分熵，$h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z}\right)$是在载波已知条件下$N_o$次观测水印图像的条件熵。易知，$h\left(\mathbf{Z}\middle|\mathbf{Y}^{N_o}\right)$表示了密钥在水印通信后的不确定性，而$I\left(\mathbf{Z},\mathbf{Y}^{N_o}\right)$表示了密钥在水印通信过程中泄漏的信息。$h\left(\mathbf{Z}\middle|\mathbf{Y}^{N_o}\right)$越大、$I\left(\mathbf{Z},\mathbf{Y}^{N_o}\right)$越小表示水印方法的安全性越高。

5.3 ISS 水印安全性分析

根据水印算法所面临攻击的条件不同，关于水印的安全性的分析主要在KMA、KOA（Known Original Attack）和WOA三种情况下进行[46,48]。KMA是指攻击者掌握了水印图像和相应的秘密信息本身，通过多次观察来估计水印密钥；KOA是指攻击者掌握水印图像和对应的未嵌入水印的信号载体；WOA是最为困难的一种情况，即攻击者只拥有水印图像。

本文利用GSM模型描述自然图像载体小波系数的统计分布，从信息论的角度来分析ISS水印的安全性能。由于在KOA情况下攻击者已经掌握了载体图像，利用GSM模型不会带来额外的帮助，因此本文的讨论仅从KMA和WOA情况下展开。

5.3.1 KMA 条件

在KMA条件下，攻击者掌握了对水印图像的N_o次观测以及对应的嵌入信息。由于用GSM模型描述图像载体时，尺度因子$\mathbf{S}$可以由观测量估计得到，因此（5-6）和（5-7）分别具体化为（5-8）和（5-9）。

$$h\left(\mathbf{Z}\middle|\mathbf{Y}^{N_o},\mathbf{S}^{N_o},M^{N_o}\right)=h\left(\mathbf{Z}\middle|\mathbf{S}^{N_o},M^{N_o}\right)-I\left(\mathbf{Z},\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right) \tag{5-8}$$

$$I\left(\mathbf{Z},\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)=h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)-h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},\mathbf{S}^{N_o},M^{N_o}\right) \tag{5-9}$$

考虑到每次嵌入的信息与秘密载波和图像载体相独立，（5-8）可具体化为

$$h\left(\mathbf{Z}\middle|\mathbf{Y}^{N_o},\mathbf{S}^{N_o},M^{N_o}\right)=h(\mathbf{Z})-I\left(\mathbf{Z},\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right) \tag{5-10}$$

为了求得 $h\left(\mathbf{Z}\middle|\mathbf{Y}^{N_o},\mathbf{S}^{N_o},M^{N_o}\right)$，必须求出 $h(\mathbf{Z})$、$h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)$ 和 $h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},\mathbf{S}^{N_o},M^{N_o}\right)$。根据假设条件，秘密载波的微分熵 $h(\mathbf{Z})$ 容易计算，为

$$h(\mathbf{Z})=\frac{N_v}{2}\log\left(2\pi e\sigma_z^2\right) \tag{5-11}$$

其中 $\log(\cdot)$ 为自然对数。以下分别计算 $h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)$ 和 $h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},\mathbf{S}^{N_o},M^{N_o}\right)$。

(1) $h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},\mathbf{S}^{N_o},M^{N_o}\right)$ 的计算

由于假设各图像载体之间相互独立，在已知秘密载波和每次观测对应的秘密信息条件下，各次观测之间也是相互独立，因此 $h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},\mathbf{S}^{N_o},M^{N_o}\right)$ 可以分解为各次观测对应的条件熵之和。所以有：

$$\begin{aligned}&h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},\mathbf{S}^{N_o},M^{N_o}\right)\\&=\sum_{j=1}^{N_o}h\left(\mathbf{Y}_j\middle|\mathbf{Y}_{j-1},\mathbf{Y}_{j-2},\cdots,\mathbf{Y}_1,\mathbf{Z},\mathbf{S}_j,M_j\right)\\&=\sum_{j=1}^{N_o}h\left(\mathbf{Y}_j\middle|\mathbf{Z},\mathbf{S}_j,M_j\right)\end{aligned} \tag{5-12}$$

对于每一次观测，由公式（5-2）可知水印图像由载体信号、受嵌入信息调制的秘密载波（强度受 v 控制）和载体信号在秘密载波方向的投影（强度受 λ 控制）三部分构成。对于载体信号在秘密载波方向上的投影，可以利用投影矩阵来表示[106]，即：

$$\lambda\frac{\mathbf{X}^T\mathbf{Z}}{\|\mathbf{Z}\|^2}\mathbf{Z}=\lambda\mathrm{P}_v\mathbf{X} \tag{5-13}$$

其中 $\mathrm{P}_v\triangleq\frac{\mathbf{Z}\mathbf{Z}^T}{\|\mathbf{Z}\|^2}$ 是向秘密载波 $\mathbf{Z}$ 方向的投影矩阵，如图 5-1 所示。

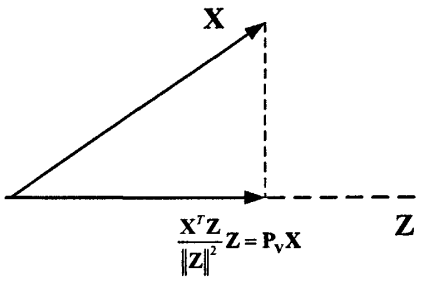


图 5-1 载体矢量向秘密载波方向的投影

Fig. 5-1 Projection of host on the secret carrier

根据投影矩阵构成特点，易知$\mathbf{P}_v$具有以下性质[106]：

（i）$\mathbf{P}_v$是对称矩阵；

证明：由$\mathbf{ZZ}^T$为对称阵，$\|\mathbf{Z}\|^2$为标量可知。

（ii）$\mathbf{P}_v$具有幂等性，即$\mathbf{P}_v^2=\mathbf{P}_v$；

证明：易知$\mathbf{P}_v^2=\frac{\mathbf{ZZ}^T}{\|\mathbf{Z}\|^2}\frac{\mathbf{ZZ}^T}{\|\mathbf{Z}\|^2}=\frac{\mathbf{Z}\|\mathbf{Z}\|^2\mathbf{Z}^T}{\|\mathbf{Z}\|^4}=\frac{\mathbf{ZZ}^T}{\|\mathbf{Z}\|^2}=\mathbf{P}_v$

（iii）$\mathbf{P}_v$的特征值只可取 1 和 0 两个值；

证明：由$\mathbf{P}_v^2=\mathbf{P}_v$可知，对于$\mathbf{P}_v$的特征方程，有$\mathbf{P}_v\mathbf{u}=\alpha\mathbf{u}$和$\mathbf{P}_v^2\mathbf{u}=\alpha\mathbf{u}$，其中$\mathbf{u}$为$\mathbf{P}_v$的特征向量，$\alpha$为$\mathbf{u}$对应的特征值。因此可得以下等价关系：$\alpha\mathbf{u}=\mathbf{P}_v^2\mathbf{u}=\mathbf{P}_v\mathbf{P}_v\mathbf{u}=\mathbf{P}_v\alpha\mathbf{u}=\alpha\mathbf{P}_v\mathbf{u}=\alpha^2\mathbf{u}\quad\forall\mathbf{u}\neq 0$。所以$\alpha=1$或$\alpha=0$。

（iv）$\mathbf{P}_v$的秩与迹相等，即$rank(\mathbf{P}_v)=tr(\mathbf{P}_v)$，其中$rank(\bullet)$和$tr(\bullet)$分别表示求矩阵的秩和迹。

证明：由于幂等阵$\mathbf{P}_v$的特征值只取 0 和 1（见$\mathbf{P}_v$的性质 iii），而且任何一个矩阵的秩等于其非零特征值的个数，所以$\mathbf{P}_v$的秩等于其值为 1 的特征值的个数。另外，根据$n\times n$矩阵的迹与特征值的关系，有$tr(\mathbf{P}_v)=\sum_{i=1}^{n}\alpha_i$，可知$\mathbf{P}_v$的迹等于特征值的和。因此幂等矩阵的迹等于其值为 1 的特征值的个数。所以，幂等矩阵$\mathbf{P}_v$的秩与迹相等。

对于 ISS 水印方法中 $\mathrm{P_v}$ 的定义，由 $\|\mathbf{Z}\|^2 \triangleq \sum_{i=1}^{N_v} Z_i^2 = tr\left(\mathbf{Z}\mathbf{Z}^T\right)$ 可知 $\mathrm{P_v}$ 的迹为：

$$tr\left(\mathrm{P_v}\right) = tr\left(\frac{\mathbf{Z}\mathbf{Z}^T}{\|\mathbf{Z}\|^2}\right) = \frac{1}{\|\mathbf{Z}\|^2} tr\left(\mathbf{Z}\mathbf{Z}^T\right) = 1 \tag{5-14}$$

因此 $\mathrm{P_v}$ 的秩也为 1，$\mathrm{P_v}$ 只有一个为 1 的特征值。对于 $N_v > 1$，$\left|\mathrm{P_v}\right| = 0$；对于 $N_v = 1$，$\left|\mathrm{P_v}\right| = 1$。

根据 ISS 的嵌入方程（5-2）和 GSM 模型的特点，在已知 GSM 模型尺度参数的条件下，载体信号服从高斯分布；$\frac{\mathbf{X}_j^T\mathbf{Z}}{\|\mathbf{Z}\|^2}\mathbf{Z}$ 表示高斯矢量在秘密载波方向上的投影，仍然服从高斯分布。所以，在已知秘密载波和观测信号对应的嵌入信息条件下，第 j 次观测信号服从以 $(-1)^{M_j} v\mathbf{Z}$ 为均值的高斯分布，其协方差矩阵可由式（5-15）定义。

$$\begin{aligned}
\Sigma_j\Big|_{\mathbf{Z}=\mathbf{z},\mathbf{S}=\mathbf{s}_j,\mathbf{M}=m_j} &\triangleq E\left[\left(\mathbf{Y}_j - E\left(\mathbf{Y}_j\right)\right)\left(\mathbf{Y}_j - E\left(\mathbf{Y}_j\right)\right)^T \middle| \mathbf{z},\mathbf{s}_j,m_j\right] \\
&= E\left[\left(\mathbf{X}_j - \lambda\frac{\mathbf{X}_j^T\mathbf{z}}{\|\mathbf{z}\|^2}\mathbf{z}\right)\left(\mathbf{X}_j - \lambda\frac{\mathbf{X}_j^T\mathbf{z}}{\|\mathbf{z}\|^2}\mathbf{z}\right)^T\right] \\
&= E\left[\left(\mathbf{X}_j - \lambda\mathrm{P_v}\mathbf{z}\right)\left(\mathbf{X}_j - \lambda\mathrm{P_v}\mathbf{z}\right)^T\right]
\end{aligned} \tag{5-15}$$

以上协方差矩阵的行列式可以求出，为

$$\begin{aligned}
&\left|\Sigma_j\right|_{\mathbf{Z}=\mathbf{z},\mathbf{S}=\mathbf{s}_j,\mathbf{M}=m_j} \\
&= \left|E\left[\left(\mathbf{X}_j - \lambda\mathrm{P_v}\mathbf{X}_j\right)\left(\mathbf{X}_j - \lambda\mathrm{P_v}\mathbf{X}_j\right)^T \middle| \mathbf{z},\mathbf{s}_j,m_j\right]\right| \\
&= \left|E\left[\left(\mathbf{X}_j - \lambda\mathrm{P_v}\mathbf{X}_j\right)\left(\mathbf{X}_j^T - \lambda\mathbf{X}_j^T\mathrm{P_v}^T\right) \middle| \mathbf{z},\mathbf{s}_j,m_j\right]\right| \\
&= \left|E\left[\left(\mathbf{I} - \lambda\mathrm{P_v}\right)\mathbf{X}_j\mathbf{X}_j^T\left(\mathbf{I} - \lambda\mathrm{P_v}^T\right) \middle| \mathbf{z},\mathbf{s}_j,m_j\right]\right| \\
&= \left|\left[\left(\mathbf{I} - \lambda\mathrm{P_v}\right)E\left(\mathbf{X}_j\mathbf{X}_j^T\right)\left(\mathbf{I} - \lambda\mathrm{P_v}^T\right) \middle| \mathbf{z},\mathbf{s}_j,m_j\right]\right| \\
&= \left[\left|\mathbf{C}_{\mathbf{X}_j}\right|\left|\left(\mathbf{I} - \lambda\mathrm{P_v}\right)\right|\left|\left(\mathbf{I} - \lambda\mathrm{P_v}^T\right)\right| \middle| \mathbf{z},\mathbf{s}_j,m_j\right] \\
&= \left[\left|\mathbf{C}_{\mathbf{X}_j}\right|\left|\left(\mathbf{U}\mathbf{U}^T - \lambda\mathbf{U}\Gamma_v\mathbf{U}^T\right)\right|\left|\left(\mathbf{U}^T\mathbf{U} - \lambda\mathbf{U}^T\Gamma_v\mathbf{U}\right)\right| \middle| \mathbf{z},\mathbf{s}_j,m_j\right]
\end{aligned}$$

$$
\begin{aligned}
&=\left[\left|\mathbf{C}_{\mathbf{x}_j}\right|\left\|\mathbf{U}\right\|\left(\mathbf{I}-\lambda\Gamma_v\right)\left\|\mathbf{U}^T\right\|\left\|\mathbf{U}^T\right\|\left(\mathbf{I}-\lambda\Gamma_v\right)\left\|\mathbf{U}\right\|\mathbf{z},\mathbf{s}_j,m_j\right]\\
&=\left[\left|\mathbf{C}_{\mathbf{x}_j}\right|\left\|\left(\mathbf{I}-\lambda\Gamma_v\right)\right|^2\mathbf{z},\mathbf{s}_j,m_j\right]\\
&=\left[\left|\mathbf{C}_{\mathbf{x}_j}\right|\left(1-\lambda\right)^2\mathbf{z},\mathbf{s}_j,m_j\right]
\end{aligned}
\tag{5-16}
$$

其中 $\Gamma_v=\begin{bmatrix}1&0&\cdots&0\\0&0&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&\cdots&0&0\end{bmatrix}$，为 $\mathbf{P}_v$ 的特征值矩阵，$\mathbf{U}$ 为对应于 Γ_v 的特征向量构成的酉矩阵。由于使用了标量 GSM 模型，$\left|\mathbf{C}_{\mathbf{x}_j}\right|_{\mathbf{s}=\mathbf{s}}=\sigma_u^{2N_v}\prod_{i=1}^{N_v}s_{j,i}^2$，所以

$$
\begin{aligned}
&h\left(\mathbf{Y}_j\middle|\mathbf{Z}=\mathbf{z},\mathbf{S}_j=\mathbf{s}_j,M_j=m_j\right)=\frac{1}{2}\log\left[\left(2\pi e\right)^{N_v}\left|\mathbf{C}_{\mathbf{x}_j}\right|\left(1-\lambda\right)^2\right]\\
&=\frac{1}{2}\log\left[\left(2\pi e\right)^{N_v}\left(1-\lambda\right)^2\sigma_u^{2N_v}\prod_{i=1}^{N_v}s_{j,i}^2\right]
\end{aligned}
\tag{5-17}
$$

观察到（5-17）式与 m_j 的取值无关，所以，

$$
\begin{aligned}
&h\left(\mathbf{Y}_j\middle|\mathbf{Z},\mathbf{S}_j,M_j\right)\\
&=\int_{-\infty}^{\infty}p\left(\mathbf{Z}\right)h\left(\mathbf{Y}_j\middle|\mathbf{Z}=\mathbf{z},\mathbf{S}_j=\mathbf{s}_j,M_j=m_j\right)d\mathbf{Z}\\
&=\int_{-\infty}^{\infty}p\left(\mathbf{Z}\right)\frac{1}{2}\log\left[\left(2\pi e\right)^{N_v}\left(1-\lambda\right)^2\prod_{i=1}^{N_v}s_i^2\sigma_u^2\right]d\mathbf{Z}\\
&=\frac{1}{2}\log\left[\left(2\pi e\right)^{N_v}\left(1-\lambda\right)^2\sigma_u^{2N_v}\prod_{i=1}^{N_v}s_{j,i}^2\right]
\end{aligned}
\tag{5-18}
$$

将（5-18）式带入（5-12）式可得：

$$
h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},\mathbf{S}^{N_o},M^{N_o}\right)=\frac{1}{2}\sum_{j=1}^{N_o}\log\left[\left(2\pi e\right)^{N_v}\left(1-\lambda\right)^2\sigma_u^{2N_v}\prod_{i=1}^{N_v}s_{j,i}^2\right]
\tag{5-19}
$$

(2) $h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)$ 的计算

根据 ISS 水印的嵌入方程（5-2）式，在已知载体信号的尺度参数和秘密信息时，第 j 次观测矢量是三个高斯随机矢量之和：载体信号、秘密载波、载体信号在秘密载波上的投影，其中载体信号在秘密载波上投影矢量的协方差矩阵与秘密载波相关[67]。因此嵌入了水印的信号仍然为零均值的高斯分布。

(i) 当$N_o=1$时

考虑在已知秘密信息和载体信号尺度参数的条件下，每一次观测矢量的各维之间相互独立。有：

$$h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)=h\left(\mathbf{Y}|\mathbf{S},M\right)=\sum_{i=1}^{N_v}E\left[h\left(Y_i|\mathbf{S}=\mathbf{s},M=m\right)\right] \tag{5-20}$$

其中Y_i表示观测水印图像的第i维（此时仅有一幅观测的水印图像）。为了求出$h\left(Y_i|\mathbf{S}=\mathbf{s},M=m\right)$，必须先得到$\left(Y_i|\mathbf{S}=\mathbf{s},M=m\right)$的方差。

$$\begin{aligned}
&E\left[Y_i^2|\mathbf{S}=\mathbf{s},M=m\right]=E\left[\left(X_i+v(-1)^m Z_i-\lambda\frac{\mathbf{X}^T\mathbf{Z}}{\|\mathbf{Z}\|^2}Z_i\right)^2\right]\\
&=E\left[X_i^2+2(-1)^m vZ_iX_i-2v(-1)^m\lambda\frac{\mathbf{X}^T\mathbf{Z}}{\|\mathbf{Z}\|^2}Z_i^2-2\lambda\frac{\mathbf{X}^T\mathbf{Z}}{\|\mathbf{Z}\|^2}Z_iX_i+\lambda^2\left(\frac{\mathbf{X}^T\mathbf{Z}}{\|\mathbf{Z}\|^2}Z_i\right)^2+v^2Z_i^2\right]\\
&=s_i^2\sigma_u^2-2\lambda s_i^2\sigma_u^2E\left[\frac{Z_i^2}{\|\mathbf{Z}\|^2}\right]+\lambda^2s_i^2\sigma_u^2E\left[\frac{Z_i^4}{\|\mathbf{Z}\|^4}\right]+\lambda^2\sigma_u^2E\left[\sum_{l=1,l\neq i}^{N_v}\frac{s_l^2Z_l^2Z_i^2}{\|\mathbf{Z}\|^4}\right]+v^2\sigma_Z^2
\end{aligned} \tag{5-21}$$

由于$\mathbf{Z}$为N_v维的高斯随机向量，则$\frac{\mathbf{Z}}{\|\mathbf{Z}\|}$为单位半径的$N_v$维球面上呈均匀分布的随机向量，$\frac{Z_i}{\|\mathbf{Z}\|}$则为$\frac{\mathbf{Z}}{\|\mathbf{Z}\|}$的一维边际分布，$E\left[\frac{Z_i^2}{\|\mathbf{Z}\|^2}\right]$、$E\left[\frac{Z_i^4}{\|\mathbf{Z}\|^4}\right]$和$E\left[\sum_{l\neq k}\frac{Z_l^2Z_i^2}{\|\mathbf{Z}\|^4}\right]$则分别为$\frac{Z_i}{\|\mathbf{Z}\|}$的二阶原点矩、四阶原点矩和四阶混合矩。这三个统计量的计算需要用到以下引理（5-1）至（5-4）[107]。

引理 5-1 设$\mathbf{u}^{(d)}$是 d 维球面上均匀分布的随机变量，则

$$E\left(\mathbf{u}^{(d)}\right)=0\qquad Cov\left(\mathbf{u}^{(d)}\right)=\frac{1}{d}\mathbf{I}_d$$

引理 5-2 设$\mathbf{u}^{(d)}$服从 d 维单位球面上均匀分布，则$\mathbf{u}^{(d)}$的边际分布存在，$\left(U_1,\cdots,U_k\right)$的边际密度为

$$\frac{\Gamma(d/2)}{\Gamma\left((d-k)/2\right)\pi^{k/2}}\left(1-\sum_{i=1}^{k}u_i^2\right)^{(d-k)/2-1}$$

其中，$\sum_{i=1}^{k} u_i^2 < 1,\ 1 \le k < d$。

引理 5-3 设 $\mathbf{u}^{(d)} = (U_1, \cdots, U_d)^T$ 服从 d 维单位球面上均匀分布，则 $(U_1^2, \cdots, U_k^2)^T$ 服从 Dirichlet 分布，即

$$(U_1^2, \cdots, U_k^2)^T \sim D_d(1/2, \cdots, 1/2; (d-k)/2),\ 0 < k < d$$

引理 5-4 设 $\mathbf{y} = (Y_1, \cdots, Y_{d-1})^T \sim D_{d-1}(\alpha_1, \cdots, \alpha_{d-1}; \alpha_d)$，则

$$E(Y_j) = \frac{\alpha_j}{\alpha}$$

$$Var(Y_j) = \frac{\alpha_j(\alpha - \alpha_j)}{\alpha^2(\alpha+1)}$$

$$Cov(Y_i, Y_j) = -\frac{\alpha_i \alpha_j}{\alpha^2(\alpha+1)}$$

其中 $\alpha = \sum_{i=1}^{d} \alpha_i$，$\alpha_i > 0$。

由于 $Z_i' = \frac{Z_i}{\|\mathbf{Z}\|}$ 的分布为 N_v 维单位球面上的均匀分布，根据引理 5-2，令 $d = N_v$，$k = 1$，则 Z_i' 的概率密度函数为

$$\frac{\Gamma(N_v/2)}{\Gamma((N_v-1)/2)\pi^{1/2}} (1 - u_i^2)^{(N_v-1)/2-1}$$

根据引理 5-3，令 $d = N_v$，$k = N_v - 1$，则 $(Z_1'^2, \cdots, Z_{N_v-1}'^2)^T \sim D_{N_v-1}(1/2, \cdots, 1/2; 1/2)$ 其中 $D_{N_v-1}(1/2, \cdots, 1/2; 1/2) = D_{N_v-1}(\alpha_1, \cdots, \alpha_{d-1}; \alpha_{N_v})$。再应用引理 5-4，可得：

$$E(Z_j'^2) = \frac{\alpha_j}{\alpha}$$

$$Var(Z_j'^2) = \frac{\alpha_j(\alpha - \alpha_j)}{\alpha^2(\alpha+1)}$$

$$Cov(Z_k'^2, Z_l'^2,) = -\frac{\alpha_k \alpha_l}{\alpha^2(\alpha+1)}$$

其中，$\alpha=\sum_{i=1}^{N_v}\alpha_i=\frac{N_v}{2}$。因此可得

$$E\left(Z_j'^2\right)=\frac{\alpha_j}{\alpha}=\frac{1/2}{N_v/2}=\frac{1}{N_v}$$

$$Var\left(Z_j'^2\right)=\frac{\alpha_j\left(\alpha-\alpha_j\right)}{\alpha^2\left(\alpha+1\right)}=\frac{1/2\left(N_v/2-1/2\right)}{N_v^2/4\left(N_v/2+1\right)}=\frac{2\left(N_v-1\right)}{N_v^2\left(N_v+2\right)}$$

$$Cov\left(Z_k'^2,Z_l'^2,\right)=-\frac{1/4}{N_v^2/4\left(N_v/2+1\right)}=-\frac{2}{N_v^2\left(N_v+2\right)}$$

考虑到 $Var\left[Z_j'^2\right]=E\left[Z_j'^4\right]-\left(E\left[Z_j'^2\right]\right)^2$，所以有：

$$E\left[Z_j'^4\right]=Var\left[Z_j'^2\right]+\left(E\left[Z_j'^2\right]\right)^2$$
$$=\frac{2\left(N_v-1\right)}{N_v^2\left(N_v+2\right)}+\left(\frac{1}{N_v}\right)^2=\frac{3}{N_v\left(N_v+2\right)}$$

由于 $Z_i'^2$ 与 $Z_l'^2$ 不相关，$Cov\left(Z_k'^2,Z_l'^2\right)=E\left[Z_i'^2,Z_l'^2\right]-E\left[Z_i'^2\right]E\left[Z_l'^2\right]$，则

$$E\left[Z_i'^2,Z_l'^2\right]=Cov\left(Z_i'^2,Z_l'^2\right)+E\left[Z_i'^2\right]E\left[Z_l'^2\right]$$
$$=-\frac{2}{N_v^2\left(N_v+2\right)}+\frac{1}{N_v}\cdot\frac{1}{N_v}$$
$$=\frac{1}{N_v\left(N_v+2\right)}$$

由以上 N_v 维球面均匀分布的随机向量的特性可计算出(5-21)式中各统计量，得到：

$$E\left[Y_i^2\left|\mathbf{S}=\mathbf{s},M=m\right.\right]$$
$$=s_i^2\sigma_u^2-2\lambda s_i^2\sigma_u^2E\left[\frac{Z_i^2}{\|\mathbf{Z}\|^2}\right]+\lambda^2s_i^2\sigma_u^2E\left[\frac{Z_i^4}{\|\mathbf{Z}\|^4}\right]+\lambda^2\sigma_u^2E\left[\sum_{l=1,l\neq i}^{N_v}\frac{s_l^2Z_l^2Z_k^2}{\|\mathbf{Z}\|^4}\right]+v^2\sigma_Z^2$$
$$=s_i^2\sigma_u^2-2\lambda s_i^2\sigma_u^2\frac{1}{N_v}+\lambda^2s_i^2\sigma_u^2\frac{3}{N_v\left(N_v+2\right)}+\lambda^2\sigma_u^2\sum_{l=1,l\neq i}^{N_v}\frac{s_l^2}{N_v\left(N_v+2\right)}+v^2\sigma_Z^2$$
$$=s_i^2\sigma_u^2+\frac{\sigma_u^2}{N_v\left(N_v+2\right)}\left[-2\lambda s_i^2\left(N_v+2\right)+2\lambda^2s_i^2+\lambda^2\sum_{l=1}^{N_v}s_l^2\right]+v^2\sigma_Z^2 \qquad (5\text{-}22)$$

因为假设嵌入信息等概率出现，而且（5-22）式的结果与 M 的取值无关，

所以在 KMA 条件下，一次水印图像观测的微分熵为：

$$h(\mathbf{Y}|\mathbf{S},M)=E\left[h(\mathbf{Y}|\mathbf{S}=\mathbf{s},M=m)\right]$$
$$=\frac{1}{2}\sum_{i=1}^{N_v}\log 2\pi e\left[s_i^2\sigma_u^2+\frac{\sigma_u^2}{N_v(N_v+2)}\left(-2\lambda s_i^2(N_v+2)+2\lambda^2 s_i^2+\lambda^2\sum_{l=1}^{N_v}s_l^2\right)+v^2\sigma_Z^2\right] \quad (5\text{-}23)$$

(ii) 当 $N_o>1$ 时

多次观测（$N_o>1$）条件下，由于各次观测中嵌入的载波相同，有[41]：

$$h\left(\mathbf{Y}^{N_o}\left|\mathbf{S}^{N_o},M^{N_o}\right.\right)$$
$$\le\sum_{i=1}^{N_v}E\left[h\left(Y_{1,i},Y_{2,i},\cdots Y_{N_o,i}\left|S_{1,i}=s_{1,i},\cdots,S_{N_o,i}=s_{N_o,i},M_{1,i}=m_{1,i},\cdots,M_{N_o,i}=m_{N_o,i}\right.\right)\right]$$
$$\le\frac{1}{2}\sum_{i=1}^{N_v}E\left(\log\left((2\pi e)^{N_o}\left|\Sigma_{\mathbf{Y}_i}\right|\right)\right) \quad (5\text{-}24)$$

其中 $Y_{1,i},Y_{2,i},\cdots Y_{N_o,i}$ 为各次观测的水印图像中第 i 维的值，$\Sigma_{\mathbf{Y}_i}$ 为 $\left(Y_{1,i},Y_{2,i},\cdots Y_{N_o,i}\right)$ 的协方差矩阵，$|\bullet|$ 表示求行列式。$\Sigma_{\mathbf{Y}_i}$ 的各个元素可由（5-25）计算。

$$\Sigma_{\mathbf{Y}_i}(j,k)=E\left[Y_{j,i}Y_{k,i}\left|S_{1,i}=s_{1,i},\cdots,S_{N_o,i}=s_{N_o,i},M_{1,i}=m_{1,i},\cdots,M_{N_o,i}=m_{N_o,i}\right.\right]$$
$$=\begin{cases}s_{j,i}^2\sigma_u^2+\dfrac{\sigma_u^2}{N_v(N_v+2)}\left[-2\lambda s_{j,i}^2(N_v+2)+2\lambda^2 s_{j,i}^2+\lambda^2\sum_{l=1}^{N_v}s_{j,l}^2\right]+v^2\sigma_Z^2 & j=k\\ (-1)^{m_j+m_k}v^2\sigma_Z^2 & j\ne k\end{cases} \quad (5\text{-}25)$$

为了描述方便，记

$$A_j=s_{j,i}^2\sigma_u^2+\frac{\sigma_u^2}{N_v(N_v+2)}\left[-2\lambda s_{j,i}^2(N_v+2)+2\lambda^2 s_{j,i}^2+\lambda^2\sum_{l=1}^{N_v}s_{j,l}^2\right],\quad B=v^2\sigma_Z^2$$

则有：

$$\Sigma_{\mathbf{Y}_i}=\begin{bmatrix}A_1+B & (-1)^{m_1+m_2}B & \cdots & (-1)^{m_1+m_{N_o}}B\\ (-1)^{m_2+m_1}B & A_2+B & \cdots & (-1)^{m_2+m_{N_o}}B\\ \vdots & \vdots & \ddots & \vdots\\ (-1)^{m_{N_o}+m_1}B & (-1)^{m_{N_o}+m_2}B & \cdots & A_{N_o}+B\end{bmatrix} \quad (5\text{-}26)$$

因此，可以经由以下步骤求得 $\Sigma_{\mathbf{Y}_i}$ 的行列式：

$$\left|\Sigma_{\mathbf{Y}_i}\right|=\begin{vmatrix} A_1+B & (-1)^{m_1+m_2}B & \cdots & (-1)^{m_1+m_{N_o}}B \\ (-1)^{m_2+m_1}B & A_2+B & \cdots & (-1)^{m_2+m_{N_o}}B \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{m_{N_o}+m_1}B & (-1)^{m_{N_o}+m_2}B & \cdots & A_{N_o}+B \end{vmatrix}$$

$$=(-1)^{m_1}(-1)^{m_2}\cdots(-1)^{m_{N_o}}\begin{vmatrix} (-1)^{-m_1}(A_1+B) & (-1)^{m_2}B & \cdots & (-1)^{m_{N_o}}B \\ (-1)^{m_1}B & (-1)^{-m_2}(A_2+B) & \cdots & (-1)^{m_{N_o}}B \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{m_1}B & (-1)^{m_2}B & \cdots & (-1)^{-m_{N_o}}\left(A_{N_o}+B\right) \end{vmatrix}$$

$$=(-1)^{\sum_{i=1}^{N_o}m_i}\begin{vmatrix} (-1)^{m_1}(A_1+B) & (-1)^{m_2}B & \cdots & (-1)^{m_{N_o}}B \\ (-1)^{m_1}B & (-1)^{m_2}(A_2+B) & \cdots & (-1)^{m_{N_o}}B \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{m_1}B & (-1)^{m_2}B & \cdots & (-1)^{m_{N_o}}\left(A_{N_o}+B\right) \end{vmatrix}$$

$$=(-1)^{\sum_{i=1}^{N_o}m_i}(-1)^{\sum_{i=1}^{N_o}m_i}\begin{vmatrix} (A_1+B) & B & \cdots & B \\ B & (A_2+B) & \cdots & B \\ \vdots & \vdots & \ddots & \vdots \\ B & B & \cdots & \left(A_{N_o}+B\right) \end{vmatrix} \tag{5-27}$$

考虑到$(-1)^{\sum_{i=1}^{N_o}m_i}(-1)^{\sum_{i=1}^{N_o}m_i}=1$，所以有：

$$\left|\Sigma_{\mathbf{Y}_i}\right|=\begin{vmatrix} (A_1+B) & B & \cdots & B \\ B & (A_2+B) & \cdots & B \\ \vdots & \vdots & \ddots & \vdots \\ B & B & \cdots & \left(A_{N_o}+B\right) \end{vmatrix}$$

$$=\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & A_1+B & B & \cdots & B \\ 0 & B & A_2+B & \cdots & B \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & B & B & \cdots & A_{N_o}+B \end{vmatrix}$$

$$=\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ -B & A_1 & 0 & \cdots & 0 \\ -B & 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -B & 0 & 0 & \cdots & A_{N_o} \end{vmatrix}$$

$$=\begin{vmatrix} 1+\sum_{j=1}^{N_o}\frac{B}{A_j} & 1 & 1 & \cdots & 1 \\ 0 & A_1 & 0 & \cdots & 0 \\ 0 & 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_{N_o} \end{vmatrix}=\left(1+\sum_{j=1}^{N_o}\frac{B}{A_j}\right)\prod_{j=1}^{N_o}A_j \tag{5-28}$$

即可得：

$$\left|\Sigma_{\mathbf{Y}_i}\right|=\left(1+\sum_{j=1}^{N_o}\frac{B}{A_j}\right)\prod_{j=1}^{N_o}A_j$$

$$=\left(1+\sum_{j=1}^{N_o}\frac{v^2\sigma_Z^2}{s_{j,i}^2\sigma_u^2+\frac{\sigma_u^2}{N_v(N_v+2)}\left[-2\lambda s_{j,i}^2(N_v+2)+2\lambda^2 s_{j,i}^2+\lambda^2\sum_{l=1}^{N_v}s_{j,l}^2\right]}\right)$$

$$\cdot\prod_{j=1}^{N_o}\left(s_{j,i}^2\sigma_u^2+\frac{\sigma_u^2}{N_v(N_v+2)}\left[-2\lambda s_{j,i}^2(N_v+2)+2\lambda^2 s_{j,i}^2+\lambda^2\sum_{l=1}^{N_v}s_{j,l}^2\right]\right) \tag{5-29}$$

将（5-29）式代入（5-24）可得

$$h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)$$

$$\le\frac{1}{2}\sum_{i=1}^{N_v}\log\left((2\pi e)^{N_o}\left(1+\sum_{j=1}^{N_o}\frac{v^2\sigma_Z^2}{s_{j,i}^2\sigma_u^2+\frac{\sigma_u^2}{N_v(N_v+2)}\left[-2\lambda s_{j,i}^2(N_v+2)+2\lambda^2 s_{j,i}^2+\lambda^2\sum_{l=1}^{N_v}s_{j,l}^2\right]}\right.\right. \tag{5-30}$$

$$\left.\left.\cdot\prod_{j=1}^{N_o}\left(s_{j,i}^2\sigma_u^2+\frac{\sigma_u^2}{N_v(N_v+2)}\left[-2\lambda s_{j,i}^2(N_v+2)+2\lambda^2 s_{j,i}^2+\lambda^2\sum_{l=1}^{N_v}s_{j,l}^2\right]\right)\right)\right)$$

(3) $I\left(\mathbf{Z},\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)$

将本节第(1)和第(2)部分计算的结果带入（5-9）可得 KMA 条件下秘密载波

和观测量之间互信息（信息泄漏）。

当仅考虑一次观察时$N_o=1$，由（5-18）和（5-23）可得：

$$
\begin{aligned}
&I(\mathbf{Y};\mathbf{Z}|\mathbf{S},M)\\
&=\frac{1}{2}\sum_{i=1}^{N_v}\log 2\pi e\left[s_i^2\sigma_u^2+\frac{\sigma_u^2}{N_v(N_v+2)}\left(-2\lambda s_i^2(N_v+2)+2\lambda^2 s_i^2+\lambda^2\sum_{l=1}^{N_v}s_l^2\right)+v^2\sigma_Z^2\right]\\
&-\frac{1}{2}\log\left[(2\pi e)^{N_v}(1-\lambda)^2\sigma_u^{2N_v}\prod_{i=1}^{N_v}s_i^2\right]\\
&=\frac{N_v}{2}\log 2\pi e+\frac{1}{2}\sum_{i=1}^{N_v}\log\left[s_i^2\sigma_u^2+\frac{\sigma_u^2}{N_v(N_v+2)}\left(-2\lambda s_i^2(N_v+2)+2\lambda^2 s_i^2+\lambda^2\sum_{l=1}^{N_v}s_l^2\right)+v^2\sigma_Z^2\right]\\
&-\frac{N_v}{2}\log 2\pi e-\frac{1}{2}\log\left((1-\lambda)^2\sigma_u^{2N_v}\prod_{i=1}^{N_v}s_i^2\right)\\
&=\frac{1}{2}\sum_{i=1}^{N_v}\log\left[s_i^2\sigma_u^2+\frac{\sigma_u^2}{N_v(N_v+2)}\left(-2\lambda s_i^2(N_v+2)+2\lambda^2 s_i^2+\lambda^2\sum_{l=1}^{N_v}s_l^2\right)+v^2\sigma_Z^2\right]\\
&-\frac{1}{2}\log\left((1-\lambda)^2\sigma_u^{2N_v}\prod_{i=1}^{N_v}s_i^2\right)
\end{aligned}
\tag{5-31}
$$

当$N_o>1$时，由（5-19）和（5-30）可得：

$$
\begin{aligned}
&I\left(\mathbf{Z},\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)=h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)-h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},\mathbf{S}^{N_o},M^{N_o}\right)\\
&\le\frac{1}{2}\sum_{i=1}^{N_v}\log\left((2\pi e)^{N_o}\left(1+\sum_{j=1}^{N_o}\frac{v^2\sigma_Z^2}{s_{j,i}^2\sigma_u^2+\frac{\sigma_u^2}{N_v(N_v+2)}\left[-2\lambda s_{j,i}^2(N_v+2)+2\lambda^2 s_{j,i}^2+\lambda^2\sum_{l=1}^{N_v}s_{j,l}^2\right]}\right)\right.\\
&\left.\cdot\prod_{j=1}^{N_o}\left(s_{j,i}^2\sigma_u^2+\frac{\sigma_u^2}{N_v(N_v+2)}\left[-2\lambda s_{j,i}^2(N_v+2)+2\lambda^2 s_{j,i}^2+\lambda^2\sum_{l=1}^{N_v}s_{j,l}^2\right]\right)\right)\\
&-\frac{1}{2}\sum_{j=1}^{N_o}\log\left[(2\pi e)^{N_v}(1-\lambda)^2\sigma_u^{2N_v}\prod_{i=1}^{N_v}s_{j,i}^2\right]
\end{aligned}
\tag{5-32}
$$

5.3.2 WOA 条件

WOA 条件下，攻击者仅掌握对水印图像的N_o次观测。利用 GSM 模型来描述载体图像的统计分布，在 WOA 攻击下水印密钥的剩余熵可以表示为

$$
h\left(\mathbf{Z}\middle|\mathbf{S}^{N_o},\mathbf{Y}^{N_o}\right)=h\left(\mathbf{Z}\middle|\mathbf{S}^{N_o}\right)-I\left(\mathbf{Z},\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o}\right)=h(\mathbf{Z})-I\left(\mathbf{Z},\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o}\right)
\tag{5-33}
$$

其中，

$$I\left(\mathbf{Z},\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o}\right)=h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o}\right)-h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},\mathbf{Z}\right) \tag{5-34}$$

当$N_o=1$，即攻击者只掌握一次观测信息时，考虑到观测矢量各维之间的相互独立，秘密载波和观测水印图像之间的互信息可以表示为：

$$\begin{aligned}
&I\left(\mathbf{Z};\mathbf{Y}|\mathbf{S}\right)=h\left(\mathbf{Y}|\mathbf{S}\right)-h\left(\mathbf{Y}|\mathbf{S},\mathbf{Z}\right)\\
&=h\left(\mathbf{Y}|\mathbf{S},M=0\right)-h\left(\mathbf{Y}|\mathbf{S},\mathbf{Z}\right)\\
&=h\left(\mathbf{Y}|\mathbf{S},M=0\right)-h\left(\mathbf{Y}|\mathbf{S},\mathbf{Z},M\right)-I\left(\mathbf{Y};M|\mathbf{S},\mathbf{Z}\right)\\
&\geq h\left(\mathbf{Y}|\mathbf{S},M=0\right)-h\left(\mathbf{Y}|\mathbf{S},\mathbf{Z},M\right)-\log\left(2\right)
\end{aligned} \tag{5-35}$$

以上推导中利用了$h\left(Y|\mathbf{S}\right)=h\left(Y|\mathbf{S},M=0\right)$的条件。这是因为假设嵌入的秘密信息$M$取0和1的概率相同；而且在GSM模型中，$\mathbf{X}$的每一维在已知尺度因子条件下服从$X_i\sim N\left(0,s_i^2\sigma_u^2\right)$分布，$\mathbf{Z}$的每一维服从$Z_i\sim N\left(0,\sigma_z^2\right)$。由于$Z_i$的分布关于原点具有对称性，故对于$M$的不同取值，$\mathbf{Y}$将有相同的分布形式，即有：$p\left(Y|\mathbf{S},M=1\right)=p\left(Y|\mathbf{S},M=0\right)$。因此

$$\begin{aligned}
&h\left(Y|\mathbf{S}\right)=-E\left[\log p\left(Y|\mathbf{S}\right)\right]\\
&=-E\left[\log\sum_{i=1}^{2}\left(p\left(M=m_i\right)p\left(Y|\mathbf{S},M=m_i\right)\right)\right]\\
&=-E\left[\log p\left(Y|\mathbf{S},M=0\right)\right]\\
&=h\left(Y|\mathbf{S},M=0\right)
\end{aligned} \tag{5-36}$$

由（5-34）及（5-35）式可知：

$$\begin{aligned}
&I\left(\mathbf{Z};\mathbf{Y}|\mathbf{S},M\right)\geq I\left(\mathbf{Z};\mathbf{Y}|\mathbf{S}\right)=I\left(\mathbf{Z};\mathbf{Y}|\mathbf{S},M\right)-I\left(\mathbf{Y};M|\mathbf{S},\mathbf{Z}\right)\\
&\geq I\left(\mathbf{Z};\mathbf{Y}|\mathbf{S},M\right)-\log\left(2\right)
\end{aligned} \tag{5-37}$$

由（5-37）式可以看出，WOA条件下关于秘密载波的信息泄露不大于KMA条件下的信息泄露，而最多比KMA条件下的信息泄露少log2比特。该log2比特是由于嵌入的秘密信息不确定造成的。

5.3.3 与文献[67]中结论的比较

文献[67]使用高斯模型来描述自然图像载体系数的分布特点，对ISS水印的

安全性进行了分析，得到了相应的结论。以下将本文工作的结果与文献[67]得到的结论进行比较。为了比较的公平，在 GMS 模型和高斯模型下，自然图像载体（即子带小波系数）的方差应相同，即如(5-38)式所示。其中 σ_x^2 为高斯模型中小波系数分布的方差。

$$\sum_{i=1}^{N_v} s_{j,i}^2 \sigma_u^2 = N_v \sigma_x^2 \tag{5-38}$$

(i) 水印嵌入功率的比较

本文在 GSM 模型下对 ISS 线性模型中第 j 次嵌入水印的平均功率进行了计算，如公式(5-5)所示。为了比较的方便，将其再次列出如下：

$$D_w = v^2 \sigma_z^2 + \frac{\lambda^2}{N_v^2} \sum_{i=1}^{N_v} s_{j,i}^2 \sigma_u^2$$

文献[67]中计算得到的ISS线性模型中第 j 次嵌入水印的平均功率用 $D_{w-Gauss}$ 表示如下：

$$D_{w-Gauss} = v^2 \sigma_z^2 + \frac{\lambda^2}{N_v} \sigma_x^2 \tag{5-39}$$

考虑到(5-38)式的条件，并将其代入公式(5-5)可得

$$D_w = v^2 \sigma_z^2 + \frac{\lambda^2}{N_v^2} N_v \sigma_x^2 = D_{w-Gauss}$$

因此，本文的讨论与文献[67]的讨论中，对应于第 j 次嵌入水印的平均功率相同。

(ii) $h\left(\mathbf{Y}^{N_o} \middle| \mathbf{Z}, \mathbf{S}^{N_o}, M^{N_o}\right)$ 和 $h\left(\mathbf{Y}^{N_o} \middle| \mathbf{Z}, M^{N_o}\right)$ 的比较

本文的分析中，在 KMA 条件下，令 $\mathbf{s}_j$ 为 GSM 模型中载体分布的尺度向量，得到在已知嵌入信息和秘密载波条件下第 j 次观测的水印图像的微分熵可以表示为(5-17)式。为方便比较，再次列出如下。

$$h\left(\mathbf{Y}_j \middle| \mathbf{Z}, \mathbf{S}_j = \mathbf{s}_j, M_j = m_j\right) = \frac{1}{2} \log\left[(2\pi e)^{N_v} \left|\mathbf{C}_{\mathbf{x}_j}\right| (1-\lambda)^2\right]$$

$$= \frac{1}{2} \log\left[(2\pi e)^{N_v} (1-\lambda)^2 \sigma_u^{2N_v} \prod_{i=1}^{N_v} s_{j,i}^2\right]$$

在文献[67]中，假设载体为高斯分布，得到的第 j 次观测的水印图像的微分熵记为 $h_{Gauss}\left(\mathbf{Y}_j\middle|\mathbf{Z},M_j=m_j\right)$，如(5-40)式所示。

$$h_{Gauss}\left(\mathbf{Y}_j\middle|\mathbf{Z},M_j=m_j\right)=\frac{1}{2}\log\left[\left(2\pi e\right)^{N_v}\left(1-\lambda\right)^2\sigma_x^{2N_v}\right] \tag{5-40}$$

将以上两式列出并比较，可以看出对于第 j 次观测，若 GSM 模型中的尺度参数均相同，即 $\sigma_u^2 s_{j,i}^2=\sigma_x^2\ \left(\forall i,j\right)$，则 GSM 模型退化为高斯模型。此时有 $\sigma_u^{2N_v}\prod_{i=1}^{N_v}s_{j,i}^2=\sigma_x^{2N_v}$。在这种情况下，GSM 模型和高斯模型下两种表示相同。考虑 KMA 条件下已知秘密载波和对应嵌入信息条件时 N_o 次观测的条件熵，文献[67]得出的结论记为 $h_{Gauss}\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},M^{N_o}\right)$ 如式(5-41)所示，与本文得到的结论 $h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},\mathbf{S}^{N_o},M^{N_o}\right)$ 相比较，可以看出，基于高斯模型的前者可由退化后的 GSM 模型下的结论予以表示。

$$\begin{aligned}h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},\mathbf{S}^{N_o},M^{N_o}\right)&=\frac{1}{2}\sum_{j=1}^{N_o}\log\left[\left(2\pi e\right)^{N_v}\left(1-\lambda\right)^2\sigma_u^{2N_v}\prod_{i=1}^{N_v}s_{j,i}^2\right]\\ h_{Gauss}\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},M^{N_o}\right)&=\frac{N_o}{2}\log\left[\left(2\pi e\right)^{N_v}\left(1-\lambda\right)^2\sigma_u^{2N_v}\right]\end{aligned} \tag{5-41}$$

(iii) $h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)$ 的比较

本文中对 $N_o=1$ 时已知嵌入信息条件下观测量的微分熵记进行了分析，得到 $h\left(\mathbf{Y}\middle|\mathbf{S},M\right)$ 由式(5-23)表示，其中 $\mathbf{S}$ 为 GSM 模型中的尺度参数。与此相对应，文献[67]中基于载体高斯模型得到的 $h_{Gauss}\left(\mathbf{Y}\middle|M\right)$ 可由表示为式(5-42)。

$$\begin{aligned}&h\left(\mathbf{Y}\middle|\mathbf{S},M\right)=E\left[h\left(\mathbf{Y}\middle|\mathbf{S}=\mathbf{s},M=m\right)\right]\\ &=\frac{1}{2}\sum_{i=1}^{N_v}\log 2\pi e\left[s_i^2\sigma_u^2+\frac{\sigma_u^2}{N_v\left(N_v+2\right)}\left(-2\lambda s_i^2\left(N_v+2\right)+2\lambda^2 s_i^2+\lambda^2\sum_{l=1}^{N_v}s_l^2\right)+v^2\sigma_Z^2\right]\\ &h_{Gauss}\left(\mathbf{Y}\middle|M\right)=E\left[h\left(\mathbf{Y}\middle|M=m\right)\right]\\ &=\frac{N_v}{2}\log\left(2\pi e\right)+\frac{N_v}{2}\log\left(\sigma_x^2+v^2\sigma_Z^2+\sigma_x^2\frac{\lambda\left(\lambda-2\right)}{N_v}\right)\end{aligned} \tag{5-42}$$

由于在 GSM 模型退化为高斯模型时，GSM 模型中的尺度参数均相同，考虑到$\sigma_u^2 s_{j,l}^2 = \sigma_x^2 \left(\forall i,j\right)$，有(5-43)式的关系。同样可以发现，(5-42)式的结论可以由(5-23)式予以表示。

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{N_v}\log 2\pi e\left[s_i^2\sigma_u^2+\frac{\sigma_u^2}{N_v\left(N_v+2\right)}\left(-2\lambda s_i^2\left(N_v+2\right)+2\lambda^2 s_i^2+\lambda^2\sum_{l=1}^{N_v}s_l^2\right)+v^2\sigma_Z^2\right]\\
&=\frac{1}{2}\sum_{i=1}^{N_v}\log 2\pi e\left[\sigma_x^2+\frac{\sigma_x^2}{N_v\left(N_v+2\right)}\left(-2\lambda\left(N_v+2\right)+2\lambda^2+N_v\lambda^2\right)+v^2\sigma_Z^2\right]\\
&=\frac{N_v}{2}\log 2\pi e+\frac{N_v}{2}\log\left(\sigma_x^2+v^2\sigma_Z^2+\sigma_x^2\frac{\lambda\left(\lambda-2\right)}{N_v}\right)
\end{aligned}\tag{5-43}$$

当$N_o>1$时，考虑到$\sigma_u^2 s_{j,l}^2=\sigma_x^2\left(\forall i,j\right)$的退化条件，记

$$\begin{aligned}
A_j&=s_{j,l}^2\sigma_u^2+\frac{\sigma_u^2}{N_v\left(N_v+2\right)}\left[-2\lambda s_{j,l}^2\left(N_v+2\right)+2\lambda^2 s_{j,l}^2+\lambda^2\sum_{l=1}^{N_v}s_{j,l}^2\right]\\
&=\sigma_x^2+\sigma_x^2\frac{\lambda\left(\lambda-2\right)}{N_v}\\
B&=v^2\sigma_Z^2
\end{aligned}$$

则N_o次观测中第i维构成的矢量$\left(Y_{1,i},Y_{2,i},\cdots Y_{N_o,i}\right)$的协方差矩阵行列式为(5-44)式所示。

$$\begin{aligned}
\left|\Sigma_{Y_i}\right|&=\left(1+\sum_{j=1}^{N_o}\frac{B}{A_j}\right)\prod_{j=1}^{N_o}A_j\\
&=\left(1+\sum_{j=1}^{N_o}\frac{v^2\sigma_Z^2}{\sigma_x^2+\sigma_x^2\dfrac{\lambda\left(\lambda-2\right)}{N_v}}\right)\cdot\prod_{j=1}^{N_o}\left(\sigma_x^2+\sigma_x^2\frac{\lambda\left(\lambda-2\right)}{N_v}\right)\\
&=\sigma_x^{2N_o}\left(1+\frac{N_o v^2\sigma_Z^2}{\sigma_x^2\left(1+\dfrac{\lambda\left(\lambda-2\right)}{N_v}\right)}\right)\cdot\left(1+\frac{\lambda\left(\lambda-2\right)}{N_v}\right)^{N_o}
\end{aligned}\tag{5-44}$$

由此得到在已知秘密信息条件下N_o次观测的条件熵的上界为式(5-45)所示。(5-45)式即为文献[67]在高斯模型下得到的结论。

$$h\left(\mathbf{Y}^{N_o}\middle|M^{N_o}\right)$$

$$\le\frac{N_v}{2}\log\left((2\pi e)^{N_o}\left(\sigma_x^2\right)^{N_o}\left(1+\frac{\lambda(\lambda-2)}{N_v}\right)^{N_o}\left(1+\frac{N_o v^2\sigma_z^2}{\sigma_x^2\left(1+\frac{\lambda(\lambda-2)}{N_v}\right)}\right)\right) \tag{5-45}$$

(iv) 关于$I\left(\mathbf{Y};\mathbf{Z}\middle|\mathbf{S},M\right)$

通过以上的比较，当 GSM 模型退化为高斯模型时，文献[67]中有关$h\left(\mathbf{Y}^{N_o}\middle|M^{N_o}\right)$和$h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},M^{N_o}\right)$的计算结果都可以由本文中的$h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)$和$h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)$予以表示。根据 Shannon 信息论中互信息的定义，可知 KMA 条件下秘密载波和观测量之间的互信息可以表示为以上二者之差，即

$$I\left(\mathbf{Z},\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)=h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)-h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},\mathbf{S}^{N_o},M^{N_o}\right)$$

因此，文献[67]的结论可以用退化条件下本文的结论予以表示。文献[67]讨论的情况可以看作是本文结论的一种特例。

5.4 仿真与讨论

本节根据以上理论分析的结果，在 GSM 模型下对 ISS 水印的安全性进行了仿真。根据分析部分的假设条件，秘密载波服从$\mathbf{Z}\sim N\left(0,\sigma_z^2 I_{N_v}\right)$的分布，在使用相同的秘密载波的条件下，水印通信中秘密载波信息泄露的多少决定了水印的安全程度。因此，本节以多幅具有不同纹理特征的自然图像（包括 aerial，baboon，barb，boat，f16，lena，peppers 和 sailboat）的小波系数作为载体，分别计算了在不同条件下有关秘密载波的信息泄露。实验中使用双正交 9/7 小波对自然图像进行 2 层分解，随机选取 HL2, LH2 和 HH2 的子带系数作为载体。

图 5-1 显示了秘密载波信息泄露与载体衰减因子之间的关系。实验中，令观测次数$N_o=1$，载波长度$N_v=512$，DWR=25dB，载波衰减因子λ的变化从 0 至 1。图中带圆圈的实线和带方框的虚线分别表示在 KMA 条件下，基于 GSM 模型和基于高斯模型描述载体分布的方法对应的秘密载波信息泄露；虚点线表示在 WOA 条件下，基于 GSM 模型的方法中秘密载波信息泄露的下界。从图中可以

看出，随载体衰减因子的增加，秘密载波的信息泄露随之增加。由于$\lambda=0$对应于没有对载体进行衰减，此时 ISS 水印退化为传统的加性扩频水印。相比于传统的加性扩频水印，ISS 水印具有更好的鲁棒性，但是从图 5-1 中看出鲁棒性的提高是以降低安全性为代价的。相比于高斯模型描述载体的方法，基于 GSM 模型的方法显示出 ISS 水印密钥具有更多的信息泄露。这是由于利用 GSM 模型能够更为准确地刻画自然图像载体的统计特性，从而使攻击者可以更多地排除图像载体对于估计秘密载波的干扰。

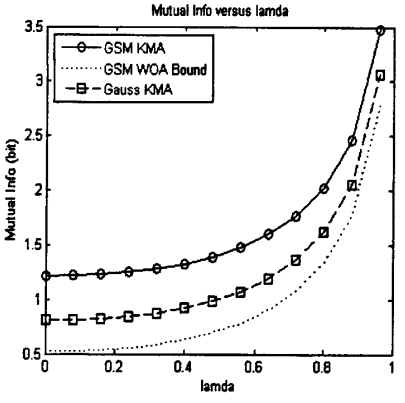


图 5-1 秘密载波信息泄露与载体衰减因子的关系

Fig. 5-1 Relationship between the information leakage of secret carrier and the host-rejection parameter

图 5-2 显示了在秘密载波长度固定为 512，载体衰减因子λ固定为 0.5，DWR 为 25dB 条件下，秘密载波信息泄露的上界与观测次数之间的关系。其中带圆圈的实线表示在基于 GSM 模型的方法中，KMA 条件下秘密载波的信息泄露上界；带方框的虚线表示基于高斯模型的分析中，KMA 条件下秘密载波的信息泄露上界。可以看出，随着观测次数的增加，关于秘密载波的信息泄露随之增加，攻击者可以通过对水印图像的多次观察积累对于秘密载波的知识。由于各次观测之间存在一定的相关性，秘密载波信息泄露的增长与观测次数的增长不呈线性关系。

图 5-3 比较了一次观测条件下，当 DWR 固定为 25dB，载体衰减因子λ固定为 0.5 时，基于 GSM 模型和高斯模型的方法在 KMA 条件下有关秘密载波的信息泄露与载波长度之间的关系。图中横轴表示秘密载波长度由 128 变化至 512，纵轴为秘密载波平均每一维上的信息泄露。由图示可以看出，随着长度的增长，秘密载波平均每一维的信息泄露随之减少，因此在 ISS 水印中使用更长的秘密载

波将能够提高水印的安全性。

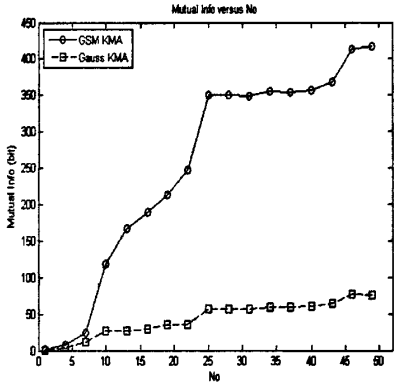


图 5-2 秘密载波信息泄露的上界与观测次数的关系

Fig. 5-2 Relationship between the upper bound of the information leakage of secret carrier and the number of observations

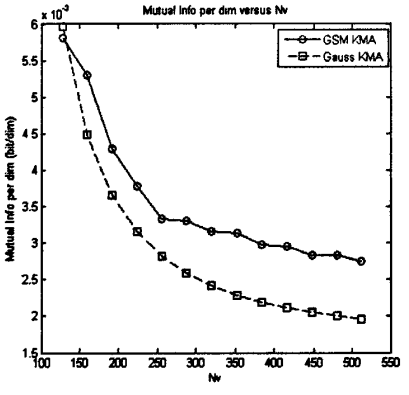


图 5-3 秘密载波信息泄露与载波长度的关系

Fig. 5-3 Relationship between the information leakage of secret carrier and the length of carrier

图 5-4 显示了秘密载波信息泄露与 DWR 变化之间的关系。实验中载波长度固定为 512，观测次数为 1 次，载体衰减因子 λ 为 0.5。图中带圆圈的实线和带方框的虚线分别表示在 KMA 条件下用基于 GSM 模型和基于高斯模型的分析方法得到的 ISS 水印密钥的信息泄露；图中虚线表示 WOA 条件下基于 GSM 模型的分析方法对应的秘密载波信息泄露的下界。可以看出随着 DWR 的增长，水印的能量相对于载体能量逐渐减弱，因此水印通信中泄露的关于秘密载波的信息随之减少，水印的安全性得以提高。

从以上仿真中可以看到，基于高斯模型的分析方法得到的关于秘密载波的信息泄露小于基于 GSM 模型分析方法得到的结果，因此前者的分析高估了 ISS 水

印的安全性。

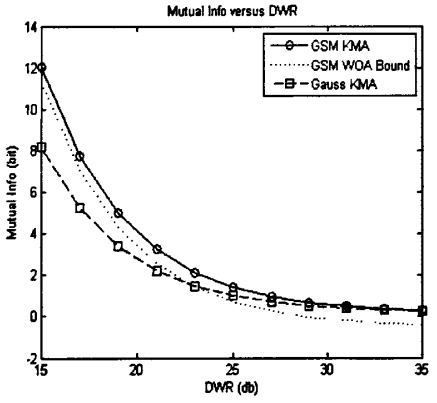


图 5-4 秘密载波信息泄露与 DWR 的关系

Fig. 5-4 Relationship between the information leakage of secret carrier and DWR

5.5 本章小结

高斯模型在协方差相同的条件下具有最大的不确定性[10]，因此不能准确捕捉自然图像载体统计分布的特殊性，导致在分析中高估了 ISS 水印的安全性。利用 GSM 模型可以准确刻画自然图像小波系数分布的特征，在一定程度上减小了对自然图像统计描述的不确定性，因此基于 GSM 模型对 ISS 水印安全性进行分析能够更充分地利用到自然图像载体的分布特点，从而得到更为贴近实际的分析结论。本文的工作利用 GSM 模型描述图像载体，从信息论的角度展开分析，得到了 ISS 水印安全与其影响因素之间的联系，并从理论上证明文献[67]中基于载体分布为高斯模型的相关分析结果可以被视为本文结论的特例。本文的工作为进一步设计新一代大容量、鲁棒和安全的水印提供了依据。

第六章 总结与展望

在社会生活高度信息化的今天，人们对于数字媒体的获取、制造、处理的能力空前提高。与此同时，人们对于信息安全问题的重视也与日俱增。在不断发展传统的信息加密技术以外，数字水印技术被广泛用于数字媒体知识产权保护、进行内容认证、媒体信息注释、设备使用控制、保密通信等领域。数字水印技术作为信号处理、信息论、统计学、计算机科学、密码学和人类视/听系统研究等学科交叉的研究方向，吸引了世界众多科研工作者的研究兴趣。

近年来，关于数字水印算法安全性的问题成为水印研究的重要方向。传统的水印研究主要关注的是水印的不可见性、水印容量和鲁棒性，而有关水印安全的研究则关注的是水印算法本身的安全问题。对于水印安全性的攻击是在Kerckhoffs准则下，即攻击者掌握了除水印密钥以外的所有关于水印的知识，通过对水印图像的多次观测，来估计水印密钥的行为。相比于对水印鲁棒性攻击的目的是为了增加检测水印信息的误码率，对水印安全进行攻击的目的则是试图掌握水印密钥并以此达到对水印算法的完全破解。因此，对水印安全性的研究尤为重要。今天，水印安全已经与水印容量、水印鲁棒性和水印不可见性一起成为数字水印研究的四大支柱。

6.1 本文工作总结

关于数字水印安全性的研究在最近几年展开。已有的研究首先着力于明确水印安全的相关定义，尤其对与水印鲁棒性的研究进行区分。在确立了水印安全性研究方向的基础上，这些工作对基于扩频序列的水印安全和基于量化索引调制的水印安全进行了分析。这些研究在取得巨大成果的同时，也暴露出一些问题。

在基于扩频序列的水印算法中，载体分布的特征对水印安全的影响至关重要。常用的图像水印系统多使用自然图像子带小波系数为载体。实际上，$N\times N$维的自然图像子带小波系数的分布只是所有可能分布的子集，有关图像模型的研究中也发现自然图像子带小波系数呈现很强的非高斯性。以往的研究在分析中假设图像载体为高斯分布，由于高斯分布在协方差相同条件下具有最大的不确定性，因此这些研究未能充分利用自然图像载体统计分布的特征，据此得到的水印

安全性能与实际情况有一定距离。针对以上问题，本文利用自然图像模型来刻画图像载体小波系数的分布特点，在此基础上研究了加性扩频水印的安全性和改进的扩频水印的安全性。本文的主要工作和结论总结如下。

（1）有关图像模型的研究发现，自然图像小波系数的分布可以由多种模型来描述，都能达到较为准确的刻画效果。相比之下，高斯尺度混合模型能够实现刻画准确性和数学可推导性的统一，因此非常适合于从统计和信息论的角度进行分析。本文的研究即采用高斯尺度混合模型来描述自然图像小波系数分布，结合具体的扩频水印算法模型实现了对水印安全的分析。

（2）本文利用基于 Fisher 信息的方法，使用 GSM 模型刻画自然图像小波系数的统计分布，对加性扩频水印系统的安全性进行了理论分析。在 KMA 和 WOA 情况下分别用对扩频载波（密钥）估计的 CRB 和 MCRB 来衡量加性水印系统的安全性，并得到加性扩频水印安全性与观测次数、扩频载波（密钥）长度、扩频载波嵌入能量以及嵌入信息分布特点之间的关系。本文的分析表明，以往的工作实际上高估了加性扩频水印系统的安全性。由于采用了 GSM 模型描述自然图像载体统计分布，使得本文分析的过程中能够充分利用到载体统计分布的非高斯性，得到的分析结论与实际的情况更为贴切。

（3）改进的扩频水印作为传统扩频水印算法的扩展，在近年来得到广泛的研究和使用。由于 ISS 水印算法中，嵌入的水印与载体相关，难以解析地得到观测量的似然函数。本文在使用 GSM 模型描述自然图像载体系数分布的基础上，利用 Shannon 互信息来衡量水印通信过程中关于水印密钥的信息泄露。在分析中利用投影矩阵来表示 ISS 水印的嵌入特点，并利用投影矩阵的性质简便地实现了安全性的相关分析，得到了 ISS 水印在 KMA 条件下和 WOA 条件下的安全性能与载波（密钥）长度、载体衰减因子、观测次数以及水印嵌入能量之间的关系。

与以往采用高斯模型描述载体分布的工作相比，本文的工作得到了对扩频水印系统安全性的更准确评价；本文的工作同时对设计新一代的安全、鲁棒水印算法具有重要的意义。

由于时间的限制，本文的工作还有很多可以改进和提高之处。第一，在自然图像模型的选择上，可以进一步使用更加精确的模型，例如文献[108]提出在 GSM 模型的基础上使用隐马尔可夫模型刻画子带间小波系数的关系来实现对自

然图像载体的更加精确描述。第二，在本文的研究中，假设秘密载波为独立同分布的高斯随机向量，尽管这是多数扩频水印算法中载波的生成办法，但是从研究的角度，可以进一步探讨呈现其它分布特点的秘密载波对于安全性的影响。第三，本文通过对加性扩频水印和 ISS 水印安全性的理论分析，定量地给出了它们的理论安全界限。由于时间的关系，本文没有研究目前最优的攻击算法与该理论界限之间的距离。对于水印安全的攻击实际上是在相关限制条件下对秘密载波信号进行估计的问题，值得进一步研究。第四，水印的安全性与鲁棒性、不可见性及水印容量是水印最为关键的四方面性能，水印算法中参数的选择涉及到对这四方面性能的联合优化。本文的工作已经给出了水印安全性和影响因素之间的定量关系，还需要通过适当的算法来求解以上的优化问题，从而设计实现新一代大容量、强鲁棒、高安全的不可见水印。

6.2 研究工作的展望

在数字水印的相关研究中，水印安全性是一个崭新的方向。在以往工作的基础上，可以在以下方面开展进一步的研究。

(1) 基于更加准确的自然图像模型研究水印安全

准确描述自然图像载体统计分布对于研究水印安全的意义已经多次说明。采用更加准确和易于实现信息论分析的统计模型将有助于进一步提高水印安全分析的准确程度。近年来关于自然图像模型的研究正不断取得进展，文献[108]在 GSM 模型的基础上，提出使用隐马尔可夫随机场描述不同尺度子带的小波系数之间的联系，实现了对于自然图像小波系数统计分布的更加准确的描述。由于描述图像载体分布性能的提高伴随着描述复杂性的增加，使得利用信息论或统计的方法分析水印安全性更加困难。因此，利用更加准确的自然图像模型的方向既是机遇，又是巨大的挑战。

(2) QIM 水印的安全性的分析

QIM 水印是一种广为应用的水印算法。目前有关 QIM 水印安全的分析仅见于文献[63]和[64]，其中假设图像载体系数的量化间隔远小于载体系数的方差

（Flat Assumption）。在此假设基础上，可以认为在每一个量化间隔之内载体系数呈均匀分布，因此每个经量化的载体系数对于密钥的信息泄露的影响都是相同的。据此得到关于 QIM 水印安全性与鲁棒性相独立。但是，实际上当量化间隔很小时，QIM 水印的鲁棒性很差，这样就影响了该水印方法的实际使用价值。因此，具有实际使用价值的 QIM 水印并不能符合以上的假设条件（Flat Assumption）。对于自然图像载体的小波系数，落入每一个量化间隔内的载体不能完全满足均匀分布。特别是由于绝对值在零附近的小波系数更多，而在绝对值大的区域内小波系数很少，每个量化间隔内的小波系数分布不同，每个量化区域对于密钥信息泄露的贡献也不同。因此，具有实际意义的 QIM 水印安全性与量化网格的设计和载体统计特性有关，也与该水印的鲁棒性能有关。

由于理论假设与实际情况不同而导致分析结论不同的例子很多，例如以往的研究在高斯模型基础上分析得到基于扩频的水印在加性强噪声信道中的性能要比基于 QIM 的水印方法更优[109]。文献[110]考虑图像载体实际的分布，对 DC-DM QIM 水印的性能进行了更加准确的分析，得出 DC-DM QIM 水印的可达速率并不逊于传统的扩频水印。结合图像载体特性对具有实际应用意义的 QIM 水印安全性进行分析也是一项值得研究的课题。

(3) 水印嵌入位置的估计

现有的关于水印安全的分析实际上都是建立在一个很强的假设之上，即攻击者知道水印嵌入的位置。尽管根据 Kerckhoffs 原理，攻击者应该能够掌握除了水印密钥以外的关于水印算法的所有知识，但是实际中水印嵌入位置与水印安全性之间的关系确值得深入研究。以往关于数字水印的研究表明，嵌入水印的位置可能会影响到水印的鲁棒性、不可见性等性能。那么嵌入水印位置的不同是否会影响到攻击者估计水印密钥的难易程度？目前学术界还没有在此方向得到一些较为满意的结论。尽管这个方向的研究将非常困难，但是非常有意义。

(4) 通用的水印安全性分析

目前对于水印安全性的研究都是针对具体的水印算法，例如对加性扩频水印、对改进的扩频水印以及对基于量化索引调制的水印安全性进行研究，得到的

结论也都是针对于具体的水印算法。是否存在一种更加一般或通用的水印安全分析？能否得到一些能适用于多种甚至所有的水印系统的安全性能的结论？这样的分析应该从信息论的角度出发，对于各种水印嵌入的特点予以概括并得到一些“普适”的结论。

(5) 估计算法

与水印安全研究相对应的就是对水印算法的攻击，也就是对于数字水印密钥的估计问题。尽管对于水印密钥的估计并不是本文所讨论的问题，但是有关水印密钥估计算法的研究是一个非常重要的研究方向。目前对水印密钥的估计算法主要有基于最大似然估计的算法和基于独立分量分析的估计算法，分别应用于KMA 和 WOA 条件下。由于关于水印安全性能分析的结论多是给出水印算法安全性能的界，并没有提出使用怎样具体的估计算法，因此研究更好的估计算法来逼近理论估计的界将具有重要的意义。

(6) 水印性能的联合优化

水印安全性已经与水印容量、水印不可见性以及水印鲁棒性共同成为水印研究的四大支柱。新一代水印算法的设计需要同时考虑到水印的以上性能。在加性扩频水印算法和改进的扩频水印算法中，分析得出水印的安全性能与水印嵌入的参数有关，例如扩频载波（密钥）的长度、水印嵌入能量、嵌入衰减因子等等。同时，这些参数的取值也影响到水印容量、不可见性和鲁棒性。为了得到高性能的水印算法，必须同时综合考虑水印的这些性能。因此，在水印算法的参数选取上，需要以这些性能要求作为约束条件，对参数取值进行优化。目前的研究工作中，对于水印安全与其它性能的联合优化的研究还刚刚开始，还需要进行深入研究。例如可以结合一些新型的优化算法来实现水印参数选取和性能优化。

(7) 安全水印

文献[27]对于水印安全的等级进行了定义并根据 Kerckhoffs 准则在 WOA 条件下划分出不安全水印、密钥安全水印、子空间安全水印和隐写安全水印。按照这样的定义，本文研究的扩频水印算法都属于不安全的水印。现有的关于水印安

全性分析的文献也多为针对不安全水印算法。文献[27]中提出一种称为“自然水印”的算法。这种水印算法以 ISS 水印为基础，具有在嵌入水印前后图像小波系数的统计分布不变的性质，因此属于隐写安全的水印。自然水印的鲁棒性很差，并不是一种成熟实用的水印方案。Mathon 等人在文献[27]的基础上提出基于载体模型的安全水印嵌入算法，能够在保证水印安全的同时实现失真的最小化[111]。设计和研究不同安全等级的水印方法是一个非常有意义的研究方向，目前在这方面的研究还比较少，进一步工作的空间非常广阔。

参考文献

1 I. J. Cox, M. L. Miller, J. A. Bloom. 数字水印. 王颖，黄至蓓等译. 北京：电子工业出版社， 2003.

2 黄继武，谭铁牛. 图像隐形水印综述. 自动化学报，2000， 26（5）：645-655.

3 孙圣和，陆哲明，牛夏牧. 数字水印技术及应用. 科学出版社，2004 年.

4 康显桂. 鲁棒图象水印方法的研究. 中山大学博士学位论文，中国，广州，2004.

5 M. Wu, B. Liu. Watermarking for image authentication. IEEE International Conference on Image Processing. Chicago, Illinois, October 4-7, 1998. Vol. 2, pp. 437-441.

6 T. Lin, E. J. Delp. A review of fragile image watermarks. Multimedia and Security Workshop at ACM Multimedia 99. Orlando, Florida, USA. ACM Press, 1999, pp. 25-29.

7 新浪网，新闻图片造假门，http://news.sina.com.cn/z/tupianzj/index.shtml

8 J. Fridrich. Image watermarking for tamper detection. IEEE International Conference on Image Processing. Chicago, Illinois, October 4-7, 1998, vol. 2, pp. 404-408.

9 吕述望，王彦，刘振华. 数字指纹综述. 中国科学院研究生院学报，2004，21(3), 289-298.

10 P. Moulin, R. Koetter. Data-hiding codes. Proc. of the IEEE, 2005, vol. 93(12), pp. 2083-2126.

11 N. Johnson, Z. Duric, and S. Jajodia. Information Hiding: Steganography and Watermarking: Attacks and Countermeasures. Boston, MA, Kluwer Academic Publishers, 2000.

12 J. G. Proakis, M. Salehi 著, 樊昌信, 改编. 现代通信原理. 电子工业出版社, 2007.

13 P. Moulin, A. Ivanovic. The zero-rate spread-spectrum watermarking game. IEEE Transactions on Signal Processing, , April 2003, vol. 51, no. 4, pp. 1098-1117.

14 H. S. Malvar, D.A. F. Florencio. Improved Spread Spectrum: a new modulation technique for robust watermarking. IEEE Transactions on Signal Processing, April 2003, vol. 51, no. 4, pp. 898-905.

15 B. Chen, G. W. Wornell. An information-theoretic approach to the design of robust

digital watermarking systems. Proceedings of IEEE International Conference On Acoustics, Speech, and Signal Processing (ICASSP). Phoenix, AZ, March 1999.

16 B. Chen, G. W. Wornell. Quantization index modulation: A class of provable good methods for digital watermarking and information embedding. IEEE Transactions on Information Theory, May 2001, vol. 47, no. 4, pp. 1423-1433.

17 M. H. M. Costa. Writing on dirty paper. IEEE Transactions on Information Theory. May 1983, vol.29, no.3, pp. 439-441.

18 L. Pérez-Freire, P. Comesaña, J. R. Troncoso-Pastoriza, and F. Pérez-González, Watermarking security: a survey. Transactions on Data Hiding and Multimedia Security I, October 2006, 4300: 41-72.

19 R. G. van Schyndel, A. Z. Tirkel, C. F. Osborne. A digital watermark. Proc. IEEE Int. Conference on Image Processing, Austin, Texas, USA 86-89, 1994.

20 曹志刚，钱亚生. 现代通信原理. 清华大学出版社，2005 年 6 月.

21 I. J. Cox, J. Kilian, F. T. Leighton, T. Shamoon. Secure spread spectrum watermarking for multimedia. IEEE Transactions on Image Processing, 1997, vol.12, no.6, pp.1673-1687.

22 J. Fridrich. Key-dependent random image transforms and their applications in image watermarking. Proc. International Conference on Imaging Science, Systems, and Technology, Las Vegas, NV, USA (1999), pp. 237-243.

23 G. H. Golub, C. F. Van Loan. Matrix Computations. 3rd ed. Baltimore, MD: Johns Hopkins, 1996.

24 J. J. Eggers, R. Bauml, R. Tzschoppe, B. Girod. Scalar costa scheme for information embedding. IEEE Transactions on Signal Processing, April 2003, vol. 51, no. 4, 1003-1019.

25 T. Furon, B. Macq, N, Hurley, G. Silvestre. JANIS: Just another N-Order side-informed watermarking scheme. IEEE International Conference on Image Processing, ICIP'02. Rochester, NY, USA (2002), vol 3, pp. 153-156.

26 G. J. Simmons. The prisoners' problem and the subliminal channel. Proc. Advances Cryptology, 1984, pp. 51-67.

27 F. Cayre, P. Bas. Kerckhoffs-based embedding security classes for WOA data hiding.

IEEE Transactions on Information Forensics and Security, , March 2008, vol. 3, no. pp. 11-15.

28 R. Chandramouli, M. Kharrazi, N. Memon. Image steganography and steganalysis: concepts and practice. T. Kalker et al. (Eds.), LNCS 2939, IWDW 2003, pp. 35-49, Springer-Verlag, Berlin Heidelberg 2004.

29 A. Kerckhoffs. La cruptgraphie militaire. Journal des Sciences Militaires, IX (Jan.,Feb.) 1883: 5-38, pp. 161-191.

30 刘鸿霞, 夏春和. 图像隐写分析现状研究. 计算机工程与设计, 2006 年 1 月, 第 27 卷, 第 1 期, 21-25.

31 P. Sallee. Model-based methods for steganography and steganalysis. International Journal of Image Graphics, Jan. 2005, vol. 5, no. 1, pp.167-190.

32 F. Collin. Encryptpic. http://www.winsite.com/bin/Info?500000033023

33 G. Pulcini. Stegotif. http://www.geocities.com/SiliconValley/9210/gfree.html

34 T. Sharp. Hide 2.1, 2001. http://www.sharpthoughts.org

35 A. Westfeld, A. Pfitzmann, High capacity despite better steganalysis (F5-A steganographic algorithm). Lecture Notes in Computer Science, 2001, vol.2137, Springer-Verlag, Berlin.

36 N. Provos. Defending against statistical steganalysis. 10th USENIX Security Symposium, 2001.

37 A. Westfeld, A. Pfitzmann. Attacks on Steganographic Systems. Lecture Notes in Computer Science. 2000, vol. 1768, pp. 61-75. Springer-Verlag, Berlin.

38 J. Harmsen, W. Pearlman. Steganalysis of additive noise modelable information hiding. Conference on Security and Watermarking of Multimedia Contents V. Jan 21-24, 2003 Santa Clara, California, USA .

39 K. Sullivan, U. Madhow, S. Chandrasekaran, B. Manjunath. Steganalysis for Markov cover data with applications to images. IEEE Transactions on Information Forensics and Security, June 2006, vol.1, no. 2, pp.275-287.

40 H. Farid. Detecting hidden messages using higher-order statistical models. Proc. IEEE International Conference on Image Processing, New York, Sep. 2002, pp. 905-908

41 G. Xuan, Y. Q. Shi, J. Gao, D. Zou, C. Yang, Z. Zhang, P. Chai, C. Chen, and W. Chen.

Steganalysis based on multiple features formed by statistical moments of wavelet charateristic functions. Proc. Information Hiding Workshop, Barcelona, Spain, Jun. 2005, pp. 262-277.

42 S. Lyu, H. Farid. Steganalysis using higher-order image statistics. IEEE Transactions on Information Forensics and Security, March 2006, vol. 1, No. 1, pp. 111-119.

43 G. R. Xuan, Y. Q. Shi, C. Huang, D. D. Fu, et al. Steganalysis using high-dimensional features derived from co-occurrence matrix and class-wise no-principal components analysis (CNPCA). Y. Q. Shi and B. Jeon (Eds.): IWDW 2006, LNCS 4283, 2006, pp. 49-60, Srpinger-Berlag Berlin.

44 Y. Wang, P. Moulin. Optimized feature extraction for learning-based image steganalysis. IEEE Transactions on Information Forensics and Security. 2007, vol. 2, No. 1, pp. 31-45.

45 P. Moulin, J. A. O'Sullivan. Information-theoretic analysis of information hiding. IEEE Transaction on Information Theory. Mar. 2003, vol.49, No.3. pp. 563-593.

46 P. Comesaña, L. Pérez-Freire, F. Pérez-González. Fundamentals of data-hiding security and their application to spread spectrum analysis. Proc. Information Hiding Workshop, 2005, pp.146-160. Lecture. Notes Computer Science. Springer-Verlag.

47 J. Killian, F. T. Leighton, L. R. Matheson, T. Shamoon, R. E. Tarjan. Resistance of watermarked documents to collusional attacks. Technical report, NEC Research Institute, Princeton, JN, 1997.

48 F. Cayre, C. Fontaine, T. Furon. Watermarking security: theory and practice. IEEE Trans. Signal Processing. Oct. 2005, vol. 53. no. 10, pp. 3976-3987.

49 I. J. Cox, J. P. M. G. Linnartz. Public watermarks and resistance to tampering. Proc. IEEE Int. Conf. on Image Processing. 1997, vol. 3, Santa Barbara, California, USA. 3-6.

50 C. Cachin. An information-theoretic model for steganography. D. Aucsmith, ed.: 2nd Int. Workshop on Information Hiding, IH'98. 1998. vol. 1525 of Lecture Notes in Computer Science. Portland, OR, USA, Springer Verlag, 306-318.

51 T. Mitthelholzer. An information-theoretic approach to steganography and watermarking. A. Pfitzmann, ed.: 3rd Int. Workshop on Information Hiding, IH'99. 1999. vol. 1768 of Lecture Notes in Computer Science, Dresden, Germany, Springer Verlag 1-17.

52 T. Kalker. Considerations on watermarking security. IEEE international Workshop on Multimedia Signal Processing, Cannes, France, 2001. 201-206

53 T. Furon, et al. Security Analysis. European Project IST-1999-10987 CERTIMARK, Deliverable D.5.5, 2002.

54 R. A. Fisher. On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society 222, 1922, pp. 309-368.

55 H. Cramer. Mathematical Methods of Statistics. Princeton, NJ: Princeton University Press, 1946.

56 C. Rao, S. Das Gupta. Selected Paters of C. R. Rao. New York: Wiley. 1994.

57 T. K. Moon, W. C. Stirling. Mathematical Methods and Algorithm for Signal Processing. Prentice Hall, Upper Saddle River, NJ, USA, 2000.

58 A. Hyvarinen, E. Oja. Independent component analysis: A tutorial. Neural Networks. 2000, vol. 12, no. 4-5, pp. 411-430.

59 A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Networks. May 1999, vol. 10, no. 3, pp. 626-634.

60 T. M. Cover, J. A. Thomas. Elements of Information Theory. Wiley series in Telecommunications, 1991.

61 J. Q. Ni, R. Y. Zhang, C. Fang, J. W. Huang, and C. T. Wang. Watermarking Security Incorporating Natural Scene Statistics. Lecture Notes in Computer Science, Information Hiding 2008, vol. 5284/2008. pp. 132-146. Springer Berlin / Heidelberg.

62 R. A. Choudrey. Variational Method for Bayesian Independent Component Analysis. Ph.D. Thesis, Oxford University, 2002.

63 L. Pérez-Freire, F. Pérez-González, T. Furon and P. Comesaña. Security of Lattice-Based Data Hiding Against the Known Message Attack. IEEE Trans. Information Forensics and Security. December 2006, vol. 1, pp. 412-439.

64 L. Pérez-Freire and F. Pérez-González. Security of lattice-based data hiding against the Watermarked Only Attack. IEEE Transactions on Information Forensics and Security. 2008, vol.3, no. 4, pp.593-610.

65 P. L. Combettes. The foundations of set theroretic estimation. Proceedings of the IEEE, Feb 1993, vol. 81, no.2, pp.182-208.

66 J. H. Conway, N. J. A. Sloane. Voronoi regions of lattices, second moment of polytopes, and quantization. IEEE Transactions on Information Theory. March 1982, vol. IT-28, no. 2, pp. 211-226.

67 L. Pérez-Freire, F. Pérez-González. Spread Spectrum Watermarking Security. IEEE Transactions on Inoformation Forensics and Security. 2009, vol. 4 (1), pp. 2-24.

68 A. S. Householder. Unitary triangularization of a nonsymmetric matrix. Journal ACM, 1958, vol. 5, no. 4, 339-342.

69 H. Tijms. Understanding Probability: Chance Rules in Everyday Life. Cambridge: Cambridge University Press, 2004.

70 J. Portilla, V. Strela, M. J. Wainwright, E. P. Simocelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE Transactions on Image Processing. November 2003, vol. 12, no. 11, pp.1338-1351.

71 S. W. Liu, E. P. Simoncelli. Modeling multiscale subbands of photographic images with fields of Gaussian scale mixtures. IEEE Transactions on Pattern Analysis and Machine Intelligence, April 2009, vol. 31, issue 4, pp. 693-706.

72 D. L. Ruderman. The statistics of natural images. Computation in Neural Systems. 1999, 5, pp. 517-548.

73 E. P. Simoncelli. Statistical modeling of photographic images. Handbook of Video and Image Processing. 2nd edition, ed. Alan Bovik, 2005, Academic Press.

74 M. J. Wainwright, Eero. P. Simoncelli. Scale mixtures of Gaussians and the statistics of natural images. Advances in Neural Information Processing Systems 12, ed. S.A. Solla, T. K. Leen and K.R. Muller, 2000, pp.855-861, MIT Press.

75 E. T. Jaynes. Where do we stand on maximum entropy? The Maximal Entropy Formalism, ed. R. D. Levine, M. Tribus. 1978, MIT Press.

76 A. van der Schaaf and J. H. van Hateren. Modeling the power spectra of natural images: Statistics and information. Vision Research, 1996, vol. 28, no. 17, pp. 2759-2770.

77 P. Moulin, J. Liu. Analysis of multiresolution image denoising schemes using a generatlized Gaussian and complexity priors. IEEE Trans. Info. Theory. 1999. vol. 45 pp. 909-919.

78 E. P. Simoncelli. Bayesian denosing of visual images in the wavelet domain. Bayesian

Inference in Wavelet Based Models, eds. P. Muller, B. Vidakovic. 1999. pp. 291-308. Springer-Verlag, New York.

79 S. W. Liu, Eero. P. Simoncelli. Statistical modeling of images with fields of Gaussian scale mixtures. Advances in Neral Information Processing Systems, eds, B. Scholkopf, J. Platt andT. Hofmann, May 2007, vol. 19, pp, MIT Press.

80 H. R. Sheikh, A. C. Bovik, G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. IEEE Transactions on Image Processing, December 2005, vol. 14, no. 12, pp.2117-2128.

81 S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. IEEE Trans. Pattern Anal. Machine Intell., 1989, vol. 11, no. 7, pp. 674-693.

82 G. Van de Wouwer, P. Scheunders, D. Van Dyck. Statistical texture characterization from discrete wavelet representations. IEEE Transactions on Image Processing. 1999, vol. 8, no.4, pp. 592-598.

83 E. P. Simoncelli, E. H. Adelson. Noise removal via Bayesian wavelet coring. Proc. IEEE int. Conf. Image Proc. (ICIP), Lausanne, Switzerland, Oct. 1996, pp. 379-382.

84 M. H. Pi, C. S. Tong, S. K. Choy, H. Zhang, A fast and effective model for wavelet subband histograms and its application in texture image retrieval. IEEE Transactions on Imgae Processing. 2006, vol.15, no.10, pp. 3078-3088.

85 H. Yuan, X. P. Zhang. Multiscale fragile watermarking based on the Gaussian mixture model. IEEE Transaction on Image Processing. 2006, vol.15, no.10. pp. 3189-3200.

86 R. Hogg, J. McKean and A. Craig. Introduction to Mathematical Statistics. Upper Saddle River, NJ: Pearson Prentice Hall, 2005, pp. 359-364.

87 K. B. Petersen, O. Winther. The EM algorithm in independent component analysis. ICASSP 2005, pp.169-172.

88 Y. Wan, D. B. Shi. Joint exact histogram specification and image enhancement through the wavelet transform. IEEE Transactions on Image Processing. 2007, vol. 16, No. 9, pp.2245-2250.

89 M. S. Crouse, R. D. Nowak, R. G. Baraniuk. Wavelet-based statistical signal processing using hidden Markov models. IEEE Transactions on Signal Processing. 1998, vol.46, no.

4, pp. 886-902.

90 张荣跃，倪江群，黄继武. 基于小波域隐马尔可夫模型的鲁棒多比特图像水印算法. 软件学报. 2005，16(7)：1323-1332.

91 王春桃. 基于模型的稳健图像数字水印关键技术研究. 中山大学博士论文，2007.

92 R. O. Duda, P. E. Hart, D. G. Stork 著，李宏东，姚天翔等译. 模式分类，机械工业出版社，2006.

93 M. J. Schervish. Theory of Statistics. New York: Springer. 1995.

94 李道本. 信号的统计检测与估计理论. 科学出版社，2004 年第二版.

95 P. Stoica, A. Nehorai. MUSIC, maximum likelihood, and Cramer-Rao bound. IEEE Transactions on Acoustics, Speech and Signal Processing. May 1999, vol. 37, no. 5, pp. 720-741.

96 H. L. van Trees. Detection, Estimation, and Modulation Theroy. John Wiley and Sons, 1968.

97 郑君里, 应启珩, 杨为理. 信号与系统. 高等教育出版社. 2000 年第 2 版.

98 陈前斌, 蒋青, 于秀兰. 信息论基础. 高等教育出版社. 2007.

99 P. Stoica and B. C. Ng. On the Cramer-Rao bound under parametric constraints. IEEE Signal Processing Lett. 1998. vol. 5, no. 7, pp. 177-179.

100 N. A. D'Andrea, U. Mengli, and R. Reggiannini. The modified Carmer-Rao bound and its application to synchronization problems. IEEE Transactions on Communications, Feb./Mar./Apr. 1994, vol. 42, pp.1391-1399.

101 R. W. Miller, C. B. Chang. A modified Carmer-Rao bound and its applications. IEEE Transactions on Information Theory. May 1978, vol. IT-24, no. 3, pp. 398-400.

102 Ya-lun Chou. Statistical Analysis. Holt International. 1969.

103 F. Gini, R. Reggiannini, and U. Mengali. The modified Cramer-Rao bound in vector parameter estimation. IEEE Transactions on Communications. Jan. 1998, vol. 46, no. 1, pp.52-60.

104 张东, 倪江群, 李大捷. 基于 GSM 模型的扩频水印安全性分析. 自动化学报. 2009, 35 (7) 841-850.

105 D. Zhang, J. Q. Ni, D. J. Lee. GSM based security analysis for Add-SS watermarking. Advances in Visual Computing, 4th International Symposium, ISVC 2008, LNCS 5359,

pp.400-409.

106 张贤达. 矩阵分析与应用. 清华大学出版社. 2005.

107 T. Fang, S. Kotz, K. W. Ng. Symmetric Multivariate and Related Distributions. London, New York: Chapman & Hall, 1990.

108 S. W. Lyu. Modeling multiscale subbands of photographic images with fields of Gaussian Scale Mixtures. PAMI. April 2009, vol. 31 no. 4, pp. 693-706.

109 F. Pérez-González, F. Balado, and J. R. Hernández. Performance analysis of existing and new methods for data hiding with known-host information in additive channels. IEEE Transactions on Signal Processing, Special Issue on Signal Processing for Data Hiding in Digital Media & Secure Content Delivery. April 2003, vol.51, no. 4, pp. 960-980.

110 L. Pérez-Freire, F. Pérez-González, and S. Voloshynovskiy. An accurate analysis of scalar quantization-based data-hiding. IEEE Transactions on Information Forensics and Security. March 2006, vol.1, no.1, pp.80-86.

111 B. Mathon, P. Bas, F. Cayre and F. Pérez-González. Distortion optimization of model-based secure embedding schemes for data-hiding. Information Hiding: 10th International Workshop, IH 2008, Santa Barbara, CA, USA, May 19-21, 2008,

附录一：主要英文缩写全称对照表

英文缩写	全称
Add-SS Watermarking	Additive Spread-Spectrum Watermarking
ASS Watermarking	Attenuated Spread Spectrum Watermarking
BPSK	Binary Phase Shift Keying
CRB	Cramer-Rao Bound
DC-DM QIM	Distortion-Compensated Dither-Modulation QIM
DWR	Document to Watermark Ratio
EM	Expectation Maximization
FIM	Fisher Information Matrix
GGM	Generalized Gaussian Model
GMM	Gaussian Mixture Model
GSM	Gaussian Scale Mixture
ICA	Independent Component Analysis
ISS Watermarking	Improved Spread-Spectrum Watermarking
KMA	Known Message Attack
KOA	Known Original Attack
LSB	Least Significant Bit
MCRB	Modified Cramer-Rao Bound
QIM	Quantization Index Modulation
VB-ICA	Variational Bayesiam ICA
WOA	Watermarked Only Attack

附录二：在学期间发表论文列表

1. 张东, 倪江群, 李大捷. 基于 GSM 模型的扩频水印安全性分析. 自动化学报. 2009 年，第 35 卷，第 7 期，841－850 页。（对应论文第四章）
2. Dong Zhang, Jiangqun Ni, Dah-Jye Lee. Security analysis for Spread-Spectrum watermarking incorporating statistics of natural images. Advances in Visual Computing. 4th International Symposium, ISVC 2008. LNCS 5359, pp.400-409, 2008. （对应论文第四章）
3. Dong Zhang, Jiangqun Ni, Dah-Jye Lee, Jiwu Huang. GSM based security analysis for Add-SS watermarking. 7th International Workshop on Digital Watermarking, Busan, Korea, Nov, 2008. （对应论文第四章）
4. Dong Zhang, Jiangqun Ni, Qiping Zeng, Jiwu Huang. Security analysis on improved Spread-Spectrum watermarking incorporating statistics of natural images. Science in China Series F: Information Sciences. （已投稿，对应论文第五章）
5. 倪江群，唐承佩，张东，冯国聪. CAN 通信协议及其硬件实现. 通信学报，2008 年第 29 卷第 5 期，107-113 页.
6. 张东，黄宇恒. 基于 FPGA 的多路模拟数据采集接口设计. 仪表技术，2005 年第 6 期, 16-18,37 页.

致　谢

本论文是在倪江群教授的指导下完成的，在此向倪老师致以最诚挚的感谢。从论文的选题、研究中的讨论到论文的完成，倪老师都倾注了极大的心血。倪老师深厚的学术造诣、严谨的治学态度和勇于创新的探索精神都使我受益匪浅。感谢倪老师为我创造了多次参加学术会议的机会并支持我赴美国进行交流访问研究。

感谢 Brigham Young University 电气与计算机工程系的李大捷教授邀请我进行交流访问研究。李老师在学术上的指导和在生活上的关心，帮助我圆满完成了在美国的研究任务。感谢李老师资助我参加多次国际学术会议。

本人的工作得到了实验室同事和同学的支持和热情帮助。在此向王春桃博士、张荣跃博士、岳薇工程师、唐承佩博士和实验室的同学们表示衷心感谢。

感谢魏兆一博士对我的帮助。在美国的一年中，魏博士帮助我克服了许多生活上的困难。与魏博士在学术上的讨论使我受益良多。感谢在美国期间所有帮助过我的朋友们。

多年来，家人的支持一直是我努力科研并最终完成论文的动力。父母、妻子和岳父母为了我能够专心学习和科研，作出了很大的牺牲。女儿也时常调皮地送上笑脸，为我带来无尽欢乐。亲人的关爱是我一生最为珍贵的财富，谢谢你们！

最后，感谢国家留学基金委对我在美国交流访问研究的资助。