

基于纠错码的信息隐藏及其实现方法^{*}

马永辉¹, 徐长勇², 平西建³, 刘翠卿³

(1. 空军雷达学院 信息处理工程系, 武汉 430019; 2. 空军装备研究院 雷达所, 北京 100085; 3. 解放军信息工程大学 信息工程学院, 郑州 450002)

摘要: 信息隐藏是信息安全领域的重要研究内容之一。介绍了基于纠错码的信息隐藏的一般原理, 阐述了主要实现方法。针对一种基于纠错码编码冗余的信息隐藏算法框架, 给出了以 JPEG 图像为载体的实现方法, 并分析了其特点。最后指出了值得进一步研究的问题。

关键词: 信息隐藏; 数字隐写; 数字水印; 纠错码; 鲁棒性

中图分类号: TP309; TP391 **文献标志码:** A **文章编号:** 1001-3695(2012)07-2686-04

doi:10.3969/j.issn.1001-3695.2012.07.078

Survey and realization on information hiding based on error-correcting code

MA Yong-hui¹, XU Chang-yong², PING Xi-jian³, LIU Cui-qing³

(1. Dept. of Information Processing Engineering, Air Force Radar Academy, Wuhan 430019, China; 2. Institute of Radar & Electronic Counter Measures, Equipment Academy of Air Force, Beijing 100085, China; 3. Institute of Information Engineering, PLA Information Engineering University, Zhengzhou 450002, China)

Abstract: Information hiding is one of the most important research subjects in the field of information security. This paper summarized several information hiding methods using error-correcting code, then proposed an algorithm based on the encoding redundancy of error-correcting code, and gave the realization in JPEG encoded images and the characteristics are analyzed.

Key words: information hiding; steganography; watermarking; error-correcting code; security; robustness

0 引言

随着互联网通信技术和多媒体信息处理技术的发展, 信息隐藏成为信息安全领域的重要研究内容之一。信息隐藏利用人的视觉、听觉等感知冗余以及多媒体数据的统计冗余, 将秘密信息以一定的方式嵌入到公开的数字载体中。这些载体通常是网络上传输的数字媒体文件, 如图像、音频、视频或文档等^[1]。信息隐藏主要包括数字隐写和数字水印两个技术分支, 其中数字水印主要用于版权保护, 所嵌入的水印是与载体相关的作者、所有者、出版时间等信息, 其保护对象是嵌入信息的载体。数字隐写则是将秘密信息以不引起第三方注意的方式嵌入到载体中, 并在公开网络中进行传输, 因此其保护对象是嵌入到载体中的信息, 主要用于隐蔽通信。与传统的密码技术相比, 由于嵌入信息后对外表现的是载体信号的内容和特征, 隐藏了通信的存在性, 因而更不容易被第三方发现。

数字水印要求嵌入水印信息后不影响载体的商用价值, 而且所嵌入的水印应该能够抵抗各种处理或攻击, 即应该具有鲁棒性。而作为一种新的隐蔽通信手段, 数字隐写利用载体来传输秘密信息, 通常需具有两方面的重要特性, 即隐藏容量和安全性。隐藏容量保证了能够传输足够多的数据, 而在安全性方面, 首先要求秘密信息嵌入不影响载体的外在感官表现, 其次

要求嵌入信息后载体的统计特性保持不变。

近年来, 随着信息隐藏技术的不断发展以及与之相对抗的信息隐藏攻击技术的不断进步, 研究者在提高数字水印的鲁棒性以及数字隐写的安全性等方面采取了多种措施。其中, 纠错码的引入使信息隐藏在性能提升方面取得了较好的效果。纠错码通常用于通信时提高信息传输的可靠性, 在信息码中加入一定数量的多余码元(即监督码元), 信息码和监督码元之间满足一定的约束关系, 在传输中发生错误时, 利用这一约束关系可以检错并且纠错。纠错码在信息隐藏技术的应用可以追溯到 Crandall 提出的矩阵编码^[2], 目前的应用还包括利用纠错码提高鲁棒性以及携带秘密信息等。

本文研究纠错码在信息隐藏技术中的应用, 介绍了常见的基于纠错码的信息隐藏方法, 针对基于纠错码编码冗余的信息隐藏算法框架, 给出了一种实现方法, 并对该方法进行分析, 最后指出了进一步的研究方向。

1 基于纠错码的信息隐藏

1.1 利用纠错码提高鲁棒性

先对要嵌入载体的信息进行纠错编码, 然后将编码后的信息嵌入载体数据。这样, 在载密数据遭受攻击的情况下, 接收方仍可以利用纠错码的纠错性能恢复嵌入信息, 因而主要用于

收稿日期: 2011-11-25; 修回日期: 2012-01-06 基金项目: 国家自然科学基金资助项目(60970142)

作者简介: 马永辉(1978-), 男, 河北定州人, 硕士研究生, 主要研究方向为信息处理等(mayh1978@163.com); 徐长勇(1980-), 男, 安徽凤阳人, 博士, 主要研究方向为网络对抗、电子对抗等; 平西建(1953-), 男, 河南新乡人, 教授, 博导, 主要研究方向为信息隐藏、图像处理、模式识别等; 刘翠卿(1964-), 女, 山东烟台人, 高级工程师, 博士, 主要研究方向为信息隐藏、信息处理等。

数字水印技术中提高水印信息的鲁棒性,其技术特点主要包括如下几个方面^[3,4]:

a)与未使用纠错码进行预处理相比,利用纠错码对水印信息进行预处理,可以明显提高水印信息的鲁棒性。但是,当攻击增大到一定程度时,纠错码的作用效果变得不明显。

b)在水印嵌入容量、嵌入强度与方法相同的情况下,在低噪声信道,卷积码的纠错能力最好;在高噪声信道,卷积码的纠错效果最差,出现越纠越错的情况。而 BCH 码的纠错能力相对比较稳定。

c)对于同一种纠错码设置不同码率,随着码率的减少,提取出的水印误码率随之减小。但是减少码率是以增加冗余为代价的,相应的编码效率要降低。

1.2 利用纠错码携带秘密信息

在有噪声信道中传输数据时,信道噪声的存在将使所收到的数据含有差错。利用信道有噪声这种现象,在纠错码的纠错范围之内将秘密信息嵌入其中,从而实现有噪声信道中的信息隐藏^[5-7],其信息隐藏原理如下:

a)在发送端,先将秘密信息同伪随机序列模二加,然后进行纠错编码,接着对载体信源进行编码,同时根据信道的特点来选择合适的码字嵌入位置,即嵌入位分析。经过嵌入位分析,将编码后的秘密信息嵌入到编码后的载体信源中,最后送入信道中进行传输。

b)在接收端,先对接收数据进行信道译码,然后对译码后的数据进行嵌入位检测,利用检测出的嵌入位在未经过信道译码的数据中提取信息。对提取出的信息进行纠错译码,然后与伪随机序列进行模二加得到秘密信息。

在信息隐藏机制不公开的情况下,这种信息隐藏方法具有一定的安全性。但是为了保证安全性,需要尽量使得信息嵌入率同信道误码率在同一数量级上,从而导致该方法的隐藏容量较小。

1.3 利用纠错码增强安全性

矩阵编码通过较少的对载体数据的改动来嵌入较多的信息,在嵌入量相同的情况下,由于对载体的修改量减小,因而减少了载体数据的统计特性变换,从而提高了安全性。Westfeld^[8]在其所提出的 F5 算法中实现了矩阵编码的思想。该方法可以在 $2^l - 1$ bit 的载体数据中,利用不多于 1 bit 的改动嵌入 l bit 秘密信息,秘密信息的嵌入率较高。矩阵编码的关键是对原始数据的序号进行二进制编码,也就是说每一个原始数据对应一个二进制数,用于编码的 l 个二进制组合可以看做一组互不相同的向量。F5 算法限定最大修改比特数为 1,编码方法比较单一,因而对秘密信息嵌入率的限制很大。

文献[9]对这一方法进行了推广,提出了利用分组码来构造任意嵌入率的方法,从而减少了对嵌入率的限制,可以有更多的编码方式选择。密写编码和混合密写编码^[1]对上述方法作了进一步的推广,这些方法都以牺牲载体数据的利用率为代价,来达到减少载体数据失真的目的,但这必然造成隐藏容量的降低。

1.4 基于纠错码编码冗余的信息隐藏

信息隐藏是根据人类的视听觉特性,利用数字媒体中存在

的冗余来隐藏信息的,而纠错码是通信中为了提高信息传输可靠性而人为加入冗余的具有严密数学结构的编码数据,因此利用纠错码的这种编码机制,纠错码在可靠传输环境下的编码冗余便成为一种可以利用的载体资源^[10,11]。

设 (n, k) 纠错码的校验矩阵为 H , c 是该纠错码中的一个码字,用 c 作为载体码字,载密码字用 r 表示,错误图样为 $e = r - c$,则

$$S^T = Hr^T = H(c + e)^T = Hc + He^T = 0 + He^T = He^T$$

称 S 为 r 或者 e 的伴随式。

由此可见,设 S 是 (n, k) 纠错码对应的伴随式,则 S^T 仅与错误图样 e 有关,而与发送的码字无关, S 完全由 e 确定。也就是说,如果在码字 c 中嵌入信息,且其引入的错误在纠错码的纠错范围之内,那么是可以通过纠错译码算法识别并获得错误图样的。由此可得基于纠错码编码冗余的信息隐藏算法原理。

从载体数字媒体中提取 n bit,以前 k 位作为信息位进行纠错编码,可得 n 位的纠错码码字。以这 n 位的纠错码码字作为载体码字 c ,设 m 是要嵌入载体码字的秘密信息,在秘密信息 m 与错误图样 e 之间建立一一对应的映射关系。载密码字用 r 表示,则 $r = c + e$,从而得到载密码字(对于二元域上的纠错码来说,这里的 $+$ 指的是模二加,下同)。接着,将载密码字嵌入到从数字媒体提取出的 n bit 中,从而得到载密数字媒体。

由于纠错码的编码体制为收发双方共知,当接收方收到载密码字 r 时,通过计算 $S^T = Hr^T$ 可得伴随式 S 。由于 S^T 仅与错误图样 e 有关,纠错码纠错范围内的错误图样 e 的求解,可以根据 $S^T = He^T$ 利用各种已有的译码方法求解,再由秘密信息与错误图样间的映射关系即可得到秘密信息。

由于载体码字的长度为 n ,设 (n, k) 纠错码可以纠正 t 个错误,则与载体码字对应的错误图样的个数为

$$N_e = C_n^0 + C_n^1 + \cdots + C_n^t \quad (1)$$

如果对这些错误图样进行二进制编码,则其最大编码子集的长度为

$$h = \lfloor \log_2 N_e \rfloor = \lfloor \log_2 (C_n^0 + C_n^1 + \cdots + C_n^t) \rfloor \quad (2)$$

其中, $\lfloor \cdot \rfloor$ 表示向下取整。显然, h 即为在每个载体码字中可嵌入的比特数。

图1给出了上述基于纠错码编码冗余的信息隐藏算法框架。

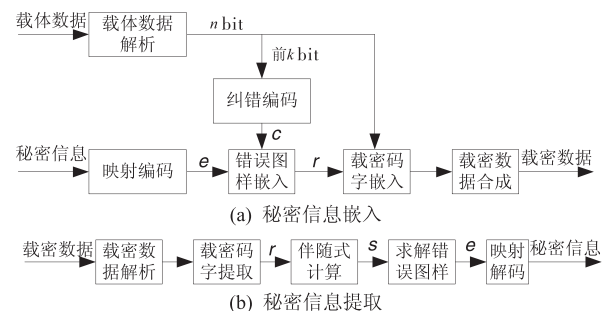


图1 基于纠错码编码冗余的信息隐藏算法

2 基于纠错码编码冗余的信息隐藏算法实现

以在 JPEG 图像压缩域的 DCT 系数中嵌入信息为例,给出一种具体的实现方法。

JPEG 图像压缩时,首先将原始图像分割成互不重叠的 8×8 块,然后对各个块进行二维 DCT,并以量化表量化,得到量化 DCT 系数块。对于量化后的 8×8 系数块,位于第 1 行第 1 列的 DC 系数包含了图像的主要信息(如亮度),要对它单独编码。由于两个相邻块的 DC 系数相差很小,对其采用差分编码,其他 63 个 AC 系数中存在大量为零的值,因此以行程长度编码来对零值进行压缩。最后,对差分编码后的 DC 系数和行程长度编码后的 AC 系数,均采用熵编码得到压缩图像数据。

对于隐藏算法来说,如果修改 DC 系数,将会较大地影响图像质量,且会造成误差累积。而如果利用 AC 系数携带信息,则不存在这一问题。同时考虑到量化 DCT 系数中大部分都为零,如果修改这些系数,将会大大减少零的数量,从而引起攻击者的注意,因此不在值为零的系数中嵌入信息。也就是说,应该将秘密信息嵌入到非零的 AC 系数中。

2.1 秘密信息嵌入

秘密信息嵌入的具体步骤如下:

a) 对于秘密信息序列 M ,将其分为 h 个一组,得秘密信息组合 $m = (m_1 m_2 \cdots m_h)$ 。根据秘密信息与错误图样之间的映射关系,将各组秘密信息分别映射为错误图样 $e = (e_1 e_2 \cdots e_n)$,最终可得错误图样序列 E 。

b) 对压缩图像进行变长解码和行程长度解码,得到与图像大小相同的量化 DCT 系数矩阵。对其中的各个 8×8 块,以 zigzag 扫描的顺序提取其中的非零 AC 系数并取绝对值,同时记录各个系数的正负号,得到一维的非零 AC 系数序列 $L = (l_1, l_2, \cdots, l_p)$, p 为该块中非零 AC 系数的个数。如果 $p < n$,则不在该块中嵌入信息,继续对下一个块进行处理;如果 $p \geq n$,则在该块中嵌入信息。

c) 对于可嵌入信息的块,从 L 中选取最大的 n 个值,并以其在 L 中的顺序排列得 $L_s = (l_1, l_2, \cdots, l_n)$ 。如果 L 中存在多个相等的可以选取到 L_s 中的值,则取在 L 中排列靠前的值。接着,取 L_s 中前 k 个系数的 LSB(least significant bit,最不重要比特),并进行纠错编码,得到载体纠错码码字 $c = (c_1 c_2 \cdots c_n)$ 。

d) 从错误图样序列 E 中依次取各组错误图样 e ,则载密纠错码码字 $r = c + e$ 。

e) 通过修改 L_s 中各个系数的 LSB,把载密码码字 r 嵌入到各个 DCT 系数块中。

设 $l_i (i \in L_s)$ 是当前块中用于嵌入载密码码字的系数,LSB(l_i) 是其 LSB, $r_i (r_i \in r)$ 是要嵌入 l_i 的载密码码字中的一个比特,修改后的 l_i 为 l'_i , L_s 中各个系数的最小值为 $l_{\min}$,则嵌入规则如下:

(a) 当 $l_i = l_{\min}$ 时,若 $r_i = \text{LSB}(l_i)$,则 $l'_i = l_i$;若 $r_i \neq \text{LSB}(l_i)$,则为了实现盲提取,必须使 l_i 在修改后仍不小于 $l_{\min}$,因此只能将其加 1,即 $l'_i = l_i + 1$ 。

(b) 当 $l_i > l_{\min}$ 时,若 $r_i = \text{LSB}(l_i)$,则 $l'_i = l_i$;若 $r_i \neq \text{LSB}(l_i)$,用 L_s 表示 L 中除 L_s 以外的系数组成的集合,其最大值为 $l_{\max}$,显然 $l_i \geq l_{\max}$ 。此时,根据 $l_{\max}$ 与 l_i 的大小关系来修改 l_i :

① 当 $l_i - l_{\max} = 1$,分以下两种情况进行修改:如果 L 中在 l_i 之前存在等于 $l_{\max}$ 且不属于 L_s 的系数,则只能将 l_i 加 1,即 $l'_i = l_i + 1$;如果 L 中等于 $l_{\max}$ 的系数都在 l_i 之后,则 $l'_i = l_i - 1$ 或 $l_i + 1$;

② 当 $l_i - l_{\max} < 1$,由于 $l_i \geq l_{\max}$,则 $l_i = l_{\max}$ 。为了使 l'_i 仍为最大的 n 个系数之一,只能将 l_i 加 1,即 $l'_i = l_i + 1$;

③ 当 $l_i - l_{\max} > 1$,对 l_i 随机加减 1 以使其 LSB 与 r_i 一致,即 $l'_i = l_i - 1$ 或 $l_i + 1$ 。

根据上述各种情况修改 L_s 中各个系数,并将结果放回 L 得 L' 。将 L' 中各值乘以其各自的符号,放回其在原系数块中的位置,即得嵌入信息的系数块。

f) 对嵌入信息的 DCT 系数矩阵的各个 8×8 块进行行程长度编码和变长编码,最终由熵编码得到载密图像数据流。

2.2 秘密信息提取

在秘密信息接收端,图像数据解析部分与信息嵌入时相同,而且纠错码的编码机制是为收发双方所共知的。具体步骤如下:

a) 对于由 8×8 系数块中的非零 AC 系数组成的序列 $L = (l_1, l_2, \cdots, l_p)$,如果 $p < n$,则该块中不存在秘密信息,继续对下一个块进行处理;如果 $p \geq n$,则该块中存在秘密信息。

b) 对于存在秘密信息的块,从 L 中取最大的 n 个值,并以其在 L 中的顺序排列得 $L_s = (l_1, l_2, \cdots, l_n)$ 。提取各个系数的 LSB,即得 n 位的载密码码字 r 。

对 DCT 系数矩阵中其他所有块进行上述同样的处理,即可得载密码码字序列 R 。

c) 对于 R 中的各组码字 r ,由 $S = rH^T$ 计算伴随式 S ,再由 $S^T = He^T$ 求解 e ,最后可得错误图样序列 E 。

d) 根据通信双方约定的秘密信息与错误图样之间的一一对应映射关系表,依次对 E 中的各组错误图样进行映射解码,即可得到秘密信息序列。

2.3 算法实验结果及分析

使用五幅大小为 512×512 的标准灰度图像进行实验。首先以质量因子 75 进行 JPEG 压缩,然后分别使用 (7,4) 汉明码和 (23,12) 格雷码在其中嵌入信息。

2.3.1 隐藏容量

表 1 列出了五幅图像的隐藏容量(本文 1 和 2 分别指使用 (7,4) 汉明码和 (23,12) 格雷码时的情况,下同)。可见,对于 airplane 和 Lena 等存在较多平滑区域的图像,存在的非零 AC 系数较少,因此可嵌入信息的块更少,隐藏容量较小;而对于 baboon 和 bridge 等纹理较复杂的图像,隐藏容量相对较大。

表 1 不同情况下的 PSNR 和隐藏容量

图像	隐藏容量/bit		PSNR/dB		
	本文 1	本文 2	不嵌入	本文 1	本文 2
airplane	5 426	2 970	38.474 8	38.020 5	38.224 2
baboon	8 178	24 090	31.339 9	31.192 7	30.655 9
bridge	8 118	19 393	32.185 1	32.028 2	31.511 7
boat	6 934	4 356	36.798 8	36.318 3	36.438 2
Lena	4 424	1 419	37.833 2	37.428 9	37.603 7

2.3.2 安全性

信息隐藏的安全性通常从两个方面来进行度量,即视觉不可感知性和统计不可感知性。视觉不可感知性要求从视觉上无法区分载密对象与载体对象,即两者在图像质量方面需保持一致。图 2 显示了两幅图像 airplane 和 bridge 在秘密信息嵌入前后的变化情况。可以看出,信息嵌入没有对图像质量造成视觉上可见的影响。

同时,使用峰值信噪比 (PSNR) 来定量衡量信息嵌入对图像质量的影响。分别计算不嵌入信息时以及嵌入信息后的图像与原始未压缩图像之间的 PSNR,同样列于表 1。可见,与不嵌入时相比,嵌入信息后图像的 PSNR 下降很少,均不到 1 dB。这说明载密图像与载体图像的差别很小。

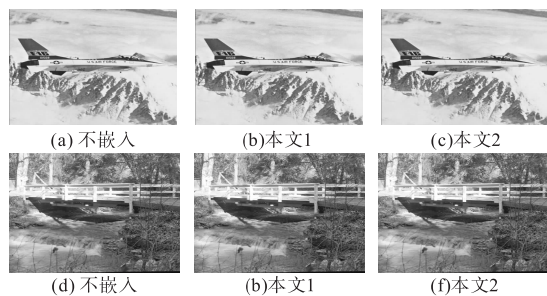


图2 信息隐藏前后的图像质量比较

对于 JPEG 压缩图像来说,其 DCT 系数直方图的统计特性通常满足:a)系数绝对值越大,其对应的直方图中的值越小,即出现频率越低;b)随着系数绝对值的升高,其出现次数下降的幅度减小。

图 3 所示是图像 baboon 的 DCT 系数直方图在信息嵌入前后的变化情况。可见,嵌入信息前后 DCT 系数直方图的统计特性是一致的。

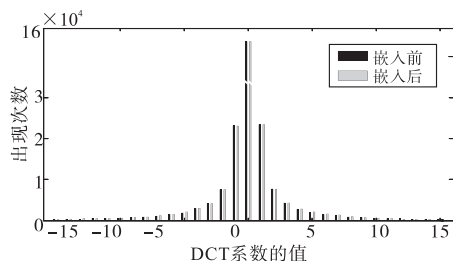


图3 信息嵌入前后的DCT系数直方图

Fridrich 等人^[12]将待检测图像删除上面 4 行或左侧 4 列,然后重新分块并对 DCT 系数以相同的量化表进行量化,以此来构造与原始图像统计特性相近的参考图像。通过比较待检测图像与估计图像之间 DCT 系数直方图和分块特性的差异,来判断待检测图像中是否含有秘密信息。为了度量本文方法对抗该攻击的效果,将载体图像、载密图像以及估计图像的每个小块中(2,3)和(3,2)位置的 DCT 系数分别取出,并统计其直方图,结果如图 4 所示。可见,不同位置的 DCT 系数直方图在不同情况下比较接近。

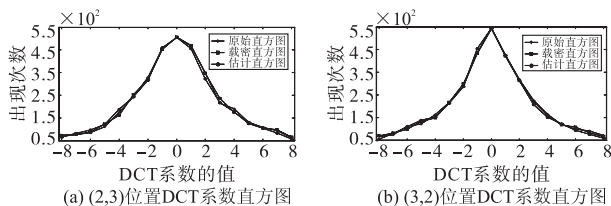


图4 不同位置的DCT系数直方图

分块特性的计算方法如下:

$$B = \sum_{i=1}^{\lfloor (M-1)/8 \rfloor} \sum_{j=1}^N |g_{8i,j} - g_{8i+1,j}| + \sum_{i=1}^M \sum_{j=1}^{\lfloor (N-1)/8 \rfloor} |g_{i,8j} - g_{i,8j+1}| \quad (3)$$

其中: M 、 N 分别是图像的高度和宽度; $g_{u,v}$ 表示像素灰度值。任意选取 100 幅不同分辨率的灰度图像,并以不小于 30 的质量因子进行压缩,然后在其中嵌入信息。分别计算载密图像的分块特性 B_s 和相应的估计图像的分块特性 B_e ,并计算两者之间的比值 $\rho = B_s/B_e$ 。图 5 显示了对这些图像计算 ρ 的结果,其

值始终保持在 1 附近。这说明载密图像与估计图像的分块特性非常接近,根据该特性难以判断图像中是否嵌入了信息。

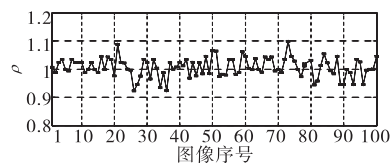


图5 载密图像与相应估计图像的分块特性之比

综合以上分析,基于纠错码编码冗余的信息隐藏方法具有较好的视觉和统计不可感知性。与利用纠错码提高鲁棒性的方法相比,其鲁棒性较差,无法应用于数字水印;与利用纠错码携带秘密信息的方法相比,其隐藏容量更高;与利用纠错码提高安全性的方法相比,其嵌入率更高。

3 结束语

在信息安全领域,信息隐藏技术的研究是一个新兴的热门课题。本文总结了基于纠错码的信息隐藏方法,并针对利用纠错码编码冗余的信息隐藏算法,给出了一种以 JPEG 压缩图像为载体的实现方法。实验结果表明,该方法具有较好的视觉和统计不可感知性,可用于秘密信息的隐藏传输。

纠错码的种类很多,根据信息隐藏的实际需要选择合适的纠错码编码类型与参数,利用数字媒体作为载体来传输秘密信息,是纠错码信息隐藏技术研究的新起点,研究并完善其他种类纠错码的信息隐藏原理,给出其具体的信息隐藏方法与应用是下一步的研究方向。

参考文献:

- [1] 王朔中,张新鹏,张开文. 数字密写和密写分析[M]. 北京:清华大学出版社,2005.
- [2] CRANDALL R. Some notes on steganography. posted on steganography mailing list [EB/OL]. (1998). <http://os.inf.tu-dresden.de/~westfeld/Crandall.pdf>.
- [3] 陈丽萍. 基于纠错码的水印技术研究[D]. 福州:福建师范大学硕士学位论文,2005.
- [4] 兰红星,陈松乔,胡爱娜,等. 基于二元树复小波域的数字水印算法[J]. 计算机应用研究, 2007,24(10):138-140.
- [5] 王天宇,刘玉君,唐冬明,等. 基于 RS 码的有扰信道信息隐藏技术的研究[J]. 信息工程大学学报,2004,5(3):85-88.
- [6] 王伟祥,刘玉君,李文雄. 利用 LDPC 码实现信道编码信息隐藏技术[J]. 信息工程大学学报, 2006,7(1):49-50, 53.
- [7] 闫雪虎. 基于纠错码的信息隐藏容量模型[J]. 计算机工程,2010,36(3):172-173,176.
- [8] WESTFELD A. F5-a steganographic algorithm; high capacity despite better steganalysis[C]//Proc of the 4th International Workshop on Information Hiding. Berlin:Springer-Verlag, 2001:289-302.
- [9] 张涛,平西建. 空域 LSB 信息伪装的隐写分析及其对策[J]. 通信学报, 2003, 24(12):156-163.
- [10] LIU C Q, PING X J, ZHANG T, et al. A research on steganography method based on error-correcting codes[C]//Proc of International Conference on Intelligent Information Hiding and Multimedia Signal Processing. 2006:377-380.
- [11] 徐长勇,平西建,刘翠卿. 安全的 MPEG 压缩视频数字隐写算法[J]. 中国图象图形学报,2009,14(11):2237-2246.
- [12] FRIDRICH J, GOLJAN M, HOGEA D. Steganalysis of JPEG images: breaking the F5 algorithm[C]//Proc of the 5th International Workshop on Information Hiding. Berlin:Springer-Verlag,2002:310-323.