文章编号: 1001-0920(2008)11-1243-06

矢量量化压缩图像中的安全隐写方法

刘光杰[a], 戴跃伟[a], 王执铨[a], 杨静宇[b]
(南京理工大学 a. 自动化学院, b. 计算机学院, 南京 210094)

摘 要: 提出一种用于矢量量化压缩图像的安全数据隐藏方案. 为降低数据嵌入引入的失真,以码字间的矢量均方差为优化指标,采用遗传算法实现码本的优化分割,并提出基于码本分割的数据嵌入算法. 采用基于自适应算术熵解码的数据映射方法,实现了嵌入前后统计特性的保持. 实验结果表明,所提出的算法在容量、失真水平和安全性方面具有较好的综合性能.

关键词: 数据隐藏; 矢量量化; 遗传算法; 相对熵
中图分类号: TP391 **文献标识码**: A

Secure steganographic method for vector quantization-compressed images

LIU Guang-jie[a], *DAI Yue-wei*[a], *WANG Zhi-quan*[a], *YANG Jing-yu*[b]
(a. Automation School, b. Computer School, Nanjing University of Science and Technology, Nanjing 210094, China. Correspondent: LIU Guang-jie, E-mail: guangj_liu@yahoo.com.cn)

Abstract: A secure data hiding scheme for vector quantization (VQ)-compressed images is proposed in this paper. To decrease the distortion caused by data embedding, the vector mean square error between codewords is taken as optimization index, the genetic algorithm is used to realize the partition of codebook, and the data embedding approach is given based on codebook partition. Meanwhile, the adaptive-arithmetic-decoder-based data mapping method is used to keep the statistical characteristics before and after data hiding. The experimental results show that the proposed scheme has better combined performance in capacity, distortion level and security.

Key words: Data hiding; Vector quantization; Genetic algorithm; Relative entropy

1 引 言

数据隐藏是通过由载体信号建立的隐藏信道进行秘密通信的技术,它包含 3 个要素:感知质量、隐藏容量与安全性. 高效的隐藏算法应保证在不破坏原始载体的感知质量的条件下,安全地隐藏更多的消息比特. 矢量量化[1](VQ)是一种有损的压缩技术,广泛用于低码率图像和音频数据的压缩. 如何更好地实现矢量量化压缩图像中的信息隐藏是十分重要的课题.

Lu[2]提出一种基于矢量量化的数字水印算法,该算法通过基于失真受限的禁忌搜索方法实现码本的分割,并通过量化嵌入机制将水印比特嵌入图像块索引值中. Lu[3]还提出通过对当前码本中码字施加扰动以获得一个扩展的码本,并利用扩展码本实现信息隐藏. 该方法的缺点是,码本的扩展对图像的压缩效率产生了较大的影响. 文献[4]提出一种基于多阶段矢量量化和多用途数字水印的算法. Wang 等人[5]提出一种基于码本分割的数据隐藏方案,该方案将数据嵌入索引值中,但文中仅提到使用密钥控制码本的分割,而没有指出如何控制码本分割带来数据嵌入的失真问题. 对此,Wang 等人[6]提出将码本分割成用来编码 0 和 1 的子码本,并以数据嵌入后的 PSNR 值作为优化指标,采用遗传算法实现码本的优化分割. 该方法的主要缺陷是:码本分割设计的优化过程要进行信息隐藏的相关计算,多次尝试嵌入的机制导致算法的时间性能较差. Wu 和 Chang[7]设计了一种将码本分割成具有 2 元素子码本的码本分割算法,同时采用量化嵌入方法将数据

收稿日期: 2007-09-08; **修回日期**: 2008-03-28.
基金项目: 国家自然科学基金项目(60374066); 中国博士后基金项目(20070421017).
作者简介: 刘光杰(1980—),男,江苏徐州人,讲师,博士后,从事信息隐藏、多媒体认证的研究; 戴跃伟(1962—),男,江苏镇江人,教授,博士生导师,从事信息安全的研究.

嵌入索引值. 然而,文献[7]中的码本分割方法使用的是较为主观的阈值比较方法,不能较好地解决嵌入引起的失真问题. 为增加嵌入容量,文献[8]提出一种可在两个矢量量化索引值中嵌入 3 个消息比特的方案. 该方案有效地增加了嵌入数据的容量,但由于要为每个图像块分配两个索引值,方案的整体性能并没有太多的提高. 文献[9]提出一种自适应的数据嵌入方法. 该方法根据要嵌入的比特数据将码本分割成大小不同的子码本,该方法实际上是一种基于非一致量化器的量化嵌入策略,在一定程度上较好地平衡了容量和失真之间的矛盾.

以上算法涉及到对矢量量化码本的分割,分割性能决定着最终的失真水平. 为避免码本分割算法的复杂性,Chang[10]提出一种基于主成分分析的码本排序策略,并将数据直接嵌入索引值的 LSB 中. 由于矢量量化会带来较明显的块效应和边缘效应,为保证数据嵌入后图像的感知质量,Chang[11]提出将信息嵌入基于搜索次序的矢量编码中. Chang[12,13]还提出将数据嵌入边缘匹配矢量量化编码中的方法. 文献[2-13]提出的方法主要侧重于平衡嵌入容量和嵌入失真之间的关系,并没有考虑嵌入数据抵抗统计隐写分析(Steganalysis)的安全性问题. 为此,本文提出用于矢量量化压缩图像的安全数据隐藏方案. 该方案基于具有统计特性保持的熵解码嵌入策略和基于基因算法的优化码本分割算法.

2 数据嵌入算法

2.1 矢量量化编码基础

矢量量化编码是一种低比特率的有损压缩技术,已被成功地用于图像和音频数据的压缩. 为了编码一个具有 $H \times W$ 个像素的灰度图像 I,将其分割成互不重叠的具有 $l \times l$ 个像素的图像块,图像压缩一般取 $l = 4$,将图像块组织成矢量 x,共 $N = H \times W/16$ 个 16 维度的矢量. 矢量量化可看作是从 R^{l^2} 空间到其有限子空间 Y 的一个映射. 这里 $Y = \{y_i, i = 1,2,\cdots,M\}$ 为编码本. 矢量量化定义了 R^{l^2} 空间的一个划分 $S = \{S_1, S_2, \cdots, S_{l_2}\}$,满足

$$S_i = \{x \in R^{l^2} : VQ(x) = y_i\}. \quad (1)$$

给定输入图像块对应的矢量 x,根据码本 Y,可用距 x 最近的码字 y_i 的索引值 i 来编码 x,若码本的大小为 M,则索引值可用 $\log_2 M$ 个比特表示. 解码端根据索引值 i 以及码本 Y,恢复相应的码字来代替原来的图像块 x. 设解码得到的图像为 I′,原始图像 I 所有的像块矢量为 $x_i (i = 1,2,\cdots,N)$,矢量量化引入的 MSE 失真可表示为

$$D(I, I') = \frac{1}{N \times l^2} \sum_{i}^{N} \| x_i - y_{VQ(x_i)} \|^2. \quad (2)$$

其中:$y_{VQ(x_i)}$ 表示 x_i 对应的码字,$\| \cdot \|$ 为欧式距离. 码本 Y 的设计是矢量量化编码中最为重要的部分,目前已有很多矢量量化码本设计方法,本文中使用的码本生成方法是最为常用的 LBG[1] 算法.

2.2 基于码本分割的数据嵌入方法

码本中相邻索引位置的码字并不一定具有较小的欧式距离,因此,若直接在索引值上进行类似于 LSB 的修改,将会对图像质量产生极大的破坏. 为了使索引值能承载消息比特,又不对嵌入后的图像质量产生较大的破坏,必须保证索引值对应的码字的变化在一个可接受的范围之内. 基于这样的考虑,文献[2-7,9]中采用了码本分割的思想实现数据的嵌入,即将部分码字组合成一些距离相近的子码本,数据嵌入改变后的索引值只能在原来索引所在子码本中进行.

设图像 I 对应的码本为 Y,码本的大小 $M = 2m$,本文将码本分割为 m 个大小为 2 的子码本 $\phi_1, \phi_2, \cdots, \phi_m$. 其中 $\phi_i = (y_{i0}, y_{i1})$,且满足如下分割条件:

$$\bigcup_{i=1}^{m} \phi_i = Y, \ \phi_i \cap \phi_j = \varnothing, \ \forall i \neq j. \quad (3)$$

设图像块 $B(i,j)$ ($i \in \{1,2,\cdots,H/l\}$, $j \in \{1,2,\cdots,W/l\}$) 对应的码字为 y,其索引为 u,并设 $y \in \phi_k$. 若嵌入的数据比特 $w = 0$,则将 y 对应的索引值 u 映射为 y_{k_0} 对应的索引值 u_{k0};若 $w = 1$,则将 u 映射为 u_{k1}. 解码端只需知道码本的分割方法,即判断嵌入在索引中的比特值. 为实现解码端的盲提取,可先将原始的码本 Y 以 $(y_{10}, y_{11}, \cdots, y_{i0}, y_{i1}, \cdots, y_{m0}, y_{m1})$ 的形式进行重排序得到码本 Y′,并根据新码本 Y′ 更新原始图像的各个索引值 u;然后,在更新后的索引值中进行数据嵌入,嵌入后索引值的奇偶性即代表嵌入的消息比特.

设矢量量化压缩图像块对应的码字为 $y_i (i = 1,2,\cdots,N)$,$y_i \in Y$,经过数据嵌入后图像块对应的码字为 $y_i' (i = 1,2,\cdots,N)$,$y_i' \in Y$,则可定义如下矢量均方差嵌入失真(VMSE):

$$D = \frac{1}{N} \sum_{i}^{N} \| y_i - y_i' \|^2. \quad (4)$$

若原始压缩图像对应码字 $(y_{10}, y_{11}, \cdots, y_{i0}, y_{i1}, \cdots, y_{m0}, y_{m1})$ 的索引值的统计分布为 $(p_{10}, p_{11}, \cdots, p_{i0}, p_{i1}, \cdots, p_{m0}, p_{m1})$,设所有索引值上均进行了数据的嵌入,且设嵌入具有码字 y_{i0}, y_{i1} 中的消息比特 0 和 1 的概率为 q_i 和 $1 - q_i$,则根据上面描述的基于子码本的数据嵌入方法,可将式(4)重写为

$$D = \sum_{i}^{m} (p_{i0}(1 - q_i) + p_{i1} q_i) \| y_{i0} - y_{i1} \|^2. \quad (5)$$

3　安全数据隐藏方法

3.1　统计安全性指标

Cachin[14] 提出了一种信息隐藏安全性分析的信息论模型. 该模型给出了在基于统计分析的假设检验下隐写系统安全性的度量方法. 设载体信号集合为 C,其上的概率分布为 P_c,隐藏后的信号集合为 S,其上的概率分布为 P_s. 若假设攻击方最多只能知道关于 P_c, P_s 的信息,则隐写系统的安全性可定义为在同一集合上的两个分布的相对熵

$$D(P_c \| P_s) = \sum_{x \in C \cap S} P_c(x) \log(P_c(x)/P_s(x)).$$

当 $D(P_c \| P_s) = 0$ 时,隐写系统具有绝对安全性;当 $D(P_c \| P_s) = \varepsilon$ 时,隐写系统 ε 为安全.

设数据嵌入前后对应不同码字 $Y = \{y_i, i = 1, 2, \cdots, M\}$ 的图像块的统计分布为 P_c 和 P_s,可定义如下安全性指标:

$$J = D(P_c \| P_s). \tag{6}$$

保证指标 J 尽可能小需在数据嵌入的同时尽可能保证嵌入数据后的分布同原始的分布保持一致. 数据嵌入前后的统计特性一般采用两种方法. 方法一是使用部分载体信号承载消息比特,通过改变另部分信号校正被数据嵌入改变了的统计分布. Provos 的 Out Guess[15] 算法是这类方法的最早尝试,文献[16]采用了类似的思想并结合 QIM 嵌入算法实现对统计特性的补偿;方法二是要求数据嵌入算法本身具有统计分布保持的特性,这类算法中比较有代表性的是文献[17]提出的基于直方图保持数据映射(HPDM)和切换数据映射(SDM)的方法和文献[18]提出的基于模型的隐写算法(MBS). 张新鹏等人[19] 也提出了一种通过控制随机加减 1 比率的嵌入算法,并将其用于 BMP 和 JPEG 图像信息隐藏中. 在这两类算法中,第一类算法因为补偿的需要而减少了嵌入的有效数据量,因此在需较大容量的设计要求下,算法性能要差于第二类算法. 事实上,文献[18]提出的基于模型的隐写算法思想给出了一种通用安全嵌入算法的设计思路. 本文主要借鉴文献[18]中基于熵解码数据分布保持方法,但与其不同的是,本文将统计分布信息混同要嵌入的消息比特一起嵌入载体图像中,而无需关于系数统计模型的建立和参数估计.

3.2　基于自适应算法解码器的统计分布保持方法

设经过码本分割并重排序的码本为 $Y' = (y_{10}, y_{11}, \cdots, y_{i0}, y_{i1}, \cdots, y_{m0}, y_{m1})$,且设对应码字 $(y_{10}, y_{11}, \cdots, y_{i0}, y_{i1}, \cdots, y_{m0}, y_{m1})$ 的索引值的统计分布为 $(p_{10}, p_{11}, \cdots, p_{i0}, p_{i1}, \cdots, p_{m0}, p_{m1})$,设数据嵌入所有的字 y_{i0}, y_{i1} 对应的索引值中,并设比特 0 和 1 的分布概率分别为 q_i 和 $1 - q_i$,记经过数据嵌入后码字 $(y_{10}, y_{11}, \cdots, y_{i0}, y_{i1}, \cdots, y_{m0}, y_{m1})$ 对应索引值的统计分布为 $(p_{10}', p_{11}', \cdots, p_{i0}', p_{i1}', \cdots, p_{m0}', p_{m1}')$. 根据 2.2 节描述的数据嵌入算法, p_{i0}', p_{i1}' 和 p_{i0}, p_{i1} 之间应具有如下关系:

$$\begin{aligned} p_{i0}' &= q_i(p_{i0} + p_{i1}), \\ p_{i1}' &= (1 - q_i)(p_{i0} + p_{i1}). \end{aligned} \tag{7}$$

根据式(7),式(6)可重新写为

$$J = \sum_{i=1}^{m} \left[p_{i0} \log\left(\frac{p_{i0}}{p_{i0} + p_{i1}} \cdot \frac{1}{q_i} \right) + p_{i1} \log\left(\frac{p_{i1}}{p_{i0} + p_{i1}} \cdot \frac{1}{1 - q_i} \right) \right]. \tag{8}$$

若使式(8)取最小值 0,只需满足

$$q_i = \frac{p_{i0}}{p_{i0} + p_{i1}}. \tag{9}$$

嵌入的消息比特一般需经通过压缩和加密的操作,因此消息比特流一般服从均匀分布,为使嵌入到 y_{i0}, y_{i1} 对应索引 u_{i0}, u_{i1} 中的比特流具有 $(q_i, 1 - q_i)$ 的 0-1 分布,需要一个能产生具有指定长度和指定分布特性的可逆映射 F. 文献[18]提出的经过修改的自适应算术编码的解码器可自适应地将源比特流中读取的 r 比特映射为具有指定分布 $(q, 1 - q)$,指定码长 t 的 0-1 序列. 指定算术编码器的参数 $(q, 1 - q)$,输入长度为 t 的 0-1 序列可准确无误地恢复原来的 r 比特数据,自适应算法编解码器的这些性质正好满足对可逆映射 F 的要求.

假设码字 $(y_{10}, y_{11}, \cdots, y_{i0}, y_{i1}, \cdots, y_{m0}, y_{m1})$ 对应编码索引的统计直方图的频数为 $(h_{10}, h_{11}, \cdots, h_{i0}, h_{i1}, \cdots, h_{m0}, h_{m1})$,将消息序列 W 嵌入 VQ 编码索引值中的过程如下:

Step1:首先根据 3.3 节提出的算法对码本进行分割、重排,并更新压缩编码中对应各个码字的索引值,令 $i = 1$.

Step2: 若 $i > m$,则终止迭代,进入 Step5;否则,统计具有码字 y_{i0} 和 y_{i1} 的共 $h_{i0} + h_{i1}$ 个编码索引,计算要嵌入在这些编码索引中的"0"比特的概率 $q_i = h_{i0}/(h_{i0} + h_{i1})$,将其用 10 位的二进制小树表示(如将 0.252 通过取整 0.252×2^{10} 并通过 10 进制到 2 进制的转化变成"0100000010").

Step3: 将参数 $(q_i, 1 - q_i)$ 输入到解码器 F 中,指定输出 0/1 序列长度 $h_{i0} + h_{i1} - 10$,解码器从 W 中自适应地读入 r_i 个比特.

Step4: 将 q_i 对应的 10 位小数和 $h_{i0} + h_{i1} - 10$ 个消息比特在密钥 K_1 的控制下,用 2.2 节中介绍的方法嵌入 $h_{i0} + h_{i1}$ 个编码索引中.

Step5: 将消息序列 W 的读取指针后移 r_i 个位

置,并计 $i = i + 1$,返回 Step2.

Step6: 将码本中码字的位置在密钥 K_2 的控制下重新打乱,并根据置乱的关系更新编码中的所有索引值.

信息提取的过程可看作是嵌入过程的反过程,首先根据密钥 K_2 恢复码本顺序;其次,对嵌入每个子码本 ϕ_i 码字索引中的数据根据 K_1 进行逐一提取;然后,根据提取出的 q_i,通过算术编码器重新恢复嵌入的数据;最后,合并各个数据段得到最终的秘密消息数据.

3.3 基于遗传算法的最优码本分割方法

结合式(5)和(9),可得到数据嵌入引入的 VMSE 具有如下形式:

$$D = \sum_{i}^{m}\left[\frac{2p_{i0}p_{i1}}{p_{i0}+p_{i1}}\|y_{i0}-y_{i1}\|^2\right]. \tag{10}$$

面向最小化 VMSE 的码本最优分割是一个组合优化问题. 考虑尺寸 M 的码本,以式(3)的形式进行分割,则可能的分割方式为 $2^{-M/2-2}M!$,对只有 128 个码字的码本而言,这个数目也高达 2.0905×10^{196}. 对于这类高维度的组合优化问题,一般采用具有全局优化能力的随机搜索方法,如模拟退火、遗传算法或蚁群算法. 本文采用遗传算法进行码本分割的优化求解,下面讨论遗传算法涉及到的 3 个核心问题.

(1) 解的编码方式

遗传算法一般采用序号编码而并非一般的 0/1 序列编码,并通过和问题背景相关的遗传算子来保证解的合理性. 设未分割码本为 $Y=(y_1,y_2,\cdots,y_{2m})$,其对应的分布为 $P=(p_1,p_2,\cdots,p_{2m})$,解定义为 $Z=(z_1,z_2,\cdots,z_{2m})$,$z_i\in\{1,2,\cdots,2m\}$,$z_i\neq z_j$,$\forall i\neq j$. Z 可看作是码字索引的全排列,$z_i$ 指代码字 y_{z_i},(z_{2i-1},z_{2i}) 表示 Y 的一个子码本. 因此,1 到 $2m$ 间的整数排列就对应着划分 Y 的解.

(2) 适应函数的选取

根据式(10),D 可看作是 Z 的函数,记为 $D(Z)$,适应函数可定义为

$$f(Z) = D_{MAX} - D(Z). \tag{11}$$

其中

$$D(Z)=\sum_{i}^{m}\frac{2p_{z_{2i-1}}p_{z_{2i}}}{p_{z_{2i-1}}+p_{z_{2i}}}\|y_{z_{2i-1}}-y_{z_{2i}}\|^2,$$

$$D_{MAX}=\max_{i\neq j}\|y_i-y_j\|^2.$$

容易证明对所有的 Z,有 $D(Z)\leq D_{MAX}$.

(3) 遗传算子

本文选用的遗传算子分别为选择算子、交叉算子和变异算子,它们定义如下:

1) 选择算子

设每一代的种群规模为 N,种群的选取采用轮盘赌的方法,个体 Z_j 被选中的概率为

$$p_p = f(Z_j)/\sum_{k}^{N}f(Z_k). \tag{12}$$

2) 交叉算子

本文采用双亲双子的交叉算子. 设父代个体为 $Z_1=(z_1^1,z_2^1,\cdots,z_{2m}^1)$,$Z_2=(z_1^2,z_2^2,\cdots,z_{2m}^2)$,在交叉中任取 Z_1 中的一个子码本 $(z_{2k_1-1}^1,z_{2k_1}^1)$,从 Z_2 中找到包含码字 $z_{2k_1-1}^1,z_{2k_1}^1$ 的两个子码本 $(z_{2k_{21}-1}^2=z_{2k_1-1}^1,z_{2k_{21}}^2)$ 和 $(z_{2k_{22}-1}^2=z_{2k_1}^1,z_{2k_{22}}^2)$;然后,将这两个子码本交换,得到两个码本 $(z_{2k_{21}-1}^2,z_{2k_{22}-1}^2)$ 和 $(z_{2k_{21}}^2,z_{2k_{22}}^2)$. 交换操作产生的子代个体 Z_2' 包含了 Z_2 的大部分子码本以及 Z_1 的子码本 $(z_{2k_1-1}^1,z_{2k_1}^1)$,用同样的方法从 Z_2 中随机选取子码本 $(z_{2k_2-1}^2,z_{2k_2}^2)$,对 Z_1 中相应的两个子码本进行交换,即产生 Z_1'. 此交叉算子可迭代使用多次作为两个个体的总体交叉.

3) 变异算子

设个体为 $Z=(z_1,z_2,\cdots,z_{2m})$,它的变异可描述为:任取 Z 的一个子码本 (z_{2k-1},z_{2k}),从 $y_{z_{2k-1}}$ 的 θ 个最近邻码字中任意选择码字 y_j,从 Z 中找到 y_j 所在的子码本 $[z_{2k'-1}(y_{z_{2k-1}}=y_j),z_{2k'}]$. 将 (z_{2k-1},z_{2k}) 和 $(z_{2k'-1},z_{2k'})$ 交换产生两个新的子码本 $(z_{2k-1},z_{2k'-1})$ 和 $(z_{2k},z_{2k'})$,代替原来的两个子码本即可得到个体 Z 的变异 Z′,这一过程可迭代进行,完成在一次变异中多个子码本的改变.

参数 θ 是一个可调整的量,θ 较小时算法收敛较快,但会出现早熟现象而停止于局部最优值,θ 过大会影响收敛的速度. 这里取 $\theta = m/8$,即码本大小的 1/16.

本文采用选择 10 % 的最优个体,利用选择和交叉算子生成另外 90 % 的新个体;然后,在所有新个体上进行变异生成新一代种群. 这样的操作有利于保持一些最优基因不会因为过度交叉而产生破坏. 根据上面给出的编码方式和遗传算子,给出如下的基于遗传算法的码本分割流程:

Step1: 设置初始种群的大小为 $N=120$,令 $i=1$,表示第 1 代种群,任取 $(1,2,\cdots,2m)$ 的 120 种排序方式作为第 1 代种群个体 pop(1),令 $i=i+1$.

Step2: 若 $i>\text{MaxGen}$,则终止迭代,进入 Step5;否则,从 pop$(i-1)$ 的 120 个个体中选择 12 个适应值最高的个体直接遗传至 pop(i).

Step3: 根据式(12),通过轮盘赌方式选择 54 对个体通过上面给出的交叉算子进行交叉(迭代 4 次,交叉概率 $p_c=100\%$),汇同 12 个直接遗传的个体组成 120 个新个体.

Step4: 对新个体以概率 $p_m=0.1$ 进行变异(迭

代 2 次),得到第 i 代种群 pop(i),令 $i = i + 1$,转到 Step2.

Step5: 输出当前 pop 中适应性最好的解作为码本的最优分割.

4　实验及结果分析

为测试本文提出的算法,对 4 幅标准 512 ×512 的灰度测试图像 Lena, Plane, Peppers, Boat 进行测试. 图像分块大小为 4 ×4,码本的尺寸为 256,通过 LBG 算法得到每个图像的码本;然后,进一步进行索引值查找产生压缩编码图像. 嵌入的数据 W 为由随机数发生器产生的足够长的服从均匀分布的 0/1 序列. 在所有索引值中进行全嵌入,根据嵌入数据的长度表示最大隐藏容量. 图 1 给出了隐写后的 Lena 和 Boat 图像,对于 256 大小码本的矢量压缩图像,图 1 的结果还是能接受的.

(a) Lena隐写图像

(b) boat 隐写图像

图 1　Lena 和 Boat 对应的隐写图像

表 1　本文算法的嵌入容量和嵌入失真

性能	Lena	F16	Peppers	Boat
容量/bit	12775	13161	13147	13318
VMSE	794.6	982.4	927.5	1172.8
$PSNR_1$/dB	28.60	27.87	27.81	26.54
$PSNR_2$/dB	31.65	30.58	31.21	29.08

表 1 给出了采用本文设计算法在 4 幅图像中嵌入数据的最大容量;相对于矢量压缩图像的矢量均方差(VMSE);相对于原始未压缩图像的峰值信噪比(PSNR1);作为比较,本文还列出了矢量量化图像相对于原始图像的 PSNR2. 由表 1 可知,优化码本分割的隐写仅引起了 3 dB 左右的质量损失.

同已有方法相比,本文算法在设计时引入了统计特性保持的机制. 为了进行比较,在嵌入过程中略去熵解码的过程,直接将 W 嵌入索引值中. 表 2 给出了具有熵解码过程的安全指标 J_1 和未使用熵解码过程的安全指标 J_2.

表 2　安全性指标比较

J	Lena	F16	Peppers	Boat
J_1	0.007443	0.007302	0.009341	0.006577
J_2	0.149221	0.118035	0.112083	0.108656

表 2 表明,本文算法具有很好的一阶统计特性保持能力,能抵抗基于一阶统计特性的隐写分析方法.

5　结　　论

本文提出一种用于矢量量化压缩图像隐写方法. 设计了基于码本分割的数据嵌入方法,为使嵌入失真尽可能小,将遗传算法用于码本的优化分割. 同以往矢量量化中的信息隐藏方法不同,本文将抗统计隐写分析的安全性要求引入了算法的设计中,采用了基于自适应算术熵解码的数据映射方法,将要嵌入的数据消息比特映射为具有指定分布特性的 0/1 序列. 实验证明,本文提出的算法在容量、失真水平和安全性方面具有较好的综合性能.

本文方案采用的基于码本分割的嵌入算法本质上是一种量化嵌入算法. 分割后的子码本都具有相同的尺寸,因此嵌入是一致量化的,尚不能达到最佳的综合性能. 另外,若分析者使用包含更高阶统计特性的隐写分析方法,则仅具有一阶安全性的方法并不够安全. 因此,在后续研究中,将考虑具有高阶保持能力的更为安全和基于非一致性量化的更高容量、更低失真的嵌入方法.

参考文献(References)

[1] Linde Y, Buzo A, Gray R M. An algorithm for vector quantizer design[J]. IEEE Trans on Communications, 1980, 28(1): 84-95.

[2] Lu Z M, Sun S H. Digital image watermarking technique based on vector quantization[J]. Electronics Letters, 2000, 36(4): 303-304.

[3] Lu Z M, Pan J S, Sun S H. VQ-based digital image watermarking method[J]. Electronics Letters, 2000, 36(14): 1201-1202.

[4] Lu Z M, Xu D G, Sun S H. Multipurpose image watermarking algorithm based on multistage vector quantization[J]. IEEE Trans on Image Processing, 2005, 14(6): 822-831.

[5] Wang F S, Pan J S, Jain L C, et al. A VQ-based image-in-image data hiding scheme[C]. Proc of 2004 IEEE Int Conf on Multimedia and Expo. Taipei: IEEE, 2004: 2191-2194.

[6] Wang F S, Jain L C, Pan J S. VQ-based watermarking scheme with genetic codebook partition[J]. J of Network and Computer Applications, 2007, 30(1): 4-23.

[7] Wu H S, Chang C C. A novel digital image watermarking scheme based on the vector quantization technique[J]. Computers and Security, 2005, 24(6): 460-471.

[8] Yu Y H, Chang C C, Hu Y C. A steganography

method for hiding data in VQ encoded images[C]. Proc of 2004 Int Symposium on Intelligent Multimedia, Video and Speech Processing. Hong Kong: IEEE, 2004: 358-361.

[9] Du W C, Hsu W J. Adaptive data hiding based on VQ compressed images[J]. IEE Proc Visual Image Signal Processing, 2003, 150(4): 233-238.

[10] Chang C C, Lin P Y. A compression-based data hiding scheme using vector quantization and principle component analysis[C]. Proc of 2004 Int Conf on Cyberworlds. Tokyo: IEEE, 2004: 369-375.

[11] Chang C C, Chen G M, Lin M H. Information hiding based on search-order coding for VQ indices[J]. Pattern Recognition Letters, 2004, 25(11): 1253-1261.

[12] Chang C C, Tai W L, Lin M H. A reversible data hiding scheme with modified side match vector quantization[C]. Proc of the 19th Int Conf on Advanced Information Networking and Application. Washington: IEEE, 2005: 947-952.

[13] Chang C C, Lu T C. Reversible index-domain information hiding scheme based on side-match vector quantization[J]. J of Systems and Software, 2006, 79(8): 1120-1129.

[14] Cachin C. An information-theoretic model for steganography[J]. Information and Computation, 2004, 192(1): 41-56.

[15] Provos N. Defending against statistical steganalysis[C]. Proc of the 10th USENIX Security Symposium. Washington: USENIX, 2001: 323-336.

[16] Solanki K, Sullivan K, Madhow U, et al. Statistical restoration for robust and secure steganography[C]. Proc of IEEE Int Conf on Image Processing. Genova: IEEE, 2005: 1118-1121.

[17] Eggers J J, Bauml R, Girod B. A communications approach to image steganography[C]. Proc of SPIE Security and Watermarking of Multimedia Contents IV. San Jose: SPIE, 2002, 4675: 26-37.

[18] Sallee P. Model-based steganography[C]. Lecture Notes in Computer Science: IWDW2003. Seoul: Springer, 2004, 2939: 154-167.

[19] Zhang X P, Wang S Z, Zhang K W. Steganography with least histogram abnormality[C]. Lecture Notes in Computer Science: MMM-ACNS2003. Petersburg: Springer, 2003, 2776: 395-406.

(上接第 1242 页)

参考文献(References)

[1] 张勇德, 黄莎白. 多目标优化问题的蚁群算法研究[J]. 控制与决策, 2005, 20(2): 170-173.
(Zhang Y D, Huang S B. On ant colony algorithm for solving multiobjective optimization problems[J]. Control and Decision, 2005, 20(2): 170-173.)

[2] Kennedy J, Eberhart R C. Particle swarm optimization[C]. Proc IEEE Int Conf on Neural Networks. Piscataway: IEEE Service Center, 1995: 1942-1948.

[3] 谢晓锋, 张文俊, 杨之廉. 微粒群算法综述[J]. 控制与决策, 2003, 18(2): 129-134.
(Xie X F, Zhang W J, Yang Z L. Overview of particle swarm optimization[J]. Control and Decision, 2003, 18(2): 129-134.)

[4] Coello C A C, Lechuga M S. MOPSO: A proposal for multiple objective particle swarm optimization[C]. Proc IEEE Int Conf on Evolutionary Computation. Piscataway: IEEE Service Center, 2002, 2: 1051-1056.

[5] Hu X, Eberhart R. Multiobjective optimization using dynamic neighborhood particle swarm optimization[C]. Proc IEEE Int Conf on Evolutionary Computation. Honolulu, 2002, 2: 1677-1681.

[6] Fieldsend J E, Singh S. A multi-objective algorithm based upon particle swarm optimization, an efficient data structure and turbulence[C]. Proc UK Workshop on Computational Intelligence. Birmingham, 2002: 37-44.

[7] Coello C A C, Pulido G T, Lechuga M S. Handling multiple objectives with particle swarm optimization[J]. IEEE Trans on Evolutionary Computation, 2004, 8(3): 256-279.

[8] Reyes-Sierra M, Coello C A C. Multi-objective particle swarm optimizers: A survey of the state-of-the-Art[J]. Int J of Computational Intelligence Research, 2006, 2(3): 287-308.

[9] Ratnaweera A, Halgamuge S K, Watson H C. Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients[J]. IEEE Trans on Evolutionary Computation, 2004, 8(3): 240-255.

[10] Deb K, Pratap A, Agarwal S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-Ⅱ[J]. IEEE Trans on Evolutionary Computation, 2002, 6(2): 182-197.

[11] Van Veldhuizen D A, Lamont G B. Multiobjective evolutionary algorithm research: A history and analysis[R]. Ohio: Air Force Institute of Technology, 1998.

[12] Schott J. Fault tolerant design using single and multicriteria genetic algorithm optimization[D]. Cambridge: Massachusetts Institute of Technology, 1995.