

文章编号: 1672-2892(2010)02-0180-06

基于蜜蜂进化算法的 JPEG 隐写方法

戴少华, 王建军

(复旦大学 电子工程系, 上海 200433)

摘 要: 为提高嵌入效率, 缩短算法运行时间, 提出了一种基于蜜蜂进化算法的变换域隐写方法。该方法首先利用由蜜蜂进化算法搜索到的最优置换矩阵将秘密信息进行置换; 接着修改了 JPEG 标准量化表以便能够嵌入较多的信息; 将置换结果嵌入到量化后 DCT 系数的 DC 和中低频 AC 区域中; 最后利用 JPEG 熵编码得到隐藏了信息的 JPEG 格式文件。实验结果表明, 与直接嵌入方法相比较, 该方法提高了含密图像的质量和安全性, 与基于遗传算法的信息隐藏方法相比, 该算法所花费时间更少。

关键词: 信息隐藏; 蜜蜂进化型遗传算法; JPEG 隐写; 量化表

中图分类号: TN911.73

文献标识码: A

JPEG steganography based on Bee Evolutionary Genetic Algorithm

DAI Shao-hua, WANG Jian-jun

(Department of Electronics Engineering, Fudan University, Shanghai 200433, China)

Abstract: In order to improve the embedding efficiency and reduce the operating time, a novel steganographic method, based on JPEG and Bee Evolutionary Genetic Algorithm(BEGA), was proposed. The proposed method first derived an optimal substitution matrix by BEGA for transforming the secret messages. By means of the substitution strategy, the quality of stego-images was improved. Next, it modified the standard JPEG quantization table for the purpose of accommodating more secret messages. The transformed messages were then hidden in the cover-image with its DC-to-middle frequency components of the quantized DCT coefficients modified. Finally, a JPEG file was generated through JPEG entropy coding. The experimental results showed that the proposed method had achieved a better image quality and security than steganographic method without optimal substitution, and it had costed less computation time than the genetic algorithm-based information hiding method.

Key words: data hiding; Bee Evolutionary Genetic Algorithm; JPEG steganography; quantization table

数据隐藏是将秘密信息隐藏在掩体介质内的一种方法, 以便使没有授权的用户无法意识到隐藏信息的存在。数字水印和隐写术是它的 2 个主要应用^[1], 前者注重鲁棒性和安全性, 而后者注重嵌入容量和视觉不可见性。本文的研究属于隐写术的范畴。常见的隐写嵌入算法有时空域和变换域两大类: 时空域算法多使用最低有效位(Least Significant Bits, LSB)技术, 将图像的最低位平面用秘密信息位替换以达到嵌入的目的; 变换域隐写算法的研究主要集中在离散傅里叶变换域(DFT)、离散余弦变换域(DCT)和离散小波变换域(DWT)等, 将掩体信息变换到频率域中进行嵌入^[2]。JPEG 是一种常用的图像文件格式^[3], 如果使用 JPEG 图像作为数据隐藏的载体, 得到含有秘密数据的 stego 图像将不容易引起拦截(攻击)者的注意^[4]。例如, 著名的信息隐藏软件 Jpeg-Jsteg 就是基于 JPEG 图像格式的信息隐藏。在时空域隐写算法研究领域, Wang 等人^[5]改进了一般的 LSB 替换方法, 提出了一种基于遗传算法的最优 LSB 替换, 即用遗传算法搜索出一个最优的映射, 然后将秘密信息通过该映射进行信息置换后再嵌入到掩体图像。Li 等人^[6]提出了基于粒子群的寻找最优替代隐写算法。蜜蜂进化算法是近年来新提出的一种基于群体智能的优化算法, 它通过模仿蜂群独特的觅食、婚飞等行为来求解复杂的优化问题。在搜索效率和收敛速度上, 蜜蜂进化算法(BEGA)都优于或近似于同类的进化算法^[7-8]。目前, 将 BEGA 算法应用到信息隐藏领域的研究工作还比较少, 因此本文提出了一种基于 BEGA 算法^[7]的 JPEG 隐写方法, 该方法利用 BEGA 算法快速搜索

到一个理想置换矩阵将秘密信息置换, 并且修改 JPEG 标准量化表, 将置换结果嵌入到量化后的 DCT 系数的直流量(Direct Current, DC)和中低频交流分量(Alternating Current, AC)区域中, 分别用全局优化和分块优化方法加以实现。实验表明, 该方法得到的 stego 图像质量较直接嵌入方法更好, 且具有较高的安全性, 与基于遗传算法的信息隐藏方法相比, 该算法所花费时间更少。

1 Jpeg-Jsteg 及最优 LSB 替换算法

1.1 Jpeg-Jsteg 及其改进算法

Jpeg-Jsteg 是一种基于 JPEG 的常用信息隐藏工具。首先把掩体图像分为不重叠的 8×8 子块, 对每个子块进行 DCT 变换, 并对变换得到的 DCT 系数进行量化, 所采用的量化表见图 1 的 JPEG 标准量化表; 其次, Jpeg-Jsteg 算法将待隐藏的信息进行加密, 将加密结果嵌入到量化后值不为 0,1 或 -1 的 DCT 系数的 LSB 中, 其嵌入顺序是按图 2 的 zigzag 扫描顺序进行; 最后, 用 JPEG 的熵编码(包括哈夫曼编码、游程编码及差分脉冲编码调制(Differential Pulse Code Modulation, DPCM))对嵌入秘密信息后的每一子块进行编码, 从而得到一个含有秘密信息的 JPEG 文件。

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Fig.1 JPEG standard quantization table
图 1 标准 JPEG 量化表

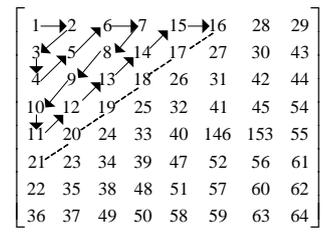


Fig.2 Zigzag scan order
图 2 Zigzag 扫描顺序

Jpeg-Jsteg 算法能隐藏的信息量很有限, 因为经过标准 JPEG 量化表量化后, 子块中大部分值都变成 0, 可以用于隐藏信息的 DCT 系数很少。针对 Jpeg-Jsteg 算法存在的这一弱点, Chang 等人^[4]提出了一种改进算法, 即 JPEG 量化表修改方法(JPEG and Quantization Table Modification, JQTM)。他们将 JPEG 标准量化表进行了修改以便嵌入更多的信息, 修改后的量化表见图 3。为了能在量化后的每个 8×8 的 DCT 系数子块的中频区域中嵌入秘密信息, 修改后的量化表中频位置上的系数都设置为 1。如果不这样修改而继续采用标准的 JPEG 量化表去量化和反量化系数, 则嵌入的信息所导致的误差将被放大, stego 图像质量会有很大的下降。子块中位于中频区域量化后的每个 DCT 系数的最低 2 位将被嵌入信息, 即隐藏 2 bit 秘密信息, 每个子块嵌入位置及顺序见图 4。

16	11	10	16	1	1	1	1
12	12	14	1	1	1	1	55
14	13	1	1	1	1	1	69
14	1	1	1	1	87	80	62
1	1	1	1	68	109	103	77
1	1	1	64	81	104	113	92
1	1	78	87	103	121	120	101
1	92	95	98	112	100	103	99

Fig.3 Quantization table of JQTM method
图 3 JQTM 算法的量化表

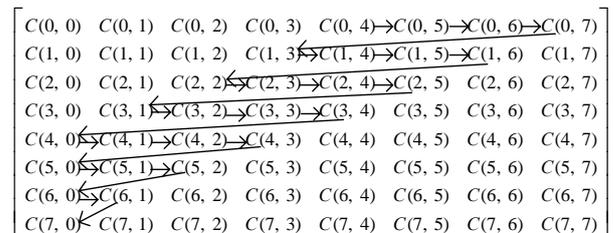


Fig.4 Embedding order of JQTM method
图 4 JQTM 算法的嵌入顺序

1.2 最优 LSB 替换方法

Wang 等人改进了传统的 LSB 替换方法, 提出了一种基于遗传算法的最优 LSB 替换算法^[5]。该算法首先运用遗传算法搜索出理想的置换矩阵, 将 LSB 替换算法中分解得到的 k 比特/像素的图像 S' 进行重新置换之后才嵌入掩体图像中。 $M = \{m_{i,j}, 0 \leq i, j \leq 2^k - 1\}$ 表示置换矩阵, $m_{i,j}$ 表示置换矩阵 M 第 i 行第 j 列的元素:

$$m_{i,j} = \begin{cases} 1 & i \text{ 置换成 } j \\ 0 & \text{不做置换} \end{cases} \quad (1)$$

当 $m_{i,j}=1$ 时, 灰度值 i 就被置换为 j 。为了确保置换过程是可逆的, 置换矩阵 M 的每行和每列有且只有一个值为 1。

下面举例说明秘密信息的置换过程, 以及置换结果的恢复过程。例如, 对于 $k=2$ 的情况, 假设有一组秘密信息为 $\{3,3,1,2,2,1,0,1\}$, 且有如图 5 所示的置换矩阵 M 及其对应的转置矩阵 M' 。根据式(1)可知, 经过置换矩阵 M 的置换操作后: 数值 0 置换成 0, 1 置换成 2, 2 置换成 3, 3 置换成 1。这样 $\{3,3,1,2,2,1,0,1\}$ 变为 $\{1,1,2,3,3,2,0,2\}$ 。置换过程是可逆的, 只要用置换矩阵的转置 M' 去置换所得到的结果数据, 就能够恢复出原始数据。

		0	1	2	3			0	1	2	3
M	0	1	0	0	0		0	1	0	0	0
	1	0	0	1	0		1	0	0	0	1
	2	0	0	0	1		2	0	1	0	0
	3	0	1	0	0		3	0	0	1	0

Fig.5 Substitution matrix M and M' transpose
图 5 置换矩阵 M 和其转置矩阵 M'

置换矩阵的不同,会导致实际嵌入掩体图像的信息不同,因而最终得到的 stego 图像也会不完全一样,图像质量自然会有所差别。如果置换矩阵是单位矩阵,即对应将原始信息不做置换直接嵌入。由此得到启发,在置换矩阵的所有可能中一定存在一个最佳矩阵,使图像质量最优。试图找出这个矩阵,用来置换信息,以此来尽可能提高 stego 图像质量。

2 本文提出的方法

受到最优 LSB 替换方法的启发,将这个策略引入变换域的信息隐藏过程中。

2.1 嵌入流程

本文算法的信息嵌入框图见图 6,整个信息嵌入过程可分 4 个步骤:

a) 秘密信息置换操作:用 BEGA 算法选择一个理想的置换矩阵 M ,对秘密信息进行置换。

b) 掩体图像预处理:应用 JPEG 压缩算法的预处理过程处理掩体图像。首先,将掩体图像进行 DCT 变换和量化,所采用的量化表见图 7。该量化表与 JPEG 标准量化表和 JQTM 方法所采用的量化表都有所不同,目的是使在量化后的 DCT 系数的 DC 到 AC 中频区域都能嵌入信息。修改后的量化表中,位于中低频区域的 35 个 AC 系数被设置为 1,保证了嵌入秘密信息后,量化和反量化过程不会导致误差被放大, stego 图像能有较好的质量,人眼察觉不到秘密信息的嵌入;而 DC 位置上的值被设置为 8,使得最终生成的 JPEG 文件与原始文件相比,在 DC 系数上的误差不会过大。

c) 秘密信息嵌入:将第 a)步得到的信息置换结果嵌入到每个子块量化后 DCT 系数的 DC 和 AC 系数的中、低频部分,嵌入顺序见图 8。位于 DC 到中频 AC 区域的系数二进制形式的最低 k 位替换为 k 比特的秘密信息, k 的具体取值视所隐藏的信息量而定,实验中设定 $k=2$ 。

d) JPEG 熵编码和 JPEG 文件生成与传送:用 JPEG 的熵编码(包括哈夫曼编码、游程编码及 DPCM 编码)对嵌入秘密信息后的每个 8×8 子块进行编码,得到一个含有修改后的量化表和所有压缩数据的 JPEG 文件。将 JPEG 文件和置换矩阵 M 传送给接收方。秘密信息的提取过程则是嵌入过程的逆过程。

需要注意的是,作者在 JQTM 算法的量化表基础上,将低频部分的量化系数都改为了 1。这样虽然可以嵌入更多的信息量,但经本文算法得到的 JPEG 文件也会略大于 JQTM 方法得到的文件。

2.2 BEGA 算法原理

遗传算法是模拟生命进化机制搜索和优化,并将自然遗传学和计算机科学结合的优化方法,遗传算法由于具有很强的解决问题能力和广泛的适应性,已渗透到研究与工程的各个领域。为了进一步提高遗传算法的性能,文献[7]提出了 BEGA 思想。BEGA 的基本原理是:以随机方式产生一个对应于优化问题可行解的初始群体,并标记种群适应度最大的个体为这一代的蜂王。然后从种群中选择出一部分生命力强的个体,再加上随机生成的新个体共 $N/2$ 个个体与选择出的蜂王以一定的概率进行交配和变异,向更高的适应度进化并产生新的蜂王。通过在进化过程中对历代群体中的个体进行优胜劣汰来寻找问题的最优解。BEGA 算法较标准遗传算法(Standard Genetic

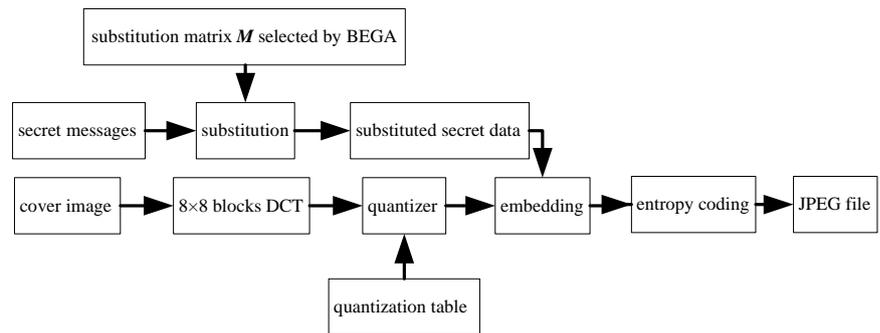


Fig.6 Diagram of embedding procedure
图 6 嵌入过程流程

8	1	1	1	1	1	1	1
1	1	1	1	1	1	1	55
1	1	1	1	1	1	69	56
1	1	1	1	1	87	80	62
1	1	1	1	68	109	103	77
1	1	1	64	81	104	113	92
1	1	78	87	103	121	120	101
1	92	95	98	112	100	103	99

Fig.7 Modified quantization table
图 7 修改后的量化表

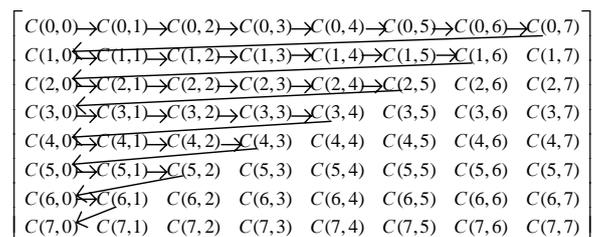


Fig.8 Proposed embedding order
图 8 信息嵌入顺序

Algorithms, SGA)^[9]主要有 2 点改进:首先是引进了蜂王,在子代进化过程中,始终记录每一代适应度最大的种群个体为蜂王,新生成的子代都会继承父代最优良的基因,因此种群的进化方向变得更加明确;其次,为维持个体的多样性,引入了蜜蜂繁殖过程中“外来种群”的思想,引入部分随机个体,以期加强遗传算法的勘探能力。

2.3 用 BEGA 优化算法搜索最优置换

本文中所要解决的问题是找出一个最优的置换矩阵,从而提高 stego 图像的质量。利用 BEGA 算法解决寻优问题,必须要将待寻优的问题中可行解的形式转化为种群个体形式,即每个置换矩阵都唯一对应一个个体 P :

$$P = [p_0 p_1 \dots p_{2^k-1}] \quad (2)$$

式中 P_i 表示置换矩阵中第 i 行中值为 1 元素所在的列。

假设有图 9 的置换矩阵 M ,那么其所对应的个体 P 为 [1 2 3 0]。

2.3.1 适应度函数

评价一个个体好坏的标准就是 stego 图像质量,本文的适应度函数 f 用 stego 图像与掩体图像之间的峰值信噪比(Peak Signal to Noise Ratio, PSNR)衡量:

$$PSNR = 10 \times \lg \left[\frac{255^2}{\frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H (z_{i,j} - c_{i,j})^2} \right] \quad (3)$$

式中: W, H 表示掩体图像宽度与高度; $z_{i,j}$ 和 $c_{i,j}$ 分别表示 stego 图像和掩体图像位于 (i, j) 处的像素灰度值。

2.3.2 遗传操作

a) 选择算子:采用轮盘赌法对个体进行选择。该法基本思路是根据个体的相对适应度,从群体中选择 $(N/2)\gamma$ 个个体以组成下一代群体中的一部分,其中 γ 代表选择算子 ($0 \leq \gamma \leq 1$)。显然,个体适应度越大,被选中的概率就越大。但是,适应度小的个体也有可能被选中,以增加下一代群体的多样性。

b) 交叉操作:交叉是把待进化的父代种群个体与蜂王的部分结构加以替换重组而产生新个体的操作,目的是寻找双亲中已有的但未能被合理利用的基因信息。本文是将选择出的 $N/2$ 个父代个体与蜂王进行交叉,产生 $N/2$ 对种群个体,交叉位置随机确定。

c) 变异操作:通过变异可引进新的基因以保证种群的多样性,在一定程度上可防止前收敛的发生。本文将每个新个体按照变异概率 P_m 判断是否需要变异,如变异,则随机交换该个体的某两位,产生新个体。

d) 有效性调整:如前所述,每一个体代表一种置换方式,显然个体中不能包含重复的元素,因此上述步骤可能会产生无效个体,对于无效个体需要进行有效性调整,使之符合要求。

2.3.3 利用 BEGA 优化算法来寻找全局最优置换矩阵的具体步骤

- 随机产生初始种群 $A(0)$,并对 γ 、变异概率 P_m 和最大迭代次数 $iter_{max}$ 赋值。
- 计算种群中每个个体对应的 $PSNR$,将最优个体保存,标记为蜂王 $Queen$ 。
- 按轮盘赌选择方法,从父代种群中选择 $(N/2)\gamma$ 个个体,再随机产生 $(N/2)(1-\gamma)$ 个新个体组成 $N/2$ 个个体。
- 将这 $N/2$ 个个体分别与蜂王 $Queen$ 进行交叉运算,并做有效性调整,得到子代种群 $B(t)$ 。
- 按变异概率对 $B(t)$ 进行变异操作得到种群 $C(t)$ 。
- 计算种群 $C(t)$ 中个体所对应的 $PSNR$,将 $PSNR$ 最大的记为 $Queen_New$ 。
- 如果 $f(Queen_New) > f(Queen)$,将 $Queen_New$ 记为下一代新蜂王, $C(t)$ 即为下一个新种群;否则,用 $Queen$ 替代 $C(t)$ 中的最差个体,再将 $C(t)$ 作为下一个新种群。
- 检验是否达到停止条件,如果达到,输出最优结果,否则转向 b)。

2.3.4 分块最优替换算法

在用 BEGA 寻找全局最优置换矩阵的基础上,又提出了利用 BEGA 算法在掩体图像的每一个 8×8 子块寻找最优置换矩阵的方法。由于针对每个子块搜索不同的置换矩阵,搜索过程细化,因此能够得到更优的效果,但是相应也需要更多的存储空间。具体方法只要将寻找全局最优置换的步骤(2.3.3)应用到每个子块即可实现,详细过程不再赘述。

$$M = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 2 & 0 & 0 & 1 \\ 3 & 1 & 0 & 0 \end{bmatrix} \quad P = [1 \ 2 \ 3 \ 0]$$

Fig.9 An example of substitution matrix M and the corresponding individual P when $k=2$
图 9 $k=2$ 时,举例说明置换矩阵 M 和个体 P 之间的对应关系

3 实验结果

实验选取 512×512 的 8 bit 灰度图像作为掩体图像(见图 10), 秘密信息采用 128×72 的 8 bit 灰度图像。采用全局的 BEGA 优化算法, 得到 stego 图像(见图 11)。同时, 还与直接嵌入方法、Wang 的基于遗传算法的方法以及分块 BEGA 算法做了比较。

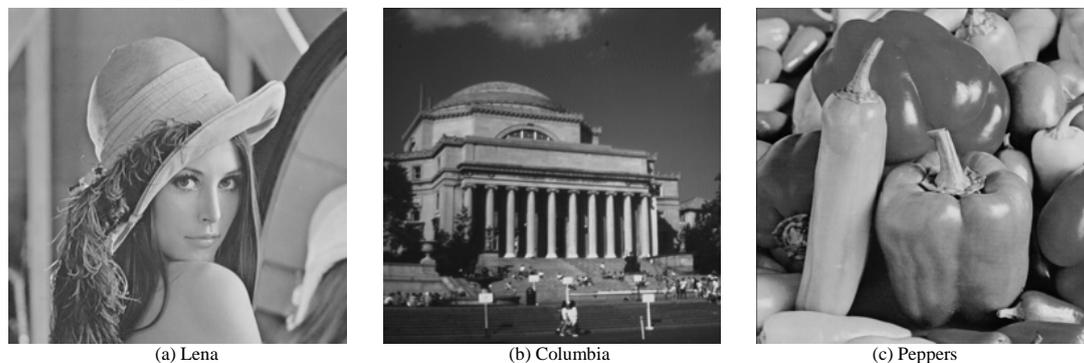


Fig.10 Cover images
图 10 掩体图像

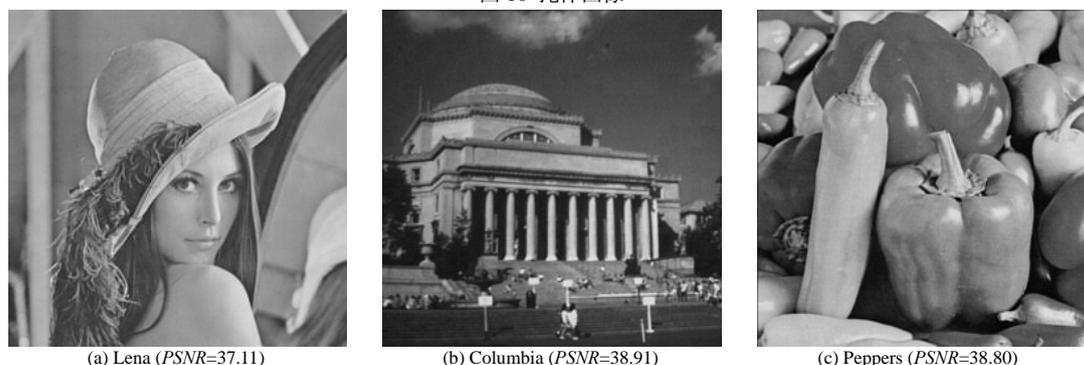


Fig.11 Stego images of the proposed method
图 11 经本文提出算法得到的 stego 图像

表 1 给出了采用秘密信息不经置换直接嵌入的方法、Wang 的 GA 算法以及本文提出的两种算法嵌入信息后所得的 stego 图像的 PSNR 值。可以看出, 经过置换后得到的 stego 图像质量优于未经置换的方法。BEGA 方法基本与 GA 方法相同, 个别会略优于 GA。另外, 由于分块优化针对每个子块搜索不同的置换矩阵, 把嵌入过程细化, 因此图像质量优于两种全局优化方法。

表 1 不同嵌入算法所得的 stego 图像 PSNR 值(单位: dB)

Table1 Comparisons of PSNR by different methods(unit:dB)

method	images					
	Lena	Columbia	Peppers	Woman	Milkdrop	Couple
non-optimal-substitution method	36.16	37.47	37.44	39.38	38.94	35.87
the GA method proposed by Wang	37.11	38.91	38.76	41.91	41.09	36.73
global optimization using BEGA	37.11	38.91	38.80	41.91	41.11	36.73
block optimization using BEGA	37.59	39.66	39.50	43.51	42.38	37.14

为了检验 BEGA 算法的性能, 在相同的计算机仿真环境下(Pentium 4, 1 G 内存, Matlab 7.6.0)进行仿真。在已知最优置换矩阵的前提下, 比较采用这两种方法找到最优解所分别消耗的时间, 见表 2。可以看出, 由于 BEGA 加强了遗传算法的开采能力, 使得全局 BEGA 优化方法更快接近最优解, 消耗时间明显少于用标准遗传算法优化的方法。

表 2 利用不同优化方法找到最优置换矩阵所花费的时间(单位: s)

Table2 Comparison of the computational time by different methods(unit:s)

method	Lena	Columbia	Peppers	Woman	Milkdrop	Couple
global optimization using SGA	128.72	133.30	133.05	128.43	136.34	142.29
global optimization using BEGA	43.41	44.60	43.89	63.17	42.66	42.86

4 结论

本文提出了一种基于 BEGA 算法的 JPEG 隐写方法。该方法修改了标准的 JPEG 量化表,将秘密信息经过置换操作后再嵌入掩体图像量化后的 DCT 系数中,其中置换方式是利用性能优异的 BEGA 算法搜索得到的。在寻找全局最优置换矩阵方法的基础上,还提出了寻找局部最优置换矩阵的分块优化方法。实验结果表明:这两种方法得到的 stego 图像质量均好于直接嵌入方法;且由于嵌入过程中采用了置换矩阵,提高了信息隐藏的安全性。与标准遗传算法相比, BEGA 算法性能更加优越,缩短了搜索时间。另外,本文算法仍存在待改进之处,今后工作将集中在对该算法信息安全性的研究,进一步提高其抗隐写分析性能。

参考文献:

- [1] Fabien Petitcolas A P, Ross J Anderson, Markus G Kuhn. Information hiding- a survey[J]. Proceedings of the IEEE, 1999, 87(7):1062-1078.
- [2] 王晨毅, 王建军. 第二代 Curvelet 变换域的信息隐藏方法[J]. 信息与电子工程, 2008, 6(2):105-110.
- [3] Pennebaker W B, Mitchell J L. JPEG: Still Image Data Compression Standard[M]. New York: Van Nostrand Reinhold, 1993.
- [4] Chang C C, Chen T S, Chung L Z. A steganographic method based upon JPEG and quantization table modification[J]. Information Science, 2002, 141(1):123-138.
- [5] Wang R Z, Lin C F, Lin J C. Image hiding by optimal LSB substitution and genetic algorithm[J]. Pattern Recognition, 2001, 34(3):671-683.
- [6] Li X, Wang J. A steganographic method based upon JPEG and particle swarm optimization algorithm[J]. Information Science, 2007, 177(5):3099-3109.
- [7] 孟伟, 韩学东, 洪炳谿. 蜜蜂进化型遗传算法[J]. 电子学报, 2006, 34(7):1294-1300.
- [8] Dervis Karaboga, Bahriye Akay. A comparative study of Artificial Bee Colony algorithm[C]// Applied Mathematics and Computation. 2009:108-132.
- [9] Goldberg D E. Genetic Algorithms in Search, Optimization and Machine Learning[M]. MA: Addison-Wesley, 1989.

作者简介:



戴少华(1986-), 男, 河南省卢氏县人, 在读硕士研究生, 主要研究方向为信息隐藏. email: shaohua.d@126.com.

王建军(1960-), 男, 陕西省乾县人, 博士, 副教授, 主要研究方向包括信息安全、图像处理及编码等。

(上接第 179 页)

作者简介:



赵 磊(1984-), 男, 河南省新郑市人, 在读硕士研究生, 主要研究领域为遥感图像处理. email: wangbin@fudan.edu.cn.

王 斌(1964-), 男, 西安市人, 博士, 教授, 主要研究方向为信号和图像处理及其在遥感数据的分析与处理、生物电磁信号的分析 and 处理中的应用。

张立明(1943-), 女, 浙江省临海市人, 教授, 主要研究方向为人工神经网络模型及其在图像识别中的应用、图像编码和处理、非线性方法。