Fidelity-Controlled Robustness Enhancement of Blind Watermarking Schemes Using Evolutionary Computational Techniques

Chun-Hsiang Huang, Chih-Hao Shen and Ja-Ling Wu

Communication and Multimedia Laboratory
Department of Computer Science and Information Engineering
National Taiwan University
Taipei, Taiwan, R. O. C.
E-mail: {bh, shen, wjl}@cmlab.csie.ntu.edu.tw

Abstract. Designing optimal watermarking schemes is inherently an interesting and difficult problem since the three most important performance requirements of digital watermarking – fidelity, robustness and watermark capacity – are conflicting with each other. Nowadays, most watermarking schemes hide the watermark information in a heuristic manner, that is, watermarks are often embedded according to predefined rules and empirical parameter settings. Therefore, the performance of digital watermarking can only be passively decided and evaluated, rather than being actively adopted as additional clues helpful to achieve better performance in embedding modules. In this paper, watermark embedding is simulated as an optimization procedure in which optimal embedded results are obtained by using important evolutionary computation techniques – the genetic algorithms. Under the condition that fixed amount of watermark bits are hidden, in this work, the minimal fidelity requirement of embedded content can be specified by users in advance and guaranteed throughout the embedding procedure. Furthermore, concrete measures of the robustness against certain attacks are treated as the objective functions that guide the optimizing procedure. In other words, a blind watermarking scheme with application-specific data capacity, guaranteed fidelity, and theoretically optimal robustness against certain types of attacks is proposed. Experimental results clearly show that the proposed scheme possesses great performance improvements over the original one. More importantly, the proposed enhancing approach has many desired architectural characteristics, such as blind detection, asymmetric embedding/detection overheads, as well as embedding and detection in different domains.

I. Problem Formulation: Optimal Watermarking

In the past decade, various watermarking schemes have been proposed, and many important advances in the field of digital watermarking have been made. A comprehensive introduction to current watermarking technologies and theoretic foundations can be found in [1]. However, designing optimal watermarking schemes is still an open problem to which satisfying solutions are not yet found [2]. The difficulties one might face while designing optimal watermarking schemes is a natural result due to the three conflicting requirements of most watermarking systems: fidelity of embedded content, robustness of the hidden information against common processes or malicious attacks, and the data capacity of hidden information. Among these requirements, capacity is often decided in advance according to the purpose of watermarking application. Therefore, as long as the predefined amount of embedded data is large enough to carry necessary information, such as identifications of content authors for owner proving or usage rules that shall be parsed by DRM-enabled consumer-electronic devices, data capacity can be simply regarded as a fixed parameter. However, even under this assumption, obtaining a reasonable trade-off between fidelity and robustness is still not as simple as one might think.

In existing watermarking schemes, perceptually acceptable embedded outcomes are often produced according to predefined embedding rules based on complicated perceptual models or assumptions, and then the robustness of that scheme is empirically experimented and evaluated by performing various attacks on the embedded media. In order to clearly illustrate the relationship between traditional watermarking schemes and their performance indexes, the performance-space view of digital watermarking schemes shall be introduced and examined in the beginning. As shown in Fig. 1, any embedded result created using a certain watermarking scheme can be expressed by a point located somewhere within the space spanned by two axes representing fidelity and robustness, respectively. Roughly speaking, it is generally agreed that the higher the required fidelity is, the lower the robustness of hidden signals against attacks will be. In fact, the region that possible embedded outcomes may locate within is consequently determined after the watermarking scheme and the original/watermark pairs are chosen. Although potentially better performance of the chosen watermarking scheme may exist, traditional watermarking schemes lack the ability to exploit better embedded outcomes.

II. Genetic-Algorithms and Watermarking

To solve the watermarking performance optimization problem within reasonable computation time, the genetic-algorithms (GAs) based optimization techniques are used in this paper. GAs are important optimization techniques belonging to the area of evolutionary computation [3]. During GA-based optimization processes, solutions to the problem can be evaluated according to objective function values representing the degree of fitness. A population of candidate solutions will be initialized as finite-length strings - the so-called chromosomes - over finite alphabet. Different from conventional single-point search methods, GAs work with a population of points in a parameter space simultaneously. In practical GA-based optimizations, three GA operators - reproduction, crossover, and mutation - are often applied to the chromosomes repeatedly. After iteratively adopting these GA operations, near-optimal solutions of desired parameters for the original problem can be obtained. A detail explanation of GAs can be found in [4].

The authors first introduced the idea of GA-based watermarking enhancement in [5]. In that paper, the performance of a DCT-domain watermarking scheme, being proposed in [6], is enhanced by applying GA operators, and the best watermark embedding positions for each DCT block can be found. [5] is the earliest publication that connects GAs together with digital watermarking. However, this enhancing scheme was criticized due to several factors that seriously limit its usage. For blind watermarking schemes, the optimal embedding positions must be delivered to the watermark detector as secret keys since the decoder cannot figure out the optimized embedding positions without the original content. However, this is usually not feasible in many important real-world applications, thus seriously limiting its applicability. As for non-blind watermarking schemes, the optimization procedure will be a time-consuming process in both embedding and detection sides. Although the authors proposed a lightweight genetic search algorithm in [7] to shorten the computation time, large amounts of computation overheads in both watermark embedding and detection are still unacceptable. In [8], a similar DCT-domain watermarking scheme is proposed, where optimization for robustness is considered. Furthermore, a GA-based spatial-domain watermarking algorithm is also proposed in [8]. However, the usage of the latter scheme is limited due to the weak robustness of its spatial domain embedding nature, and both schemes suffer from the aforementioned secret key delivery problem.

In this paper, a watermarking-performance enhancement architecture possessing theoretically optimal robustness against certain types of attacks, guaranteed fidelity and application-specific data capacity will be proposed. The proposed architecture is inherently suitable for blind watermarking, and the asymmetric embedding and detection design can effectively reduce the penalty of long computation time. Furthermore, the proposed watermarking scheme has the desirable characteristic that embedding and detection can be performed in different domain, thus both direct control of fidelity in spatial domain while embedding and strong robustness against attacks in frequency domain while detecting can be realized in a single framework.

The paper is organized as follows. Section III describes the proposed watermarking enhancement scheme and related implementation details. Experimental results, including the performance comparisons against original watermarking scheme under the assumption of equal data capacity, will be listed and explained in Section IV. Section V gives a brief discussion about the pros and cons of the proposed performance enhancing architecture, and section VI concludes this write-up.

III. The Proposed Architecture and Implementation Details

The proposed performance enhancement architecture is general and can be used to enhance the performance of various existing blind watermarking schemes. However, in order to save the implementation time and consider that frequency-domain watermarking schemes are well known for their better robustness and fidelity, a blind version of the block-DCT based image watermarking scheme originally introduced by

[6] is used to evaluate the power of this enhancing approach. [9] illustrates the details about turning the originally non-blind watermarking approach into a blind one. In this scheme, embedded watermarks are meaningful binary patterns and are randomly permuted before embedding. Watermark bits are embedded into predefined positions within the middle-band coefficients of each DCT block. The polarity between AC coefficients at selected positions and scaled DC coefficient of each block, after considering the effects of JPEG compression, is adjusted to insert the watermarking bits. Due to the blind nature of the adopted scheme in [9], the original perceptual model introduced by [6] cannot be applied directly, thus the degree of coefficient adjustment is empirically and uniformly determined. The definition of polarity for the *m*-th watermarking bit that will be embedded into one coefficient block is given by:

$$P_m = \begin{cases} 1, & if \dfrac{|AC_m|}{Q_m} \times Q_m > \dfrac{|DC|}{Scale\_Factor \times Q_{DC}} \times Q_{DC} \\ -1, & otherwised, \end{cases} \tag{1}$$

where AC_m is the AC coefficient at the position that the m-th watermark bit within each 8x8 DCT block will be embedded. DC denotes the DC value of that block. Q_m is the value in the JPEG quantization table corresponding to the position of AC_m, and Q_{DC} is the value for DC in the JPEG quantization table.

Fig. 2 depicts the flowchart of our scheme. The required fidelity of embedded content is specified by users according to different application scenarios and guaranteed throughout the embedding process. To concretely express the fidelity, the commonly used PSNR (Peak Signal to Noise Ratio) of the embedded image is adopted as the index of fidelity. In addition, the robustness becomes the performance index that we would like to enhance.

The optimization process is done in a block-by-block manner. For each 8x8 image block, a set of N initial parent chromosomes will be generated. Each initial parent chromosome is a randomly generated 8x8 block where the value of each pixel uniformly distributes over a range taking equal positive and negative extent. In fact, a parent chromosome represents a possible distortion block that stands for the difference between the original image block and the embedded one in spatial domain. Next, the energy-shaping module is applied to all initial parent chromosomes so that they can satisfy the minimum fidelity requirement specified in advance. For example, according to the definition of PSNR, if the user asks for a required PSNR value higher than 40, the maximal allowable block energy (that is, the sum of the squared value of each distortion pixel) of an 8x8 chromosome block shall be less than 416. If the energy of a randomly generated chromosome block is higher than the maximal allowable energy, the chromosome block will be uniformly scaled down to satisfy the fidelity limit. The difference between the energy of the obtained chromosome and the maximal energy limit can be further reduced by slightly adjusting randomly selected pixels. In this way, all processed parent chromosomes will result in embedded results that user-specified fidelity requirement will be guaranteed.

Each energy-shaped parent chromosome will be respectively added to the original image block to form an embedded candidate. Then the fitness value corresponding to each candidate shall be calculated. As mentioned before, the fitness value must describe the robustness of the adopted watermarking scheme against certain attacks. Therefore, the percentage of correctly extracted watermarking signals against certain attacks, named as correctly extracted rate (CER) in this paper, is undoubtedly the most intuitive index. To calculate the fitness value corresponding to each candidate chromosome, a block DCT operation is performed for each candidate block. Furthermore, a JPEG-compliant quantization/dequantization procedure is performed to each produced coefficient block to simulate the effect of JPEG compression. Every quantization step of the adopted quantization table is about half of that of the default JPEG luma quantization table. Finally, according to the adopted watermarking scheme, the fitness function value of the *n*-th initial candidate block is figured out according to the definition of Eq. (2):

$$F_n^1 = \sum_m P_m \cdot w_m \cdot (W_1 + \cdot W_2 \cdot \left| \frac{|AC_m|}{Q_m} \times Q_m - \frac{|DC|}{Scale\_Factor \times Q_{DC}} \times Q_{DC} \right|) \qquad \textbf{(2)}$$

where w_m is the m-th watermark bit going to be embedded into a predefined position of current coefficient block, and this binary signal is represented as 1 or -1. P_m represents the polarity extracted from the predefined position that the m-th watermark bit shall be embedded to. The definition of polarity is listed in Eq. (1). W_1 and W_2 are both weighting factors. W_1 controls the degree that the case “a watermark bit is correctly or wrongly extracted” contributes to the fitness value, i.e., W_1 stands for each embedded coefficient’s contribution to a visually recognizing detector. On the contrary, W_2 controls the degree that the embedded coefficient contributes to a correlation-based detector. For all experiments in this paper, W_1 is set to 100 and W_2 is 1. Other symbol definitions are the same as those given in Eq. (1).

Next, a set of N child chromosomes will be reproduced and processed according to GA-based rules. The reproduction is done by the famous roulette-wheel-method [4]. The parent chromosomes have higher fitness values are more possible to generate more offspring. As for the crossover operation, reproduced child chromosomes are randomly mated into pairs and exchange arbitrary portions of chromosomes to the other. In other words, parts of two child chromosomes are combined to form a new 8x8 distortion block. After performing the crossover operation, each pixel component of the child chromosomes has a small possibility to change from positive to negative or from negative to positive. This is the mutation operation adopted to help generating new candidates.

Now although these child chromosomes are generated based on their parent chromosomes, the adopted GA operations may result in child chromosomes violating the fidelity requirement specified in the beginning. Thus the energy-shaping procedure shall also be performed on these generated children chromosomes.

Finally, a survival-of-the-fittest policy is used to select N next-generation parent chromosomes from the set consisting of N parent chromosomes and no-more-than N child chromosomes. Since the whole set of original parent chromosomes is included in this survival competition, the fitness values of next-generation parent chromosomes will never get lower than those of previous-generation ones. These aforementioned GA optimization processes will be done repeatedly until a specified number of iterations (named as the generation number in a GA-based approach) have been performed. Finally, the chromosome with the highest fitness value will be added to the original image block, and this added block is regarded as the embedded block of best robustness.

It is worth noting that the watermark is never explicitly "embedded" to the original. On the contrary, we search for the best candidate subject to the fidelity constraint directly according to the simulated robustness performance. On the other hand, the watermark extraction process is exactly the same as the watermark detector in the original watermarking algorithm. This asymmetric behavior of the embedding and detecting modules is quite different from that of the traditional watermarking schemes. To be more specific, the proposed embedding module can be regarded as a generalized performance enhancement module depending on the given watermark detection algorithm and performance indexes. Similar optimization can be applied to various blind watermarking algorithms as long as the watermark detector is given. The involved performance indexes and attack models can be reasonably replaced, e.g. using subjective perceptual index to substitute the objective PSNR or changing the JPEG compression attack to the most probable operations that your application might encounter. In other words, more flexibility and better performance to the actual application can be obtained.

Viewing the proposed enhancement scheme by the aforementioned performance-based model, the roles played by each operation within the proposed scheme can be clearly identified. As shown in Fig. 3, crossover and mutation operators discover new embedding candidates, the energy-shaping module modifies over-distorted candidates so that the required fidelity constraint can be observed, and the survival-of-the-fittest operation guarantees that the newly generated parent chromosomes never locates at lower positions than their parents in the performance space.

IV. Experimental Results

Experiments are performed to evaluate the effectiveness of the proposed scheme. The 512x512 gray-level Lena image and a 128x128 binary watermark pattern are adopted, as shown in Figure 4. According to the dimension ratios of the original and the watermark images, 4 watermark bits will be embedded into each 8x8 block. In other words, a fixed data capacity of 16,384 bits is determined in advance.

Fig. 5 reveals some important characteristics of our enhancement schemes. According to the performance curve of the GA-based enhancement scheme, it obviously proves the assumption that the lower the specified fidelity constraint is, the higher the optimized robustness will be. In this experiment, the generation number of optimization process is set to 1000, and the mutation rate is set to 0.1. It is worth noting that even for embedded images of excellent visual quality, e.g. PSNR larger than 40, the percentage of correctly extracted watermark bits is still high enough to identify the existence of a watermark.

Fig. 6 shows the quality improvements after performing different generations of iterations. As we expected, the more the iterative operations we performed, the stronger robustness the obtained embedded results may possess. Fig. 7 lists the corresponding extracted patterns of the embedded watermark for visual evaluation.

In Fig. 8, to show that the results obtained by the proposed enhancing scheme have better performance than those created by using the original watermarking scheme, the performance curve of the non-optimized algorithm introduced in [9] is also listed for comparison. The coefficients to be embedded by the non-optimized approach are uniformly adjusted in order to create embedded results of different PSNR values. And then, JPEG compression attacks are performed on these embedded images for further evaluation of the correctly extracted rate. According to the comparison results, the proposed enhancement scheme outperforms when high fidelity is required. More importantly, for cases where embedded results the original algorithm cannot produce, such as embedded results of PSNR values higher than 42 dB, the proposed scheme can still successfully generate the needed output.

However, the seemingly-counterintuitive phenomenon that the original watermarking scheme outperforms the proposed enhancement algorithm when low fidelity is required indicates a potential weakness of evolutionary computational techniques: the obtained results may be trapped in local optimum when the search space is large. To solve this problem, simple solutions such as adopting better initial search candidates or increase the iteration number can be of help. Experimental results in Fig. 8 show great improvements obtained by adopting better initial parent chromosomes that distribute more evenly over the whole search space, e.g. using embedded results watermarked with different uniform adjustment magnitudes to produce initial parent chromosomes. In fact, this simulation also implies a more general watermarking performance enhancement philosophy – producing the embedded results based on predefined rules of existing blind watermarking schemes first and fed them in to the proposed enhancement architecture as initial search candidates. Then, the proposed GA-based enhancement architecture can effectively improve the performance of any adopted blind watermarking scheme.

It is also intuitive to assume that, the larger the iterative generation number is, the more the number of searched candidates are, and thus better embedding results can be found. According to our simulation, the increase of iteration number also results in better embedding results. However, the improvements are not as obvious as those contributed by adopting better initialization. In addition, for real-world applications, the generation number used for performance optimization is often limited by the actual computation power of embedding devices and the time constraint of the application scenarios, thus only adequate iteration numbers can be adopted.

In spite of the JPEG compression attack that has been incorporated into the design of embedding module, robustness against other attacks shall be examined too. According to our experiments, the results obtained by the proposed enhancement scheme successfully survive various other processing/attacks, such as cropping, blurring, adding noise, and scribbling.

V. Discussions

The proposed GA-based watermarking-performance enhancing architecture has a lot of advantages. First, system users can specify the required watermarking fidelity that must be guaranteed according to different application needs. Next, the asymmetric embedding/detection structure not only suits most kinds of blind watermarking schemes but also greatly alleviates the problem that evolutionary computation techniques are most often criticized – long computation time. Since the watermark detector is exactly the one used in the original watermarking scheme, many common applications of watermarking will not be affected by the required computation in the embedding process. Furthermore, the proposed watermarking scheme has the desirable characteristic that embedding and detection can be performed in different domain, thus both direct control of fidelity in spatial domain for embedding and strong robustness against attacks in frequency domain while detecting can be realized in a single framework.

An obvious problem that shall be taken into consideration is the modeling of more than two attacks while calculating the fitness value. Though the experimental results have shown robustness against other attacks, modeling multiple types of attacks and trying to optimize the performance against them are still important issues that worth further exploitation. This will be an important topic of our future research.

VI. Conclusion

In this paper, a novel watermarking performance enhancement architecture based on existing watermarking schemes and evolutionary computation techniques is proposed. The proposed scheme optimizes the robustness against certain attacks and guarantees minimum fidelity, under the condition of fixed data capacity. The proposed embedding procedures in our architecture is quite different from current watermarking schemes in concepts, and the architecture can be easily adopted to improve the performance of existing blind watermarking schemes. Experimental results show its superiority in real image watermarking applications against certain attacks, such as the JPEG compression attack.

Reference:

1. I. J. Cox, J. Bloom and M. L. Miller, *Digital Watermarking*, Morgan Kaufmann Publishers, 1st Edition, 2001
2. I. J. Cox and M. L. Miller, “The First 50 Years of Electronic Watermarking,” Journal of Applied Signal Processing, 2002, 2, pp126-132, April 2002
3. D. B. Fogel, *Evolutionary Computation toward a New Philosophy of Machine Intelligence*, IEEE Press, 1995
4. D. E. Goldberg, *Genetic Algorithm in Search, Optimization & Machine Learning*, Addison-Wesley, 1989
5. C. H, Huang and J. L. Wu, “A Watermark Optimization Technique Based on Genetic Algorithms,” SPIE Electronic Imaging 2000, San Jose, January, 2000
6. C. T. Hsu and J. L. Wu, “Hidden Digital Watermarks in Images,” IEEE Transactions on Image Processing, vol. 8, No. 1, January 1999
7. J. L. Wu, C. H. Lin and C. H. Huang, “An Efficient Genetic Algorithm for Small Range Search Problem”, *Intelligent Multimedia Processing with Soft Computing*, Springer-Verlag, pp253-280, 2005
8. J. S. Pan, H. C. and F. H. Wang, “Genetic Watermarking Techniques,” The 5th Int’l Conference on Knowledge-based Intelligent Information Engineering System & Allied Technologies.
9. C. H. Huang and J. L. Wu, "A Blind Watermarking Algorithm with Semantic Meaningful Watermarks," 34th Asilomar Conference on Signals. Systems, and Computers, Pacific Grove, October, 2000.

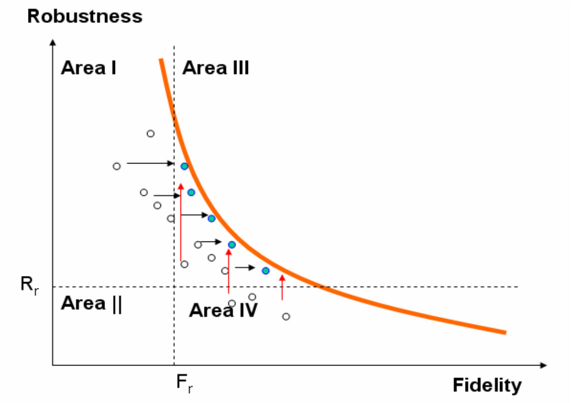

Fig. 1. The performance-space view of watermark embedding is illustrated. The curve represents the inherent performance limit of some watermarking algorithm. The watermark-embedded outcomes of a specific scheme, represented by empty circles, are determined according to predefined rules or models. Assume that if there is an application that its desired embedded results must possess robustness stronger than R_r and fidelity better than F_r, only circles locate within Area III under the curve can be of use. In fact, better embedded results, such as those shown in solid circles, may be available, but conventional watermarking schemes lack the ability to obtain them. The vertically and horizontally arrows indicate possible robustness and fidelity performance enhancements, respectively.

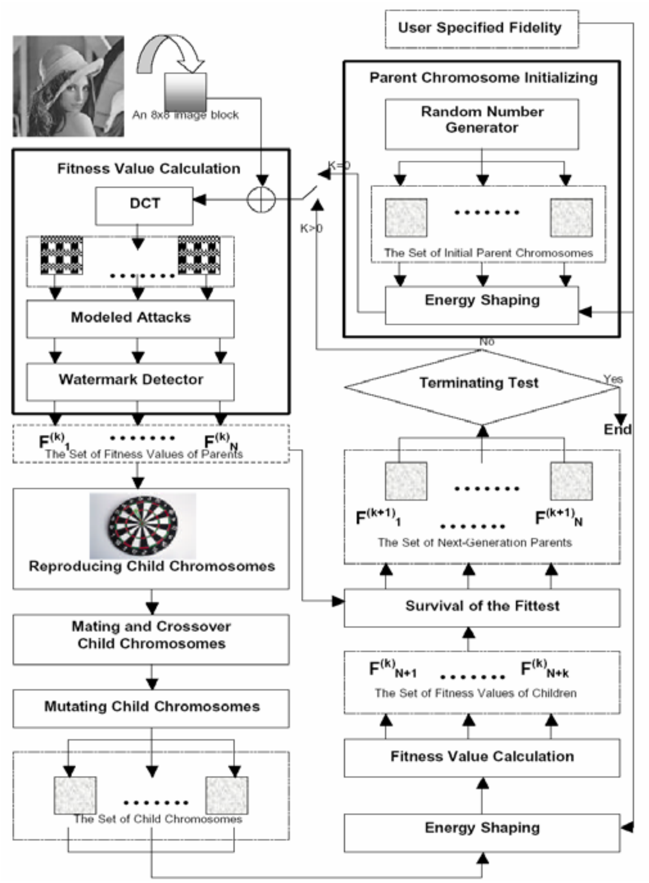
User Specified Fidelity
Parent Chromosome Initializing
Random Number Generator
The Set of Initial Parent Chromosomes
Energy Shaping
An 8x8 image block
Fitness Value Calculation
DCT
K=0
K>0
Modeled Attacks
Watermark Detector
No
Terminating Test
Yes
End
$F^{(k)}_1$ $F^{(k)}_N$
The Set of Fitness Values of Parents
$F^{(k+1)}_1$ $F^{(k+1)}_N$
The Set of Next-Generation Parents
Reproducing Child Chromosomes
Survival of the Fittest
Mating and Crossover Child Chromosomes
$F^{(k)}_{N+1}$ $F^{(k)}_{N+k}$
The Set of Fitness Values of Children
Mutating Child Chromosomes
Fitness Value Calculation
The Set of Child Chromosomes
Energy Shaping

Fig. 2. The flowchart of the proposed enhancing scheme is depicted. The blocks with dotted outlines are intermediated data sets; other blocks are required functions. This optimization process terminates after a predefined number of generations is performed.

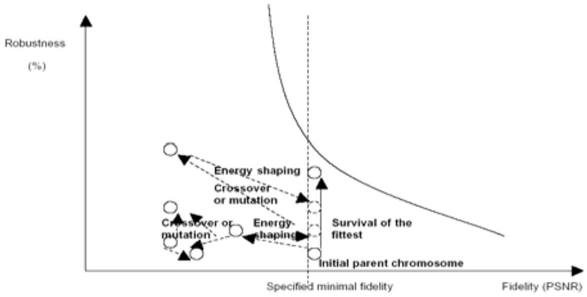

Fig. 3. From the viewpoints of performance-based watermarking model, different effects of components used in the proposed enhancing scheme are depicted.

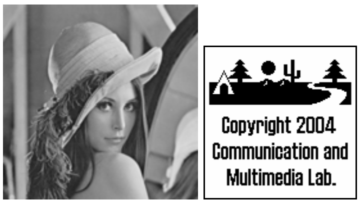

Fig. 4. The 512x512 original Lena image and the 128x128 watermark image are shown. Actual size ratio between the two images is not preserved due to layout considerations.

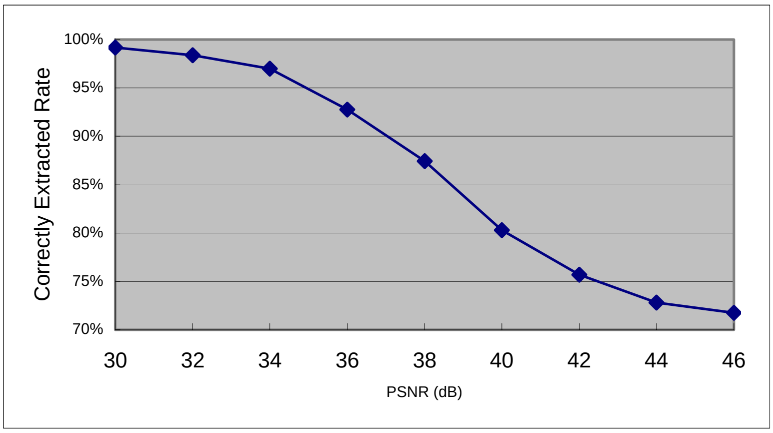

Fig. 5. The relationship between different specified minimal fidelity requirement and the corresponding robustness of embedded results, given the same number of optimizing iterations, is depicted. It clearly shows that: the higher fidelity the user demands, the worse robustness the embedded result will possess.

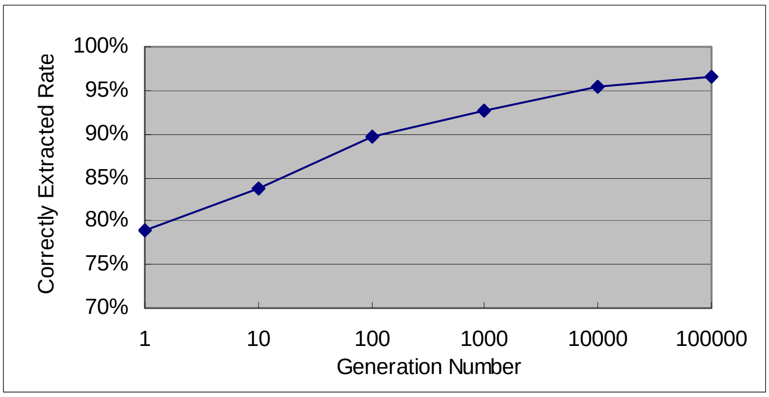

Fig. 6. The relationship between the number of optimizing iterations and the corresponding robustness performance, given that the pre-specified fidelity requirement is 36dB, is depicted. The more optimization computation we performed, the more-robust embedded results we can obtain.

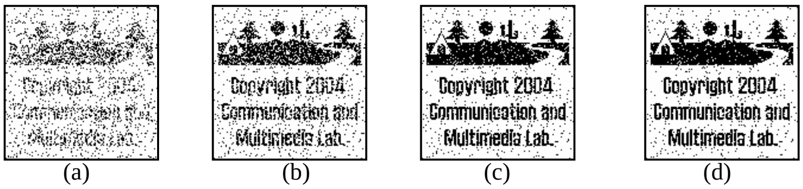

Fig. 7. Extracted watermark patterns after performing (a) 1 generations, (b) 100 generations, (c) 10,000 generations and (d) 1,000,000 generations of optimization computations, given that the pre-specified PSNR requirement is 36dB, are listed for visual evaluation.

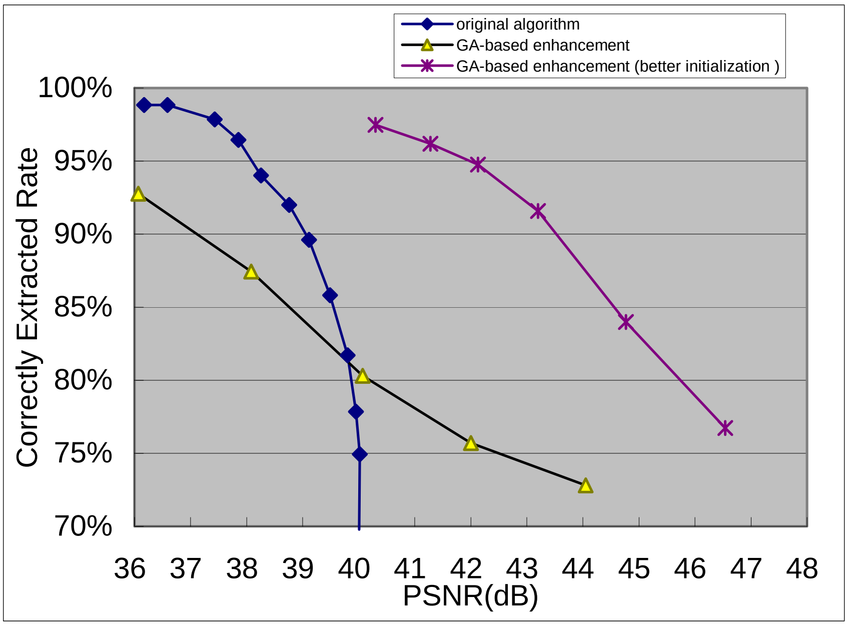

Fig. 8. The results obtained by using the proposed scheme with better initial parent chromosomes show great improvements over the original watermarking algorithm and the proposed scheme with randomly generated parent chromosome.