Watermarking Robustness Evaluation Based on Perceptual Quality via Genetic Algorithms

Giulia Boato, *Associate Member, IEEE*, Valentina Conotter, Francesco G. B. De Natale, *Senior Member, IEEE*, and Claudio Fontanari, *Member, IEEE*

***Abstract*—This paper presents a novel and flexible benchmarking tool based on genetic algorithms (GA) and designed to assess the robustness of any digital image watermarking system. The main idea is to evaluate robustness in terms of perceptual quality, measured by weighted peak signal-to-noise ratio. Through a stochastic approach, we optimize this quality metric, by finding the minimal degradation that needs to be introduced in a marked image in order to remove the embedded watermark. Given a set of attacks, chosen according to the considered application scenario, GA support the optimization of the parameters to be assigned to each processing operation, in order to obtain an unmarked image with perceptual quality as high as possible. Extensive experimental results demonstrate the effectiveness of the proposed evaluation tool.**

***Index Terms*—Digital image watermarking, genetic algorithms (GA), perceptual quality.**

I. Introduction

In THE age of information technology, it has become easier and easier to access and redistribute digital multimedia data. In this context, the scientific community started focusing on the growing problems related to copyright management and ownership proof. After the pioneering contribution by Cox *et al.* [1], digital watermarking techniques have been widely developed (see for instance [2]–[4], and the references therein) as an effective instrument against piracy, improper use or illegal alteration of contents [5]. Therefore, except for specific applications, the major constraint for a mark embedded into a cover work is robustness against manipulations, including a great variety of digital and analog processing operations, such as lossy compression, linear and nonlinear filtering, scaling, noise addition, etc. However, it is well known that designing an efficient watermarking algorithm is extremely challenging and the research is still in progress, proposing a variety of solutions and software packages [6], [7]. Consequently, the role of performance evaluation tools has become more and more important. As widely known in cryptography, benchmarking frameworks speed up the research in the field of digital watermarking and promote a continuous improvement of the existing techniques by identifying methods' weaknesses and failings [8].

In the literature, there are already several benchmarking tools, which standardize the process of evaluating a watermarking system on a large set of single attacks. The first proposed benchmarking tool is StirMark[1] [9], which applies a number of attacks (one at each time) to the given watermarked content and performs the detection process to check the presence of the mark. The average percentage of the correctly detected watermarks is used as a performance measure to compare different watermarking techniques. After StirMark, the so-called second generation of watermark benchmark has been introduced [10]. In particular, while performing the same process as Stirmark, Checkmark[2] and Certimark[3] packages provide higher quality performance assessment of the watermarking techniques under test. The novel features of these benchmarking tools are the introduction of new types of attacks, the use of a perceptual quality metric to measure the introduced degradation, the possibility to distinguish between watermark detection and decoding, and finally an application driven evaluation. The latest arrival in the benchmarking field is Optimark[4] [11], which provides a friendly graphical interface and it implements the same attacks as Stirmark, but with the possibility to create combinations of them. It supports the execution of multiple trials using images (automatically calculating the embedding strength that leads to the chosen image quality), attacks, keys, and messages selected by the user. As output it provides a set of performance indices and graphics characterizing robustness, payload, execution time, and breakdown limits of the under-test-technique. The main drawback of Optimark is the lack of possibility to expand the number of attacks.

In this paper, we present an innovative and flexible tool suitable to assess the robustness of digital watermarking techniques, by introducing a novel metric based on the perceptual quality evaluation for unmarked images. A set of attacks is chosen depending on the application the under-test algorithm is intended for (e.g., copyright or medical applications). Then genetic algorithms (GA) perform the search of optimal parameters to be assigned to each image processing operator, as well as the order they need to be applied in, to remove the watermark from the content while keeping the perceptual quality of the resulting image as high as possible. The recovered unmarked image turns

Manuscript received May 30, 2008; revised December 21, 2008. First published April 28, 2009; current version published May 15, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Wu Min.

G. Boato, V. Conotter, and F. G. B. De Natale are with the Department of Information Engineering and Computer Science, University of Trento, I-38040 Trento, Italy (e-mail: boato@disi.unitn.it; conotter@disi.unitn.it; denatale@disi.unitn.it).

C. Fontanari is with the Department of Mathematics, University of Trento, I-38100 Trento, Italy (e-mail: fontanar@science.unitn.it).

Digital Object Identifier 10.1109/TIFS.2009.2020362

[1]Available: http://www.petitcolas.net/fabien/watermarking/stirmark/

[2]Available: http://watermarking.unige.ch/Checkmark/

[3]Available: http://www.certimark.org/

[4]Available: http://poseidon.csd.auth.gr/optimark/

out to be as close as possible to the watermarked one in terms of the perceived quality, here measured by means of the weighted peak signal-to-noise ratio (WPSNR). We stress, however, that other metrics could be adopted as well.

The major difference with the existing benchmarking tools consists of the possibility to test the selected algorithm under a combination of attacks, evaluating the relative performance in terms of visual degradation perceived by the human visual system (HVS). We point out that the combination of more attacks produces a gain of quality in the unmarked image compared to the degradation introduced by one single image processing operator to remove the watermark. On the other hand, taking into account the effect of more than one attack at one time makes this problem nonlinear and multidimensional. Therefore, a suitable optimization technique as GA is needed to converge to a optimal or near-optimal solution.

Up to now, in the field of watermarking, the application of GA has been limited to the embedding procedure, in particular for the selection of suitable parameters to achieve imperceptibility and robustness [12]–[18]. Recently, a robust steganographic system, based on GA, was introduced in [19]. In order to create stego-images able to break the inspection of steganalytic systems, the authors employ GA to adjust cover-image values and create the desired statistic features. Finally, in [20], GA are exploited in the context of relational databases watermarking to optimize the decoding threshold. As far as we know, our present contribution is indeed the first systematic attempt to apply GA as a benchmarking tool (a preliminary version has been presented in [21]).

The paper is organized as follows: Section II presents our approach to measure the robustness of any watermarking system, outlines the involved optimization procedure introducing GA, and describes in detail the proposed tool. Next, Section III contains our extensive experimental results, while Section IV collects some concluding remarks.

II. Proposed Algorithm

A. Robustness Evaluation

Visual quality degradation due to the watermark embedding and the removing process is an important but often neglected issue to consider in order to design a fair watermarking benchmark. Given a pattern of possible attacks, the aim of this work is to find a near-optimal combination of them, which removes the mark minimizing the degradation perceived by the HVS [22]. Hence, we need to define a proper quality metric. In general, several metrics can be used to evaluate the artifacts but the most popular one is the peak signal-to-noise ratio (PSNR) metric. The success of this measure is due to its simplicity but several tests show that such a metric is not suitable to measure the quality perceived by HVS [23]. Since advanced watermarking techniques exploit the HVS, using the above metric to quantify the distortion caused by a watermarking process might result in a misleading quantitative distortion measurement. In the last years, more and more research has been concentrated on distortion metrics adapted to the HVS [24]. In [25], a modified version of PSNR, the so-called WPSNR, is introduced: it takes into account that HVS is less sensitive to changes in highly textured areas and introduces an additional parameter, called the noise visibility function (NVF), which is a texture masking function

TABLE I
Robustness Evaluation Metric

$Q \geq M(q)$	$R(q) \geq 1$	Robust
$Q < M(q)$	$R(q) < 1$	Non robust

$$\mathrm{WPSNR(dB)} = 10 \log_{10} \frac{I_{\mathrm{peak}}^2}{\mathrm{MSE} \times \mathrm{NVF}^2} \tag{1}$$

where I_{peak} is the peak value of the input image. The NVF can be modeled as a Gaussian to estimate the local amount of texture in the image. The value of NVF ranges from approximately zero for extremely textured areas, and up to one for clear smooth areas of an image

$$\mathrm{NVF} = \mathrm{norm}\left\{ \frac{\mathbf{1}}{\mathbf{1} + \delta_{\mathrm{block}}^2} \right\} \quad \in (0, 1] \tag{2}$$

where norm is a normalization function and $\delta_{\mathrm{block}}^2$ is the luminance variance of the 8×8 block. The NVF is inversely proportional to the local image energy defined by the local variance and identifies textured and edge areas where modifications are less visible. Therefore, for images with no high texture areas, WPSNR is almost equivalent to PSNR.

The main idea of this contribution is to evaluate the robustness of a watermarking system in terms of perceptual quality measured by WPSNR. Namely, fixed a set of admissible image processing operators, the robustness of a method is quantified as

$$R(q) = \frac{Q}{M(q)} \tag{3}$$

where Q is a fixed quality threshold, q is the perceptual quality of a watermarked image I_w, and $M(q)$ is the maximal perceptual quality of the unmarked image obtained from I_w by applying any combination of the selected attacks.

Given Q, chosen dependently on the application scenario, and the value of $M(q)$, found according to the process described in Section II-C, the robustness index $R(q)$ is evaluated according to (3). If $R(q)$ is greater than 1, then it is possible to remove the mark from the given image only degrading its maximal perceptual quality $M(q)$ under Q. As a consequence, the watermarking algorithm can be declared robust since a large degradation needs to be introduced in the image to remove the mark. On the other hand, the embedded watermark is not robust if $M(q)$ assumes values higher than the threshold Q (i.e., $R(q)$ is less than 1). It is then possible to assess the robustness of a watermarking technique according to Table I.

In this work, we attempt to maximize the function WPSNR, obtaining $M(q)$. Since we consider combinations of attacks, a suitable optimization technique is needed in order to avoid brute force computation.

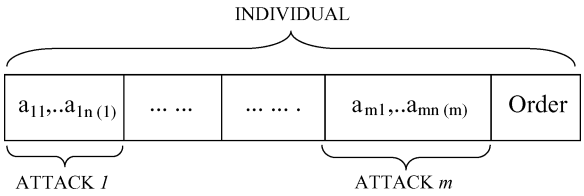

Fig. 1. Individual (chromosome) definition.

B. Genetic Algorithms

GA can be used to achieve an optimal or near-optimal solution in multidimensional nonlinear problems, such as the one to be handled in this context. GA are robust, stochastic search methods modeled on the principles of natural selection and evolution [26]. GA differ from conventional optimization techniques in that: 1) they operate on a group (population) of trial solutions (individuals) in parallel: a positive number (fitness) is assigned to each individual representing a measure of goodness; 2) they normally operate on a coding of the function parameters (chromosome) rather than on the parameter themselves; 3) they use stochastic operators (selection, crossover, and mutation) to explore the solution domain. Initially a set of individuals is encoded with chromosome-like bit strings to form an initial population. The cardinality of the set of individuals is called population size [26]. At each iteration, called generation, the genetic operators of crossover and mutation are applied to selected chromosomes in order to generate new solutions belonging to the search space. The optimization process terminates when a desired termination criterion is satisfied, for example, the maximum number of generations is reached, or the fitness value is below a fixed threshold.

GA have been widely employed to solve nonlinear optimization problems dealing with a large solution space. Although they are not guaranteed to find out the global optimum, they are less likely to get locked into a local optimum compared to traditional optimization techniques. Moreover, GA allow dealing with a larger searching space than conventional techniques; as a consequence, they are more likely to find suitable solutions to highly nonlinear and constrained problems.

The efficiency and the computational complexity of GA are heavily dependent on tuning parameters and can be calculated in terms of the number $v_{\rm op}$ of elementary operations required by the algorithms, as follows [26]:

$$v_{\rm op} = (P_C + P_M)PK_{\max} \tag{4}$$

where P_C and P_M are the crossover and mutation probability, respectively, P is the population size, and $K_{\max}$ is the number of iterations.

In this specific application, an individual represents one possible pattern of parameters to be assigned to the preselected attacks, plus the order in which they must be applied (see Fig. 1). Each attack can be parameterized by n values, according to its specification. The image processing operators must be chosen before running the tool, according to the considered application scenario and must be applied to the marked image in order to remove the embedded watermark.

The evolution of the population leads to a fine tuning of the parameterization of these attacks, such that they succeed to unmark the image while introducing a minimal degradation. Moreover, the GA support the optimization of the order of the attacks reaching an optimal or near-optimal solution. In particular, we have experimentally shown the influence of the applied order, since attacks parameterized in the same way but applied with different order may produce a loss of quality of even more than 1 dB.

C. Tool Description

In the proposed tool, GA are applied in the detection procedure of the watermarking scheme. An image previously watermarked by the algorithm to be tested and with perceived quality q is attacked with different combinations of selected image processing operators, in order to remove the embedded mark. The aim is to find a near-optimal combination of attacks to apply in order to remove the watermark, while granting a perceptual quality of the resulting image as high as possible. The algorithm robustness is then measured via (3). The optimization process is performed by GA and WPSNR is the fitness value to be maximized. We remark that the choice of this fitness function has been done to measure perceptual quality of unmarked images, but the user may adopt any other quality metric. In the following, we briefly depict the operations of the process performed by GA and reported in Fig. 2.

Step 1 Randomly generate combinations of parameters to be applied to processing operators and convert them into chromosomes. This way, an initial population is created.

The population size is typically set to 10 times the number of variables the algorithm has to deal with (length of the chromosome). Therefore, it depends on the number of values needed to parameterize each attack we want to perform in the robustness evaluation process. In Section III, we report the experimental analysis, where GA deal with four variables; thus the population size is set to 40. In this work, where the population size is not particularly large, we can reach a good efficiency of GA; in fact, we have small population over a large search space with a consequent fast convergence to optimal or near-optimal solution. Experimental results reported in Section III show in detail the efficiency of GA in terms of computational time and costs. Note that as the number of attacks to be performed increases, the population dimension and correspondingly the execution time increase as well.

Step 2 Apply each generated attack to the input image and evaluate the WPSNR of each chromosome in the current population which removes the watermark, i.e. which generates an unmarked image, and then create a new population by repeating the following steps: 1) pick as parents the chromosomes with the higher WPSNR, according to the selection rule; 2) form new children (new patterns of attacks) by applying to parents the stochastic operator of crossover with probability P_C; 3) mutate the position in the chromosome with probability P_M. In this work, widely used parameters for genetic operators have been selected (see Section III). Among all individuals of the current population which allow removing the watermark, the one that provides

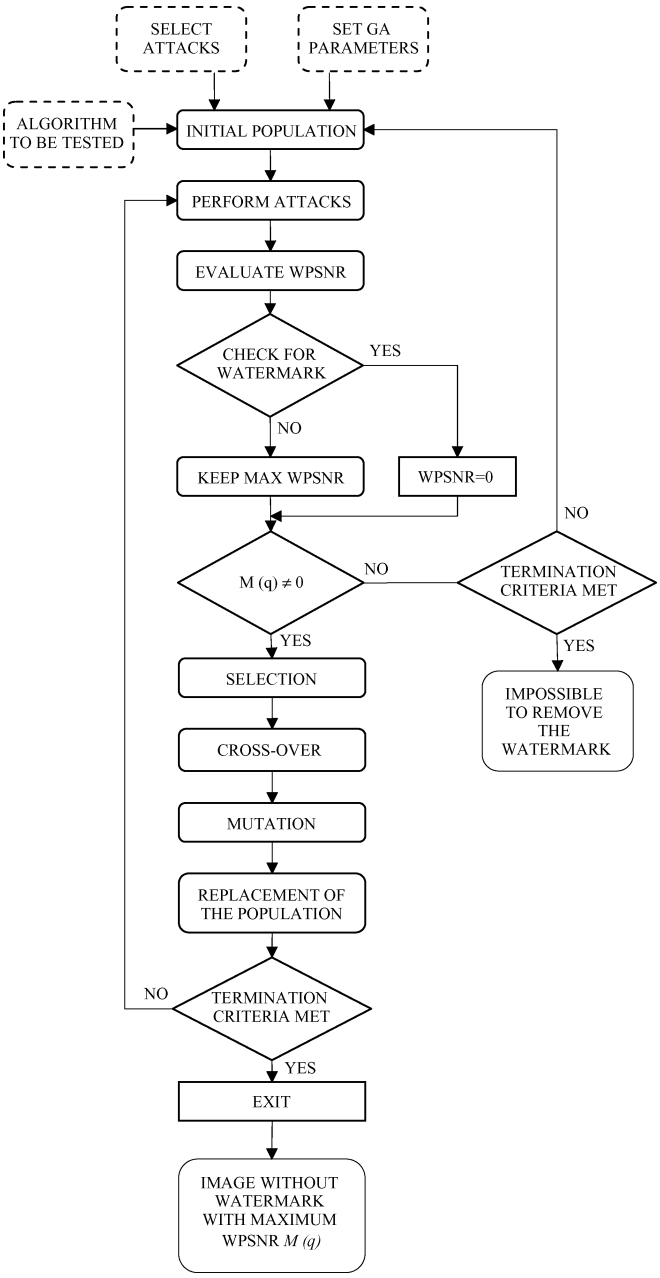

Fig. 2. Block diagram of the implemented framework.

an image with the higher WPSNR will survive to the next generation. We set to zero the fitness value of those chromosomes which do not succeed in removing the mark.

If in *Step 2* no solutions for the problem are found, i.e., none of the individuals of the population succeeds in removing the watermark, another population is re-initialized and the process is repeated until a termination criterion is met (number of generation exceeded). Consequently, the result of the test is that the analyzed watermarking technique is robust to the selected attacks.

Step 3 A new iteration with the just generated population is processed. This new population provides new attacks parameters, their corresponding fitness values are evaluated, and at every generation the individual with the highest fitness value is kept.

Step 4 The process ends when a given number of generation is exceeded (termination criteria). At that point a near-optimal combination of attacks removing the watermark from the image has been discovered. This way, the lacks of the tested algorithm with respect to the selected attacks are stressed out. At the end of the process, GA return the maximized fitness value, i.e., the maximized WPSNR $M(q)$. According to (3), it is possible to calculate the robustness index $R(q)$ and assess the global robustness performances of the watermarking technique. In particular, given the quality threshold Q, $M(q) \leq Q$ means that it is hard to remove the watermark while keeping a high perceptual quality, hence, the watermarking technique is declared to be robust. On the other hand, if $M(q) > Q$, our robustness measure indicates a serious weakness corresponding to high quality of the unmarked image.

III. Experimental Evaluation

A. Setup

In this section, we set up the robustness analysis of two perceptual-based watermarking algorithms, the former presented by Barni *et al.* in [27] and the latter proposed by Li and Cox in [28]. The main difference between them lies in the watermark recovery process, allowing watermark detection in the first case and watermark decoding in the second one.

In order to assess the robustness of these algorithms, we take into account several 512×512 grayscale common images.

The parameters for the embedding procedure are carefully selected so that the resulting watermarked images present the same WPSNR. They are then processed by the proposed GA-based tool which requires the selection of attacks, as shown in Fig. 2. Indeed, *Step 1* in Section II-C converts the attack parameters into chromosomes for initial generation. In this work, we take into consideration a combination of some (2 or 3) of the following attacks, each of them tuned by just a single parameter.

A) JPEG2000 compression, parameterized by the compression ratio ranging from 8 (no compression) down to 0.01 as a float number;
B) JPEG compression, parameterized by the quality factor ranging from 100 (no compression) down to 20 as a float number;
C) Additive white Gaussian noise (AWGN), parameterized by the noise power expressed in decibels, ranging from 0 to 40, and, for [28] parameterized in terms of standard deviation δ, ranging from 0.1 to 2;
D) Resize, parameterized by the resize factor ranging from 1 down to 0.1.
E) Amplitude Scaling, parameterized by the scaling factor, ranging from 0.1 up to 3.

The choice of the attacks will depend on the application for which the investigated algorithm is intended. We selected largely used processing operations whose combination represents a realistic scenario. Notice that the choice of both operator and parameter ranges is fully arbitrary and application driven although it affects the computational cost according to (4). Moreover, GA look also for the order the attacks are applied in, since there is no theoretical reason why the attacks should be commutative and indeed we experimentally notice

TABLE II
GA PARAMETERS SETTINGS

Population size	~ 10 times number variables
Creation function	Uniform
Fitness scaling	Proportional
Parents selection	Roulette wheel
Crossover function	Single point
Crossover probability	0.8
Elite count	1
Mutation function	Uniform
Mutation rate	0.1
Stopping criteria	1000 iterations

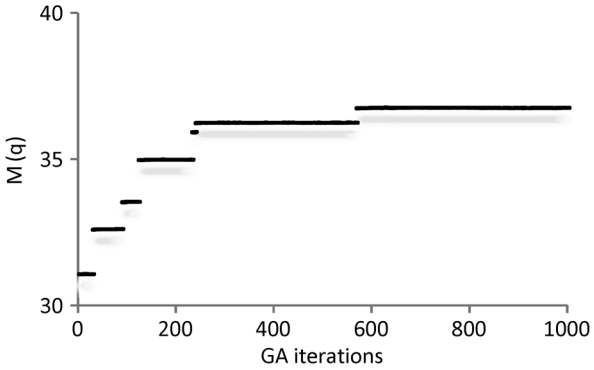

Fig. 3. Fitness trend over GA iterations.

the difference in the resulting image quality when the selected attacks are combined with different orders.

In this work, GA parameters[5] have been tuned according to standard settings [26], as reported in Table II. Fig. 3 illustrates the fitness trend in the case of algorithm [27] and the Baboon image with $q = 59$ dB (similar trends for all other simulations are not reported here). All simulations are carried on an Intel Core 2 Quad CPU at 2.4 GHz, with 2-GB Memory RAM.

B. Analysis of Algorithm [27]

This watermarking method [27] works on wavelet domain and exploits perceptual masking in order to embed the mark improving invisibility and robustness. The main advantage of this method with respect to existing algorithms operating in the discrete wavelet transform (DWT) domain is that masking is accomplished pixel by pixel by taking into account the texture and the luminance content of all image subbands. The mark w (a pseudorandom sequence) is adaptively inserted into the DWT coefficients of the three largest detail subbands, as follows:

$$\widetilde{I}_0^\theta(i,j) = I_0^\theta(i,j) + \alpha w^\theta(i,j) x^\theta(i,j) \tag{5}$$

where $\widetilde{I}_0^\theta(i,j)$ are the subband coefficients with $\theta \in 0,1,2$, α is the global parameter accounting for watermark strength, $w^\theta(i,j)$ is a weighing function considering the local sensitivity of the image to noise, and $x^\theta(i,j)$ is the mark to be embedded.

To detect the presence of the watermark, the correlation between the extracted DWT coefficients and the watermark is computed by

$$\rho = \frac{1}{3MN} \sum_{\theta=0}^{2} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \widetilde{I}_0^\theta(i,j) x^\theta(i,j) \tag{6}$$

where $2M \times 2N$ is the dimension of the host image and compared to a threshold T_ρ, chosen dependently on the admitted false alarm probability [27].

In order to evaluate the robustness of this method and compute the value $R(q)$ defined in (3):

I) tune the embedding strength α in (5) in such a way that $q = \mathrm{WPSNR}(I_w)$;
II) select the detection threshold T_ρ in such a way that the false positive probability is less than a fixed value P_{fa};
III) run the GA in order to determine $M(q)$ and set $R(q) = Q/M(q)$ as in (3).

In our simulations, the detection threshold is adaptively changed depending on the parameter α [see (5)] and imposing a probability of false alarm $P_{\mathrm{fa}} \leq 10^{-6}$ (refer to [27]).

In Table III, experimental results for the Lena and Baboon images are reported. Both are processed with the combination of three image processing operations A, C, and D described in Section III-A. For completeness sake, we report also the PSNR values for both the marked images (PSNR_m) and the unmarked ones (PSNR_u). The elapsed time for obtaining such results is almost 55 s per iteration.

We underline the weakness of the algorithm with respect to the resize operation, which plays the main role in the watermark removal process. In particular, in the case of Lena, the resulting unmarked image presents a high value of WPSNR; this highlights a robustness limitation of the algorithm under test for this image.

On the other hand, this is no longer true for Baboon: indeed, it is worth noticing that the behavior of the algorithm is image-dependent. In Figs. 4 and 5, examples of the output images (referring to Table III) are reported. The difference is due to the intrinsic nature of the watermarking algorithm. Being a perceptual method, its behavior varies depending on the texture of the content it is dealing with.

To have a whole evaluation of the robustness of the method we are analyzing referring to the single images, we calculate the robustness index $R(q)$ according to (3) following the steps described in Section II. Averaging over different watermarks we get the plots reported in Fig. 6, where the quality threshold Q is set to 40 and 35 dB. In the first case, the method results to be very robust, since $R(q) \geq 1$ for every q. In the second case, instead, the dependence of the behavior on the image content is evident. For very highly textured images, such as Baboon, the robustness of the method can be preserved using an embedding strength $\alpha \geq 0.5$. On the other hand, for the Lena image, the algorithm turns out to be not robust. The plots in Fig. 6 highlight the importance of the choice of the quality threshold Q,

[5]Available: http://www.mathworks.com/access/helpdesk/help/toolbox/gads/

TABLE III
ROBUSTNESS RESULTS OF [27] UNDER THE COMBINATION OF JPEG2000 COMPRESSION, ADDITION OF WGN AND RESIZE ATTACK

Quality parameters					Attack parameters			
q	$PSNR_m$	α	M(q)	$PSNR_u$	CR	NP	RES	Order
LENA IMAGE								
47	24.6	2.11	37.8	23.3	0.48	0	0.25	C-A-D
50	27.4	1.51	37.6	28.2	1.63	0	0.25	D-C-A
53	30.4	1.07	37.8	26.9	0.17	1	0.3	C-D-A
56	33.4	0.75	38.4	28.2	2.66	1	0.25	C-A-D
59	36.4	0.53	39.0	29.2	2.74	0	0.28	C-A-D
62	39.6	0.37	40.0	30.1	2.01	0	0.3	A-C-D
BABOON IMAGE								
47	24.3	1.57	32.5	19.4	4.3	1	0.25	C-A-D
50	27.3	1.11	32.9	20.2	4.02	0	0.26	C-A-D
53	30.4	0.78	34.4	21.0	4.14	0	0.3	A-C-D
56	33.4	0.55	34.8	21.4	3.93	0	0.31	C-A-D
59	36.4	0.39	36.5	21.9	3.86	0	0.35	D-A-C
62	39.3	0.27	37.8	22.3	3.75	0	0.38	A-C-D

Fig. 4. Lena image watermarked with $q = 47$ dB (left) and unmarked with $M(q) = 37.37$ dB (right).

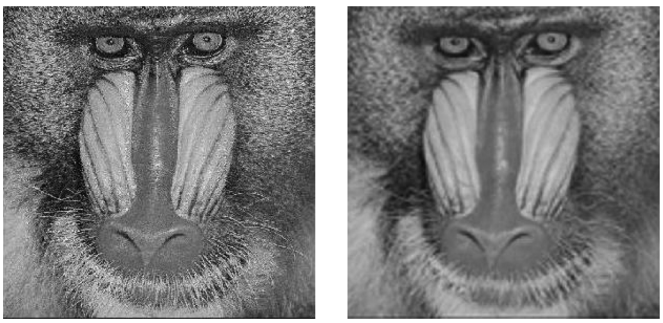

Fig. 5. Baboon image watermarked with $q = 47$ dB (left) and unmarked with $M(q) = 32.53$ dB (right).

which strictly depends on the application scenario and greatly influences the robustness assessment.

We stress that the user of the proposed tool can properly choose the signal processing operations to give as input to the framework and the results presented here are just an example of application.

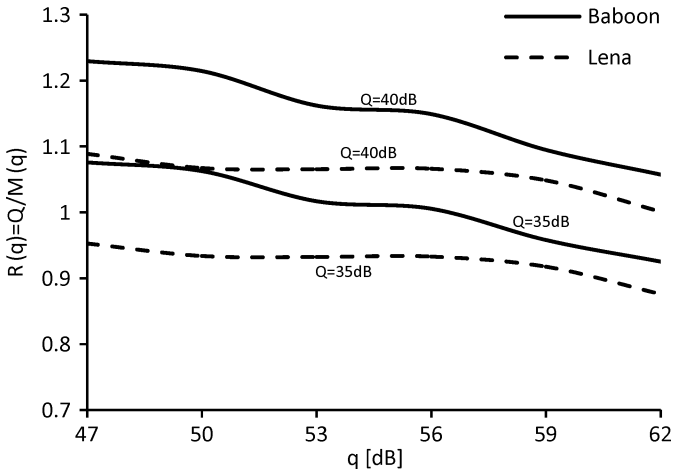

Fig. 6. Performance plots for $Q = 40$ dB and $Q = 35$ dB under the combination of JPEG2000 compression, addition of WGN and resize attack.

C. Analysis of the Algorithm [28]

The algorithm [28] is an important enhancement of traditional quantization index modulation (QIM) methods overcoming the sensitivity to volumetric changes of QIM schemes by adaptively selecting the quantization step size according to a modified version of Watson's model [29]. The authors first describe the need of dither modulation (DM) to address the problem of poor fidelity in some areas of the cover object in traditional QIM schemes due to the fixed quantization step size. DM, first proposed by Chen and Wornell as an extension of the original QIM method [30], introduces a pseudorandom dither signal reducing in such a way the perceptual artifacts caused by quantization. The embedding function embeds the message bit m_n by

$$y_n(x_n, m_n) = Q\left(x_n + d(n, m_n), \Delta\right) - d(n, m_n) \quad (7)$$

where

$$d[n,1] = \begin{cases} d[n,0] + \frac{\Delta}{2} & d[n,0] < 0 \\ d[n,0] - \frac{\Delta}{2} & d[n,0] > 0 \end{cases} \quad n = 1, 2, \ldots, L \quad (8)$$

and $d[n,0]$ is a pseudorandom signal with a uniform distribution over $[-\Delta/2, \Delta/2]$ and L is the number of samples.

To improve fidelity, Watson's perceptual model is adopted to calculate a "slack," that is the maximal distortion allowed for each discrete cosine transform (DCT) coefficient. A slack is then employed to adaptively adjust the quantization step size used in the quantization process of the DCT coefficients.

Operating in the DCT domain, slacks calculated from Watson's model are multiplied by a global constant G in order to get the final quantization step size for each DCT coefficient (notice that G must be known in the decoding phase). Moreover, G is tuned to empirically control the quality of the watermarked image.

QIM schemes are generally weak with respect to volumetric scaling. To address this problem, Watson's model has been modified so that the quantization step size is scaled linearly with respect to scaling amplitude of the volumetric attack. This way decoding can be correctly performed. Rational dither modulation (RDM) is then introduced: in particular, the authors propose to calculate the quantization step size for the current block using the slacks of previously watermarked blocks. The final perceptually adaptive RDM method is referred to as rational dither modulation–modified Watson model (RDM-MW). In the detection phase, two signals, namely $S_r(n,0)$ and $S_r(n,1)$, are calculated as follows:

$$\begin{aligned} S_r(n,0) &= Q\left(r_n + d[n,0], \Delta\right) - d[n,0] \\ S_r(n,1) &= Q\left(r_n + d[n,1], \Delta\right) - d[n,1] \end{aligned} \quad (9)$$

where r_n is the received signal. The recovered bit is the closest in the Euclidean metric to the received signal r

$$\hat{m}_n = \underbrace{\arg\min}_{l \in 0,1} \left(r_n - S_r(n,l)\right)^2. \quad (10)$$

Since one message bit is spread into a sequence of N samples, the code rate is $1/N$ and the detected message bit is determined by accumulating the two Euclidean distances for N samples, as follows:

$$\hat{m}_n = \underbrace{\arg\min}_{l \in 0,1} \sum_{h=(n-1)N+1}^{nN} \left(r_h - S_r(h,l)\right)^2 \quad n = 1, 2, \ldots, L/N. \quad (11)$$

This watermarking method implies a decoding process, which is evaluated in terms of bit-error rate (BER).

In this analysis context, we embed a message of length 8129 using a 1/31 rate repetition code, following the reference paper [28], and the BER threshold is fixed to 0.2. In order to compute the value $R(q)$ defined in (3), we proceed as follows:

I) tune the global constant G in such a way that $q = \mathrm{WPSNR}(I_w)$;
II) fix a BER threshold T_{BER} discriminating between watermarked and unwatermarked images;
III) run the GA in order to determine $M(q)$ and set $R(q) = Q/M(q)$ as in (3).

TABLE IV
EMBEDDING VALUES FOR [28]

q	G	$\mathrm{PSNR_m}$	WD	DWR
BABOON IMAGE				
47	0.886	25.08	45.9	9.51
53	0.443	31.09	23.01	15.52
59	0.218	37.22	11.46	21.65
65	0.104	43.56	5.81	27.99
71	0.041	51.06	3.29	35.49
LENA IMAGE				
47	1.149	23.28	63	8.25
53	0.574	29.27	32.02	14.24
59	0.284	35.36	16	20.33
65	0.136	41.71	8.15	26.67
71	0.053	49.47	4.48	34.45

TABLE V
ROBUSTNESS RESULTS OF [28] UNDER THE COMBINATION OF JPEG COMPRESSION, ADDITION OF WGN AND AMPLITUDE SCALING ATTACK

Quality parameters			Attack parameters			
q	M(q)	$\mathrm{PSNR_u}$	QF	δ	SF	Order
BABOON IMAGE						
47	51.4	25.8	64	1.2	0.98	E-B-C
59	55.2	37.6	91	1.3	0.98	B-E-C
71	55.4	47.0	98	0.1	0.99	B-C-E
LENA IMAGE						
47	49.4	24.5	54	1.0	0.98	C-B-E
59	54.0	36.3	89	0.9	0.98	E-C-B
71	54.4	45.3	98	1.0	0.98	E-C-B

Concerning point I, we selected G in order to have values of q and corresponding PSNR_m as reported in Table IV. The document-to-watermark ratio and the Watson distance of the marked image are also reported, following the setting choices in [28].

First we test the algorithm under the combination of attacks suggested in the reference paper: JPEG compression (B), AWGN (C), and amplitude scaling (E). Results for the Baboon and Lena images are reported in Table V. The elapsed time for obtaining such results is almost 100 s per iteration. As underlined in the description of the algorithm, we can notice a weakness with respect to JPEG compression. This attack is able to remove the mark while introducing a minimal degradation in the resulting image. This is not surprising considering QIM-based algorithms. Notwithstanding, these tests allow us to demonstrate the effectiveness of the proposed tool. Fig. 7 plots the robustness index and underlines once again the weakness of

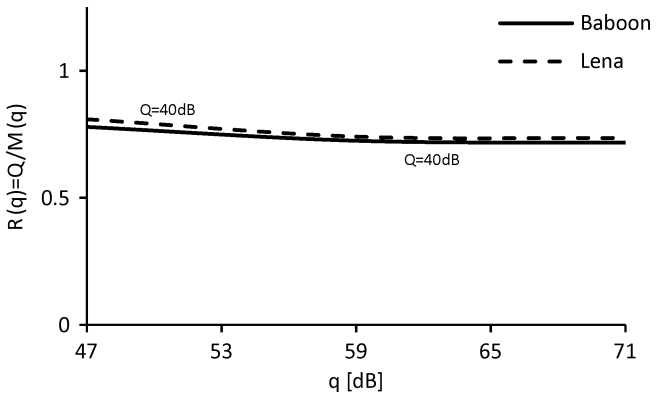

Fig. 7. Performances plot for $Q = 40$ dB under the combination of JPEG compression, addition of WGN and amplitude scaling attack.

TABLE VI
ROBUSTNESS RESULTS OF [28] UNDER THE COMBINATION OF JPEG2000 COMPRESSION, ADDITION OF WGN AND AMPLITUDE SCALING ATTACK

Quality parameters			Attack parameters			
q	M(q)	$PSNR_u$	CR	δ	SF	Order
BABOON IMAGE						
47	35.8	22.3	0.55	1.8	0.97	C-A-E
53	40.9	27.2	1.12	3.1	0.97	A-E-C
59	45.7	32.0	1.80	0.8	0.98	C-E-A
65	49.2	37.6	3.53	4.0	0.97	A-E-C
71	55.1	44.9	6.01	1.8	0.98	A-C-E
LENA IMAGE						
47	37.6	22.9	0.17	1.2	0.97	E-C-A
53	42.1	28.3	0.31	2.0	0.97	A-C-E
59	45.6	33.1	0.48	2.5	0.97	C-A-E
65	49.8	38.1	0.85	0.8	0.98	A-C-E
71	53.2	43.4	3.87	2.2	0.98	C-A-E

the algorithm, since $R(q) < 1$ for every q. Notice the similar behavior of the algorithm for different images.

A further experimental analysis is carried out in order to analyze the algorithm under the effect of JPEG2000 compression (A) instead of classical JPEG. This attack has been chosen because it is expected to become the new standard for image compression; it is, therefore, interesting to examine the robustness of watermarking algorithms with respect to this attack. The elapsed time for obtaining such results is about 120 seconds per iteration.

In Table VI, the obtained results for the Baboon and Lena images are reported. It is clear that as the quality of the marked image increases, it becomes easier and easier to remove the mark, introducing little degradation in the image.

It is worth stressing the crucial role of JPEG2000 compression and AWGN in the watermark removing process, while the amplitude scaling does not enter into play: indeed the algorithm has been designed to be resistant to this last kind of attack.

By computing the robustness index $R(q)$ of (3), we get the plot in Fig. 8. By setting the quality threshold Q to 40 dB, we

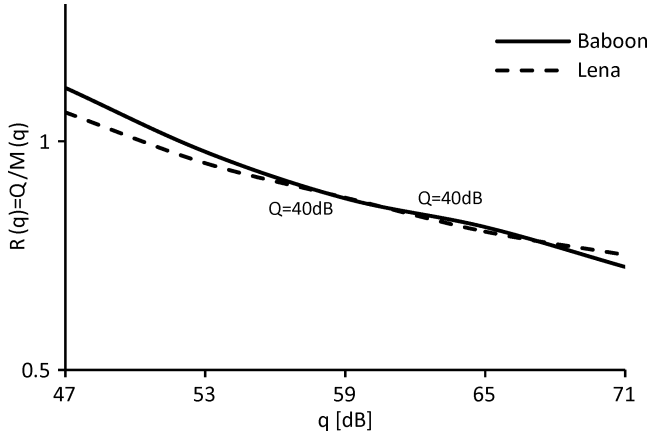

Fig. 8. Performance plot for $Q = 40$ dB under the combination of JPEG2000 compression, addition of WGN and amplitude scaling attack.

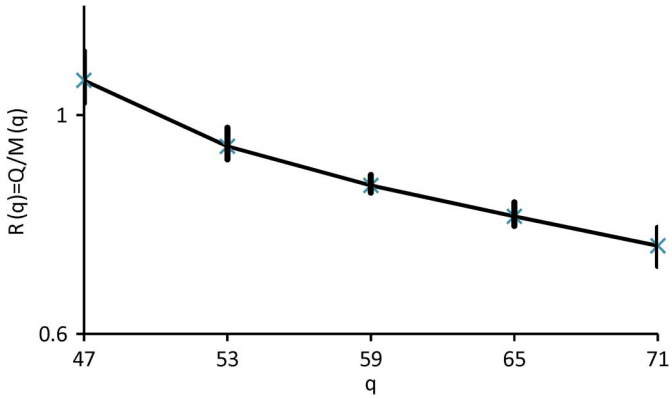

Fig. 9. Evidence of image independency.

can assess the robustness of the under-test-algorithm for values of the perceptual quality of the input image q lower than 53 dB ($R(q) > 1$ for $q < 53$). It means that the mark has to be embedded with a quite high strength to be robust in this case. Once again, notice the uniform behavior of the algorithm for different images.

This highlights a big advantage of this algorithm: it seems to be independent of the image content, thus allowing a wider application. In order to verify this statement, we have repeated the last experiment over ten different standard images.[6] Results have been plot in Fig. 9, where both mean and variance are reported for different values of q and $Q = 40$ dB. This provides clear evidence of the image independency of the algorithm, since the variance of WPSNR is at most 5%. Finally we analyze the algorithm avoiding compression. We apply the combination of the two attacks: AWGN (C) and volumetric scaling (E). GA still looks for the order the two operators need to be performed. In Table VII, results for the Baboon and Lena images are reported. The elapsed time for obtaining such results is almost 85 s per iteration.

We stress the fact that the quality of the unmarked images is in this case substantially decreased, compared to previous experiments. This is mainly explained by the choice of the selected

[6]Baboon, Lena, Boat, Cameraman, Peppers, Barbara, Goldhill, Clown, Airplane, Walkbridge.

TABLE VII
ROBUSTNESS RESULTS OF [28] UNDER THE COMBINATION OF ADDITION OF WGN AND AMPLITUDE SCALING ATTACK

Quality parameters			Attack parameters		
q	M(q)	$PSNR_u$	δ	SF	Order
BABOON IMAGE					
47	13.4	7.2	0.1	2.52	E-C
53	13.9	7.8	2.1	2.17	C-E
59	14.6	8.5	0.6	1.98	C-E
65	20.2	15.5	4.0	0.61	C-E
71	55.1	45.1	1.8	0.98	C-E
LENA IMAGE					
47	12.7	6.6	3.0	2.97	C-E
53	13.2	7.1	3.4	2.54	E-C
59	14.6	8.6	4.0	0.26	E-C
65	20.0	13.9	4.0	0.59	C-E
71	52.6	42.7	2.6	0.98	C-E

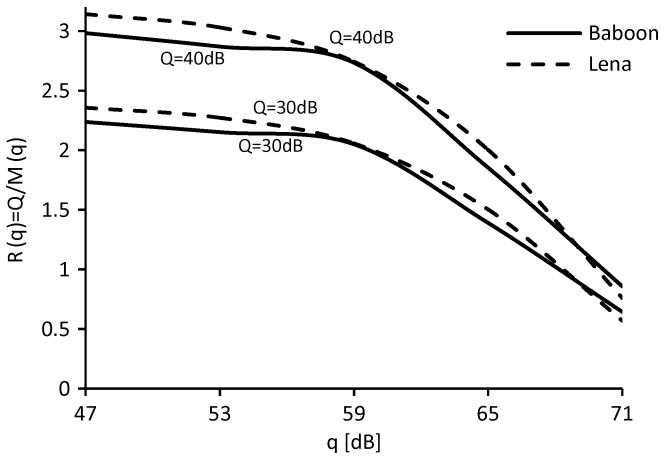

Fig. 10. Performance plots for $Q = 40$ dB and $Q = 30$ dB under the combination of addition of WGN and amplitude scaling attack.

image processing operators: in fact, the analyzed algorithm is designed to resist against the chosen attacks.

In Fig. 10, the robustness index $R(q)$ is plotted for different values of the quality threshold Q. In particular, it is shown that robustness is preserved even in the case of a low quality threshold $(Q = 30\ \text{dB})$ and the combination of both attacks, as expected from [28].

IV. CONCLUSION

In this paper, we have presented an innovative benchmarking tool to evaluate the robustness of any digital watermarking technique considering the quality of the unmarked images in terms of perceived quality. Therefore, a new metric based on WPSNR is introduced. The goal is to remove the watermark from a content while maximizing perceptual quality. So, given a set of attacks, we look for a parameterization able to remove the watermark, optimizing the WPSNR of the unmarked image. This nonlinear optimization problem is supported by GA. The effectiveness of the present tool has been demonstrated by extensive simulations pointing out the weaknesses of two well-known methods. We also point out that with the proposed tool, it is possible to fairly compare two different watermarking algorithms performing the same kind of watermark recovery (namely, either both detection or both decoding).

ACKNOWLEDGMENT

The authors would like to warmly thank Q. Li and I. J. Cox for kindly providing the code of their algorithm.

REFERENCES

[1] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.
[2] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*. London: Academic, 2002.
[3] J. Eggers and B. Girod, *Informed Watermarking*. Norwell, MA: Kluwer, 2002.
[4] M. Barni and F. Bartolini, "Watermarking systems engineering. Enabling digital assets security and other applications," in *Signal Processing and Communications Series*. Boca Raton, FL: CRC Press, 2004.
[5] E. Lin, A. Eskicioglu, R. Lagendijk, and E. J. Delp, "Advances in digital video content protection," *Proc. IEEE*, vol. 93, no. 1, pp. 171–183, Jan. 2005.
[6] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proc. IEEE*, vol. 87, no. 7, pp. 1079–1107, Jul. 1999.
[7] B. Furht and D. Kirovski, *Multimedia Watermaking Techniques and Application*. New York: Auerbach, 2006.
[8] B. Macq, J. Dittman, and E. J. Delp, "Benchmarking of image watermarking algorithms for digital rights management," *Proc. IEEE*, vol. 92, no. 6, pp. 971–984, Jun. 2004.
[9] A. P. Peticolas, R. J. Anderson, and M. G. Kuhn, "Attacks on copyright marking systems," in *Proc. 2nd Int. Workshop on Information Hiding*, Apr. 1998, pp. 219–239.
[10] S. Pereira, S. Voloshynovskiy, M. Madueno, S. Marchand-Maillet, and T. Pun, "Second generation benchmarking and application oriented evaluation," in *Proc. Int. Workshop on Information Hiding*, Apr. 2001, pp. 340–353.
[11] V. Solachidis, A. Tefas, N. Nikolaidis, S. Tsekeridou, A. Nikolaidis, and I. Pitas, "A benchmarking protocol for watermarking methods," in *IEEE Int. Conf. Image Processing*, Oct. 2001, pp. 1023–1026.
[12] N. Zhong, Z. He, J. Kuang, and Z. Zhuo, "An optimal wavelet based image watermarking via genetic algorithm," in *3rd Int. Conf. Natural Computation*, Aug. 2007, vol. 3, pp. 103–107.
[13] S. P. Maity, M. K. Kundu, and P. K. Nandi, "Genetic algorithm for optimal imperceptibility in image communication through noisy channel," in *Proc. Int. Conf. Neural Information Processing*, Nov. 2004, pp. 700–705.
[14] C. H. Huang and J. L. Wu, "A watermark optimization technique based on genetic algorithms," in *Proc. SPIE—Visual Communications Image Processing*, May 2000, pp. 516–523.
[15] C. S. Shieh, C. Huang, F. H. Wang, and J. S. Pan, "Genetic watermarking based on transform-domain techniques," *Elsevier Pattern Recognit.*, vol. 37, pp. 555–565, Mar. 2004.
[16] Y. L. Chang, K. T. Sun, and Y. H. Chen, "ART2-based genetic watermarking," in *Proc. Int. Conf. Advanced Information Networking and Applications*, Mar. 2005, vol. 1, pp. 729–734.
[17] P. Kumsawat, K. Attakitmongcol, and A. Srikaew, "A new approach for optimization in image watermarking by using genetic algorithm," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4707–4719, Dec. 2005.
[18] F. Y. Shih and F.-T. Wu, "Enhancement of image watermark retrieval based on genetic algorithms," *J. Visual Commun. Image Represent.*, vol. 16, pp. 115–133, Apr. 2005.
[19] Y.-T. Wu and F. Y. Shih, "Genetic algorithm based methodology for breaking the steganalytic systems," *IEEE Trans. Syst., Man and Cybern. B, Cybern.*, vol. 36, no. 1, pp. 24–31, Feb. 2006.
[20] M. Shehab, E. Bertino, and A. Ghafoor, "Watermarking relational databases using optimization based techniques," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 1, pp. 116–129, Jan. 2008.
[21] G. Boato, V. Conotter, and F. G. B. De Natale, "GA-based robustness evaluation method for digital image watermarking," in *Proc. IWDW 2007*, Guangzhou, Dec. 2007.

[22] J. Delaigle, C. Devleeschouwer, B. Macq, and I. Lagendijk, "Human visual system features enabling watermarking," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2002, vol. 2, pp. 489–492.
[23] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. New York: Morgan & Claypool, 2006.
[24] D. Levicky, P. Foris, and N. Nikolaidis, "Human visual system models in digital image watermarking," *Radio Eng.*, vol. 13, pp. 38–43, 2004.
[25] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T. Pun, "A stochastic approach to content adaptive digital image watermarking," in *Proc. Int. Workshop Inform. Hiding*, 1999, pp. 211–236.
[26] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley, 1999.
[27] M. Barni, F. Bartolini, and A. Piva, "Improved wavelet-based watermarking through pixel-wise masking," *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 783–791, May 2001.
[28] Q. Li and I. J. Cox, "Using perceptual models to improve fidelity and provide resistance to volumetric scaling for quantization index modulation watermarking," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 2, pp. 127–139, Jun. 2007.
[29] A. B. Watson, "DCT quantization matrices optimized for individual images," in *Human Vision, Visual Processing, and Digital Display IV*, 1993, vol. SPIE-1913, pp. 202–216.
[30] B. Chen and G. W. Wornell, "Quantization index modulation methods for digital watermarking and information embedding of multimedia," *J. VLSI Signal Process. Syst. Signal, Image, Video Technol., Special Issue Multimedia Signal Process*, vol. 47, pp. 7–33, Feb. 2001.

Giulia Boato (S'04–A'06) received the M.Sc. degree in mathematics in 2002, and the Ph.D. degree in information and communication technologies in 2005, both from the University of Trento, Italy.

In 2006, she was a visiting researcher at the University of Vigo, Spain. Currently, she is Assistant Professor of Telecommunications at the University of Trento, working in the Multimedia Signal Processing and Understanding Laboratory. Her research interests are focused on image and signal processing, with particular attention to multimedia data protection and data hiding.

Valentina Conotter received the M.Sc. degree in telecommunication engineering from the Faculty of Engineering, University of Trento, Italy. Currently, she is working toward the Ph.D. degree at the ICT International Doctorate School within the Department of Engineering and Computer Science, University of Trento, working in the Multimedia Signal Processing and Understanding Laboratory.

Her research activity is mainly focused on multimedia security and forensics, with particular attention on digital watermarking and image forensics techniques.

Francesco G. B. De Natale (M'97–SM'03) received the M.Sc. degree in electronic engineering and the Ph.D. degree in telecommunications from the University of Genoa, Italy, in 1990 and 1994, respectively.

He is Professor of Telecommunications at the University of Trento, Italy. He is the Head of the Department of Information Engineering and Computer Science (DISI), and is responsible for the Multimedia Signal Processing and Understanding Laboratory. His research interests are focused on multimedia signal processing, analysis and transmission, with particular attention to image and video data. He was General Co-Chair of the Packet Video Workshop in 2000, and Technical Program Co-Chair of the IEEE International Conference on Image Processing (ICIP) in 2005 and of the IEEE International Conference on Multimedia Services Access Netoworks (MSAN, now Mobimedia) in 2005.

In 1998, Prof. De Natale was co-recipient of the IEEE Chester-Sall Best Paper Award.

Claudio Fontanari (M'08) received the Laurea degree in mathematics from Trento University, Italy, and the Ph.D. degree in mathematics from Scuola Normale Superiore, Pisa, Italy.

He is currently Assistant Professor, Department of Mathematics, University of Trento, Italy. His research interests range from algebraic geometry to multimedia security.