A BLIND AND ROBUST WATERMARKING SCHEME USING THE WAVELET TRANSFORM

Stewart I. Fraser
Department of Engineering
University of Aberdeen
Aberdeen, AB24 3UE, UK
Email: s.i.fraser@abdn.ac.uk

Alastair R. Allen
Department of Engineering
University of Aberdeen
Aberdeen, AB24 3UE, UK
Email: a.allen@abdn.ac.uk

ABSTRACT
A novel, quantization based method for watermarking digital images is presented here. This new scheme utilises implicit visual masking by only inserting a watermark bit into wavelet coefficients of high magnitude. Also, this watermarking technique is blind (*i.e.*, neither the original uncorrupted image nor any side information is required in the recovery process) as well as being very computationally efficient. This new watermarking algorithm combines and adapts various aspects from two existing watermarking methods. Results show that the newly presented method improves upon both of these existing techniques.

KEY WORDS
Watermarking Techniques, Digital Image Watermarking, Wavelet.

1 Introduction

This paper introduces a new quantization based, blind watermarking algorithm operating within the wavelet domain. The motivation for this new algorithm was based upon various aspects from watermarking schemes presented by Dugad *et al.* [1] and Inoue *et al.* [2]. The new algorithm improves upon the Dugad algorithm in that it can survive the same malicious attacks whilst producing marked images of greater visual quality. An improvement is made upon the semi-blind Inoue scheme as the new method does not require a file containing the positions of the marked coefficients (*i.e.*, the new watermarking scheme is blind).

2 Background

Previously, Dugad *et al.* [1] presented an additive watermarking method operating in the wavelet domain. A three level wavelet transform with a Daubechies 8-tap filter was used; no watermark was inserted into the low-pass subband. Unlike some non-blind watermarking schemes [3, 4], this scheme allowed a watermark to be detected without requiring access to the uncorrupted original image (*i.e.*, it is a blind watermarking system).

The Dugad scheme also performed implicit visual masking as only wavelet coefficients with a large enough magnitude were selected for watermark insertion. Wavelet coefficients of large magnitude correspond to regions of texture and edges within an image. This has the effect of making it difficult for a human viewer to perceive any degradation to an image marked via this scheme. Also, because wavelet coefficients of large magnitude are perceptually significant, it is difficult to remove the watermark without severely distorting the marked image.

The most novel aspect of this scheme was the introduction of an *image sized watermark* consisting of pseudorandom real numbers. However, only a few of these watermark values are added to the host image. Using an image sized watermark *fixes* the locations of the watermark values; thus, there is no dependence on the ordering of significant coefficients in the correlation process for watermark detection. This is advantageous as the correlation process is extremely sensitive to the ordering of significant coefficients and any change in this ordering (via image manipulations) can result in a poor detector response.

Another watermaking algorithm operating upon significant coefficients within the wavelet domain (implemented via 5/3 taps symmetric short kernel filters) was presented by Inoue *et al.* [2]. This method takes a three level wavelet transform of the image to be watermarked and inserts the watermark into the detail coefficients at the coarsest scales (LH3, HL3 and HH3; the lowpass component LL3 is excluded).

The Inoue scheme is a quantization based watermarking technique which aims to modify wavelet coefficients of high magnitude thus embedding the watermark into edge and textured regions of an image. The quantization process used by this scheme is very straightforward and simple to implement as it requires a file to be saved detailing the locations of where the watermark bits were embedded. It is thus a semi-blind scheme as opposed to a blind scheme.

3 Advantages and disadvantages of the Dugad and Inoue watermarking algorithms

The Dugad algorithm has three main advantages: (1) It is a blind algorithm. (2) It incorporates implicit visual masking, thus, the watermark is inserted into the perceptually significant areas of an image via a simple and straightforward process. (3) It uses an image sized watermark to negate the order dependence of significant coefficients in the detection process.

There are two main disadvantages to the Dugad algorithm: (1) It embeds the watermark in an additive fashion. This is a drawback as blind detectors for additive watermarking schemes must correlate the possibly watermarked image coefficients with the known watermark in order to determine if the image has or has not been marked. Thus, the image itself must be treated as noise which makes detection of the watermark exceedingly difficult [5]. In order to overcome this, it is necessary to correlate a very high number of coefficients (which in turn requires the watermark to be embedded into many image coefficients at the insertion stage). This has the effect of increasing the amount of degradation to the marked image. (2) The detector can only tell if the watermark is present or absent. It cannot recover the actual watermark.

The Inoue algorithm has three main advantages: (1) It uses a scalar quantization process to embed the watermark. This is an advantage as quantization based watermarking schemes do not suffer from host image interference [5] (unlike additive watermarking schemes). Hence, detectors from quantization based watermarking techniques can operate successfully using a much smaller watermark than is possible for additive schemes. This has the knock-on effect of reducing the amount of degradation suffered by a marked image. (2) The detector can recover the binary watermark sequence thereby allowing the user to see it. (3) Like the Dugad scheme, the watermark is embedded into perceptually significant coefficients.

The Inoue algorithm has one main disadvantage: (1) It is a semi-blind algorithm. This is because it requires a file containing the locations of where the watermark was embedded in order for the detector to work.

4 A new approach

It is possible to share the advantages of both the Dugad and Inoue watermarking schemes whilst removing most of the disadvantages. This can be achieved by using Dugad's idea of an image sized watermark in conjunction with adapted versions of Inoue's scalar quantization insertion/detection techniques. The resultant watermarking system will be blind and quantization based. It will employ a watermark equal in size to the detail subbands from the coarsest wavelet level and only perceptually significant coefficients will be used to embed watermark bits. In summary, this new technique improves upon the Dugad method by using a quantize and replace insertion process (rather than an additive insertion process). Thus, for comparable robustness performance, the new method will produce watermarked images with less degradation than the Dugad scheme. It improves upon the Inoue scheme by having no need for a position file in the recovery process.

A flow diagram detailing the necessary steps is shown in Figure 1. Note that the quantization and dequantization steps are explained in more detail in Sections 4.1 and 4.2, respectively.

4.1 Embedding

The watermark embedding process transforms the host image into the wavelet domain (the current implementation uses Daubechies wavelets of length 4). Next, all the coefficients in the third wavelet level (excluding the LL subband) with magnitude greater than *T1* and magnitude less than *T2* are selected. A binary watermark the same size as the entire third level of the wavelet transform is created using a secret key (which is a seed to a random number generator).

The selected wavelet coefficients are then quantized in order to embed a watermark bit. The value that the selected coefficients are quantized to depends upon whether they are embedding a 1 or a 0. A selected wavelet coefficient, w^s_{ij}, will embed a 1 if the value in the watermark file at the same location, x_{ij}, is 1. Alternatively, w^s_{ij} will embed a 0 if x_{ij} is 0. Thus, the adapted Inoue quantization method for watermark insertion is:
If $x_{ij} = 0$ then $w^s_{ij} = \text{sgn}(w^s_{ij})(T1 + X1)$
If $x_{ij} = 1$ then $w^s_{ij} = \text{sgn}(w^s_{ij})(T2 - X1)$
The *X1* parameter narrows the range between the two quantization values of *T1* and *T2* in order to aid robust oblivious detection (see Section 4.2).

After all the selected coefficients have been quantized, the inverse wavelet transform is applied to all the wavelet coefficients and the watermarked image is obtained.

4.2 Detection

For oblivious detection, the wavelet transform of a possibly corrupted watermark image is taken. Then all the wavelet coefficients of magnitude greater than or equal to *T1 + X2*

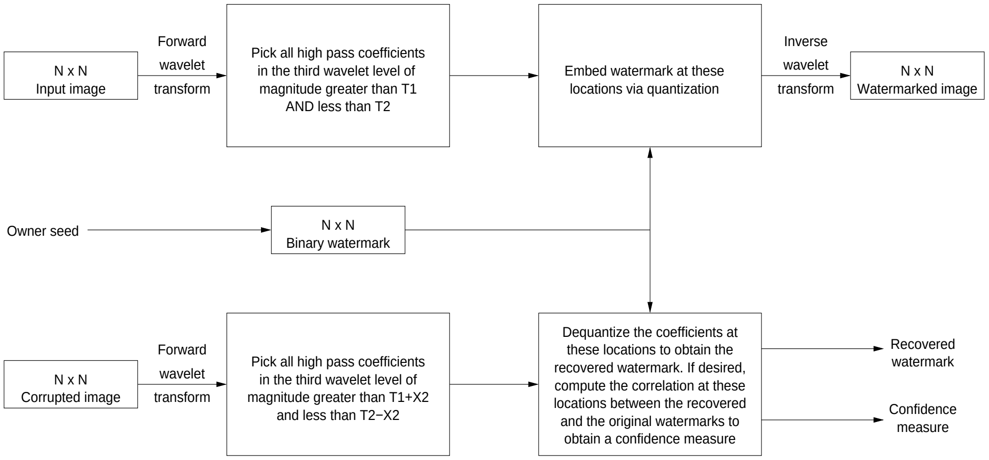
N x N
Input image
Forward
wavelet
transform
Pick all high pass coefficients in the third wavelet level of magnitude greater than T1 AND less than T2
Embed watermark at these locations via quantization
Inverse
wavelet
transform
N x N
Watermarked image
Owner seed
N x N
Binary watermark
N x N
Corrupted image
Forward
wavelet
transform
Pick all high pass coefficients in the third wavelet level of magnitude greater than T1+X2 and less than T2−X2
Dequantize the coefficients at these locations to obtain the recovered watermark. If desired, compute the correlation at these locations between the recovered and the original watermarks to obtain a confidence measure
Recovered
watermark
Confidence
measure

Figure 1. The blind quantization based watermarking scheme. The top part shows the insertion process whereas the bottom part shows the detection process.

and less than or equal to *T2* - *X2* are selected; these shall be denoted via w'^{s}_{ij}.

Note that *X2* should be less than *X1*. In the insertion process, all wavelet coefficients with a magnitude greater than *T1* and less than *T2* are selected and then quantized to either *T1* + *X1* or *T2* - *X1*. In the recovery process, all the wavelet coefficients of magnitude greater than or equal to *T1* + *X2* and less than or equal to *T2* - *X2* are selected to be dequantized. This helps ensure that all the marked coefficients are recovered and dequantized after being attacked. Also, unmarked coefficients are unlikely to drift into the range of selected coefficients after an attack. The introduction of the *X1* and *X2* parameters to the watermarking algorithm gives a degree of tolerance to the system against attacks, *i.e.*, they collaborate to give a noise margin.

A watermark bit is decoded for each of the selected wavelet coefficients via the same process described by Inoue:
If $|w'^{s}_{ij}| < (T1 + T2)/2$, a 0 bit is recovered
If $|w'^{s}_{ij}| \geq (T1 + T2)/2$, a 1 bit is recovered
The recovered watermark is then correlated with the original copy of the watermark file (obtained via the secret key) only in the locations of the selected coefficients. This allows a confidence measure to be ascertained for the presence or non-presence of a watermark in an image.

5 Results

This section outlines the results obtained by Dugad's scheme and the newly proposed scheme. It is the aim of the new scheme to be as robust as the Dugad scheme without degrading the marked images to the same extent. This newly proposed blind scheme improves upon the semi-blind Inoue scheme as it does not require a file containing the locations of the coefficients that were marked.

In order to measure the degradation suffered by host images after watermark insertion, the Peak Signal to Noise Ratio (PSNR) metric and the Watson Metric [6, 7] are used. The Watson Metric computes the Total Perceptual Error (TPE) which is an image quality metric based upon the Human Visual System (HVS). It takes contrast sensitivity, luminance masking and contrast masking into account when calculating a perceptual error value (unlike the PSNR, which merely measures the differences between pixels without considering the HVS). The higher the TPE value, the more degraded an image would appear to a human viewer. The Checkmark package [8] (*WatsonMetric.m*) was used to determine the TPE.

For all the tests in this paper, MATLAB 6.0.0.88 Release 12 was used. JPEG compression was carried out via the *imwrite* function which uses the Independent JPEG Group's (www.ijg.org) LIBJPEG library. All tests were performed upon an 8-bit (greyscale), 256×256, Lena image.

5.1 Dugad's results

The Dugad algorithm was encoded and the following parameters were set: *T1* = 40, *T2* = 50 and $\alpha = 0.2$ (the same parameters that were used in the Dugad paper). The watermarked image was then attacked with JPEG quality 5 (Q5), quality 10 (Q10) and quality 15 (Q15), median filter with kernel size 3×3 and 5×5, Gaussian noise addition ($\sigma^2 = 375$), impulse noise (normalised density of

	No attack	JPEG			Median filter		Noise addition		Cropping	Half sizing
		Q5	Q10	Q15	3×3	5×5	Gaussian	Impulse		
Average correlation	32.43	12.17	17.83	20.85	25.16	17.74	14.57	12.96	22.04	16.16
Average detector threshold	9.00	9.12	9.20	9.16	9.26	9.38	7.77	8.04	9.83	8.51
Average WM length in	2552	2552	2552	2552	2552	2552	2552	2552	2552	2552
Average WM length out	1788	1534	1560	1617	1274	933	3151	3920	1016	1957
Average PSNR (dB)	37.38	37.40	37.50	37.42	37.29	37.52	37.24	37.43	37.39	37.27
Average TPE	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
Failures (false negatives)	0	5	1	0	0	0	0	0	0	0

Table 1. Dugad results. *T1* = 40, *T2* = 50 and α = 0.2. Each test was run 30 times upon the Lena image with a different seed and the average result was then calculated. Note that WM refers to WaterMark.

0.015), cropping (from rows 60 to 190 and from columns 60 to 190) and half-sizing. Also, each of the attacks was performed 30 times upon each watermarked Lena image (using a different seed each time) and then averaged.

Further confirmation of the quantitative image quality values was obtained by using a secondary implementation of the Dugad algorithm available from [5]. For thirty trials, each with a different seed value, the average PSNR was 37.94 dB and the average TPE was 0.013 (using a block size of 16×16). This helps to validate the results in Table 1, where the average (from 300 tests; 30 trials for each of the 10 attacks) PSNR is 37.38 dB and the average TPE is 0.014 (again using blocks of size 16×16).

On average, it can be seen that the Dugad scheme is surviving all the attacks. However, for the JPEG quality 5 and 10 attacks, the watermark was not always detected (*i.e.*, a false negative reading was recorded). It can therefore be said that the Dugad scheme is reliable only for JPEG attacks of quality 15 and higher.

5.2 Our results

The same attacks used to test the Dugad algorithm (presented in Section 5.1) were used to test the new algorithm. *T1* = 115, *T2* = 200, *X1* = 20 and *X2* = 10 were the parametric values used; Figure 2 shows this watermarked image and the effect of attacking this watermarked image with various attacks. Table 2 presents the quantitative results for these various attacks.

An analysis of the probability of obtaining a false positive detector response is studied in Table 3. In this study, the lowest recorded normalised correlation (NC) value and the lowest recorded recovered watermark length from the 30 trials were saved. Using these values, it is possible to calculate the probability of obtaining a false positive reading [9]. This is a *worst case* scenario of obtaining a false positive detection as the lowest NC value and the smallest recovered watermark length are being used in the calculation (even though these two values did not occur simultaneously in any of the trials for any of the attacks).

The detector threshold value of 0.4 (in Table 2) was selected to determine the presence or non-presence of a watermark. On average, all attacks, apart from the JPEG quality 5 attack and the median 5×5 attack, are being survived. These results are replicated in Table 3 where it can be seen from the minimum NC values (for 30 trials) that a detector threshold value of 0.4 results in the new scheme being robust to all but the JPEG quality 5 attack and the median 5×5 attack.

From Table 3, it can also be seen that the chance of obtaining a false positive reading after suffering one of these attacks is extremely remote. However, the cropping attack poses a problem in that, on average, only 72 out of a possible 188 watermark bits were used by the detector, thus decreasing the reliability of the scheme. This lower number of recovered watermark bits leads to a greater chance of a false positive reading than the other survived attacks (see Table 3).

The scheme is not robust to JPEG quality 5 attacks (just like the Dugad method) nor median 5×5 attacks (unlike the Dugad method). However, both of these attacks degrade the quality of the watermarked image to a very severe extent. In order to survive the median 5×5 attack, it was found that setting *T1* = 110, *T2* = 210, *X1* = 10 and *X2* = 5 provided the necessary robustness. From 30 trials, the resultant watermarked image had an average PSNR value of 40.79 dB and an average TPE measurement of 0.006.

Thus, while surviving the same attacks as the Dugad scheme, the new scheme does not degrade the watermarked image to the same extent. From Table 2, the average (from 300 tests) PSNR value is 43.08 dB and the average TPE is 0.005. These are much better than the quantitative image quality results reported for the Dugad scheme in Section 5.1. Also, the average PSNR (40.79 dB) and the average TPE (0.006) values recorded for the new scheme that can survive median 5×5 attacks are superior to the Dugad results reported in Section 5.1.

Although only the Lena image has been used as the host image in the tests performed here, the newly presented watermarking scheme has been implemented robustly in a wide range of host images.

	No attack	JPEG			Median filter		Noise addition		Cropp- ing	Half sizing
		Q5	Q10	Q15	3×3	5×5	Gaussian	Impulse		
Average NC	1.00	0.20	0.62	0.84	0.89	0.30	0.52	0.64	0.53	0.51
Detector threshold chosen	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
Average WM length in	188	188	188	188	188	188	188	188	188	188
Average WM length out	188	153	165	170	168	163	159	154	72	142
Average PSNR (dB)	43.04	43.01	43.32	43.07	43.13	43.14	43.15	43.07	43.01	42.85
Average TPE	0.006	0.005	0.005	0.005	0.005	0.005	0.005	0.006	0.005	0.005

Table 2. Results for the new scheme (with *T1* = 115, *T2* = 200, *X1* = 20 and *X2* = 10). Each test was run 30 times upon the Lena image with a different seed and the average results were then calculated.

	No attack	JPEG			Median filter		Noise addition		Cropp- ing	Half sizing
		Q5	Q10	Q15	3×3	5×5	Gaussian	Impulse		
Minimum NC	1.00	0.03	0.56	0.78	0.82	0.23	0.45	0.62	0.43	0.42
Minimum length of recovered WM	188	144	159	161	160	157	155	139	69	128
Worst case Pfp (using minimum NC and minimum length)	$< 1.00 \times 10^{-50}$	5.30×10^{-1}	3.58×10^{-13}	1.20×10^{-25}	3.41×10^{-28}	3.24×10^{-3}	1.40×10^{-8}	8.40×10^{-14}	3.18×10^{-4}	2.46×10^{-6}

Table 3. Probability of false positive detector response for the new scheme (with *T1* = 115, *T2* = 200, *X1* = 20 and *X2* = 10). Each test was run 30 times upon the Lena image with a different seed.

6 Conclusion

A novel watermarking scheme using the Dugad method of determining the positions of marked coefficients (via an image sized/subband sized watermark) in collaboration with adapted versions of the Inoue insertion and detection techniques (using the notion of noise margins) has been presented. The new method is superior to the Dugad method in that it can survive the same attacks whilst producing marked images of higher visual quality (measured via the PSNR and Watson Metric quantitative techniques). Although the robustness of this new scheme is not quite as strong as that presented by Inoue, this can be attributed to its blind nature compared to the semi-blind nature of the Inoue method.

References

[1] R. Dugad, K. Ratakonda and N. Ahuja, A new wavelet-based scheme for watermarking images, *Proc. IEEE Intl. Conf. on Image Processing, ICIP'98*, Chicago, IL, USA, Oct. 1998, 419-423.

[2] H. Inoue, A. Miyazaki, A. Yamamoto and T. Katsura, A digital watermarking technique based on the wavelet transform and its robustness on image compression and transformation, *IEICE Trans., Special Section on Cryptography and Information Security*, E82-A, No. 1, Jan. 1999, 2-10.

[3] I. J. Cox, F. T. Leighton and T. Shamoon, Secure spread spectrum watermarking for multimedia, *IEEE Trans. on Image Processing*, Vol. 6, Dec. 1997, 1673-1678.

[4] M. Corvi and G. Nicchiotti, Wavelet-based image watermarking for copyright protection, *Scandinavian Conference on Image Analysis, SCIA '97*, Lappeenranta, Finland, June 1997, 157-163.

[5] P. Meerwald, Digital image watermarking in the wavelet transform domain, *Master's thesis*, Department of Scientific Computing, University of Salzburg, Austria, 2001.
http://www.cosy.sbg.ac.at/~pmeerw/Watermarking/

[6] A. B. Watson, DCT quantization matrices visually optimized for individual images, *Human Vision, Visual Processing and Digital Display IV, Proc. SPIE*, Vol. 1913, San Jose, CA, USA, Feb. 1993, 202-216.

[7] A. Mayache, T. Eude and H. Cherefi, A comparison of image quality models and metrics based on human visual sensitivity, *Proc. IEEE Intl. Conf. on Image Processing, ICIP'98*, Chicago, IL, USA, Oct. 1998, 409-413.

[8] S. Pereira, S. Voloshynovskiy, M. Madueo, S. Marchand-Maillet and T. Pun, Second generation benchmarking and application oriented evaluation, *Information Hiding Workshop*, Pittsburgh, PA, USA, April 2001, 340-353.

[9] D. Kundur and D. Hatzinakos, Digital watermarking using multiresolution wavelet decomposition, *IEEE ICASSP'98*, Volume 5, Seattle, WA, USA, May 1998, 2659-2662.

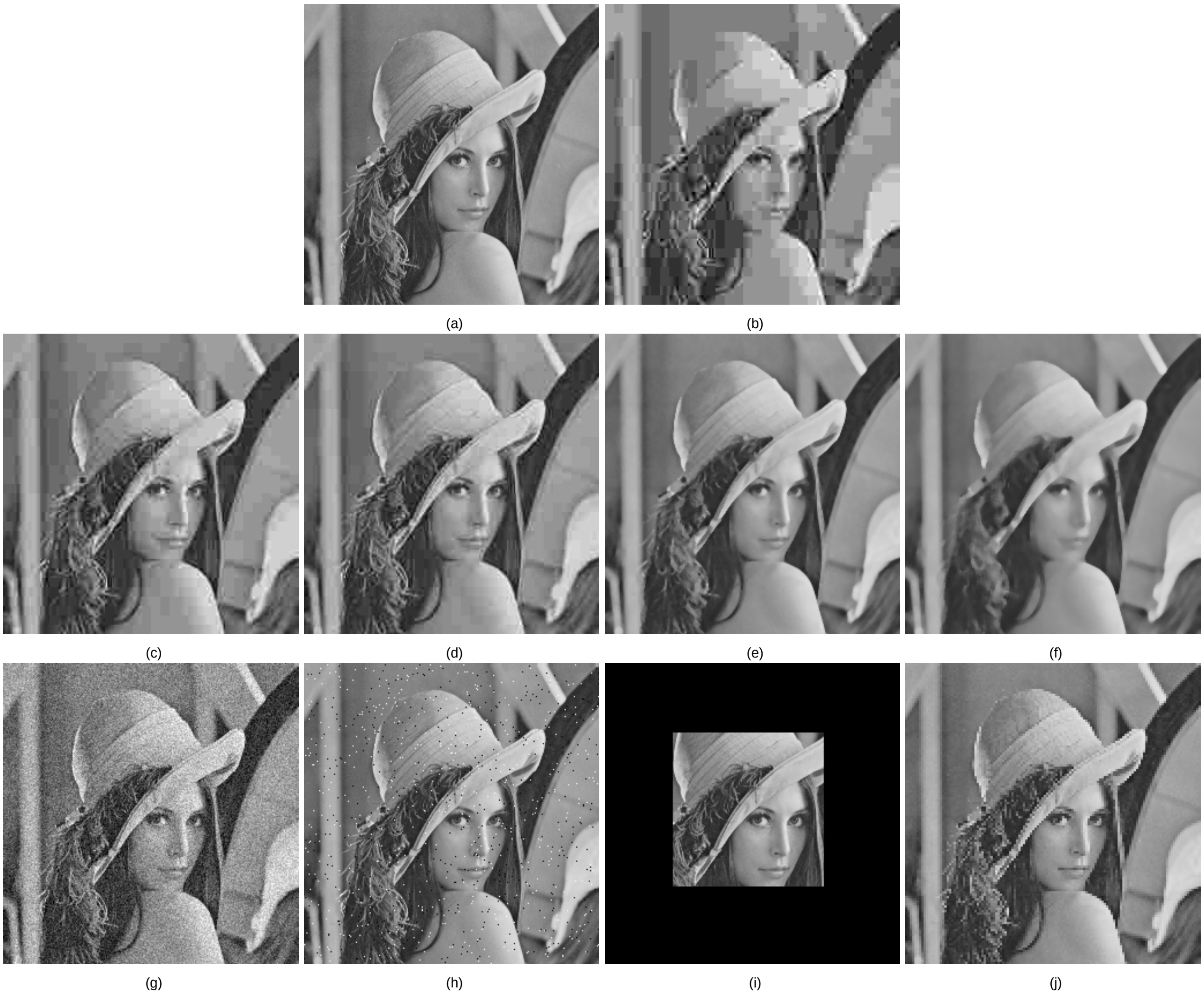

Figure 2. (a) Lena image marked via our watermarking scheme with $T1 = 115$, $T2 = 200$, $X1 = 20$ and $X2 = 10$ and attacked with: (b) JPEG quality 5, (c) JPEG quality 10, (d) JPEG quality 15, (e) median 3×3, (f) median 5×5, (g) Gaussian noise ($\sigma^2 = 375$), (h) impulse noise (normalised density of 0.015), (i) cropping and (j) half sizing (followed by resizing back to the original size).