Cover Page

1) Title of the paper:

Application of Grubbs' Test for Outliers to the Detection of Watermarks

2) authors' affiliation and address:

IRCCyN-IVC, (UMR CNRS 6597), Polytech' Nantes
Rue Christian Pauc, La Chantrerie, 44306 NANTES, France.
Tel : 02.40.68.30.52
Fax : 02.40.68.32.32

3) e_mail address:

Florent.Autrusseau@univ-nantes.fr

4) Conference & Publisher information:

ACM Workshop on Information Hiding & Multimedia and Security
http://www.ihmmsec.org/index.php/ihmmsec2014

5) bibtex entry:

```
@inproceedings{IHMmsec2014,
  author = {Urvoy, M. and Autrusseau, F.},
  title = {Application of Grubbs' test for outliers do the
detection of watermarks},
  booktitle = {2nd ACM Workshop on Information Hiding and
Multimedia Security},
  year = {2014},
  pages = {49-60}
}
```

Application of Grubbs' Test for Outliers to the Detection of Watermarks

Matthieu Urvoy
LUNAM Université, Université de Nantes
IRCCyN UMR CNRS 6597, Polytech Nantes
Rue Christian Pauc BP 50609 44306, Nantes, France
matthieu.urvoy@univ-nantes.fr

Florent Autrusseau
LUNAM Université, Université de Nantes
IRCCyN UMR CNRS 6597, Polytech Nantes
Rue Christian Pauc BP 50609 44306, Nantes, France
florent.autrusseau@univ-nantes.fr

ABSTRACT

In an era when the protection of intellectual property rights becomes more and more important, providing robust and efficient watermarking techniques is crucial, both in terms of embedding and detection. In this paper, the authors specifically focus on the latter stage. Most often, the detection consists in the comparison of a fixed and non-adaptive decision threshold to a correlation coefficient. This threshold is usually determined either theoretically or experimentally. Here, it is proposed to apply Grubbs' test, a simple statistical test for outliers, on the correlation data in order to take a binary decision about the presence or the absence of the searched watermark. The proposed technique is applied to three algorithms from the literature: the correlation data generated by the detector is fed to Grubbs' test. The obtained results show that Grubbs' test is efficient, robust and reliable. Above all, it automatically adapts to the searched watermark and can be easily applied to most types of watermarking approaches.

Categories and Subject Descriptors

I.4.9 [**Computing Methodologies**]: Image Processing and Computer Vision—*Applications*

General Terms

Watermarking, detection, Grubbs test

Keywords

Watermarking, detection, Grubbs test

1. INTRODUCTION

During the past two decades, watermarking has proven to be a practical and efficient solution to secure all kinds of digital documents. Typically, a watermarking scheme features two separate stages. On the one hand, the embedding step inserts an information, called watermark, within a host signal, in such a way that it cannot be perceived (*e.g.* invisible in images and video contents). On the other hand, the detection step assesses the presence or the absence of the watermark within the processed content. For optimal performances, it is crucial that both the embedding and the detection steps are jointly designed. Eventually, the detection results are binarized into a positive/negative output; the performances of this decision mechanism critically affect the efficiency of the overall watermarking chain.

In this study, a new decision mechanism is proposed. It can only be applied to correlation-based techniques, that is – as pointed out in [1] – most techniques: "*The example system we have used in our investigations, however fall into the class of correlation-based watermarking systems. Although this class does not include all possible watermarking systems, it does include the majority of example systems presented in this book, and, we believe, the majority of systems proposed in the literature.*"

Three kinds of correlation techniques can be encountered: the Linear Correlation (LC), the Normalized Correlation (NC) and the Correlation Coefficient (CC). Typically, the resulting values are compared to a decision threshold whose determination is often delicate and complex, but critical when it comes to ensure best performances, that is lowest False Positive (FP) and False Negative False Negative (FN) rates. Three scenarios (or hypotheses) are typically assessed to best determine the decision threshold. The detection is performed either (1) on a set of un-watermarked images, or (2) on a set of watermarked images, or (3) on a set of watermarked images but with mismatching watermarks. In this paper, notations $\mathcal{H}0$, $\mathcal{H}1$ and $\mathcal{H}2$ will respectively denote the first, second and third scenarios.

Numerous methodologies, of varying complexities, have been employed to determine the detection threshold. A common and simple practice is based on the "detector response", a diagram showing the correlation value obtained in a large number of contents, of which a single one has been watermarked. The resulting plot is used to empirically derive a threshold value. In [2], the detection threshold is set to the highest correlation value (0.23) obtained in a set of un-watermarked images. A similar approach is proposed in [3], where a large number of images (1000) are assessed under $\mathcal{H}0$ and $\mathcal{H}1$ hypotheses. The Probability Density Functions (PDFs) of the obtained correlation under $\mathcal{H}0$ and $\mathcal{H}1$ are then plotted. The threshold is set approximately halfway between $\mathcal{H}0$ and $\mathcal{H}1$ distributions, that is 0.17 in [3].

IH & MMSEC'14, June 11–13, 2014, Salzburg, Austria.

http://dx.doi.org/10.1145/2600918.2600931.

More rigorous approaches often consider the statistical properties of the PDF of the correlation under $\mathcal{H}0$, $\mathcal{H}1$ and $\mathcal{H}2$ hypotheses. These distributions are typically assumed to be Gaussian. The obtained statistical models may then be used to determine the threshold that maximizes the True Positive (TP) rate while limiting the False Positive (FP) rate to the desired value [4, 5, 6]. This is known as the Neyman-Pearson criterion.

No matter how complex are these methodologies, they all provide a fixed detection threshold. Yet, numerous factors may influence the correlation, including the host signal as well as the embedded watermark (dimensions, statistical properties, etc). A given threshold might be optimal for a given set of host signals and watermarks but may not be suited to other contents. Moreover, the recurrent assumption that processed signals are normally distributed may not hold in practice, thus inducing further detection mismatches. Besides, any variation in the embedding method (domain, equation) requires for the detection threshold to be re-estimated; in [7], for instance, a detection threshold is used for additive embedding, while another threshold is used for multiplicative embedding. When equations for the threshold are proposed, they only hold within the context of their study, that is for a particular watermarking scheme. In [4], for instance, the proposed equation (their Eq. 15) only applies to additive embedding within the Discrete Wavelet Transform (DWT) domain. In [6], a similar equation is available for multiplicative embedding in the Discrete Cosine Transform (DCT) domain. As an alternative, an adaptive detector was proposed in [8]; the detector, for each correlation peak, computes the probability that it does not belong to scenario $\mathcal{H}0$, based on a folded-normal distribution whose parameters may be estimated at detection.

In this study, a fast, universal and efficient decision method used to assess the presence or the absence of a watermark is proposed. It is based on Grubbs' test for outliers, a test that searches for eventual outlying observations in the cross-correlation data. Therefore, there is no need for a detection threshold anymore. Provided that the detection is correlation-based, the proposed method can be applied to all embedding domains and equations (additive, multiplicative, substitutive, etc), whatever the correlation is (linear or normalized, 1D or 2D). Methods computing a CC can also be easily adapted to the proposed technique. This stands in contrast with classical threshold estimation techniques, such as in [4], where derived equations solely apply to the proposed watermarking scheme. Moreover, in most previous works, several assumptions on the statistics of the host signal and the watermark coefficients are necessary. Instead, the proposed technique solely requires for the cross-correlation values to be approximately normally distributed.

In this study, the proposed method is adapted to three algorithms of the literature [9, 10, 3] that operate in different embedding domains and feature various embedding equations. This paper is organized as follows: section 2 first describes the proposed decision method; section 3 then briefly reviews the watermarking techniques in which the proposed method was evaluated. Section 4 then presents experimental results and thoroughly assesses the performances of the proposed method. Finally, section 5 concludes this work.

2. PEAK DETECTION: GRUBBS' TEST FOR OUTLIERS

According to [11], "*an outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism*". When viewing cross-correlation plots such as the ones from Fig. 1, it makes sense to consider the detection peak as an outlying value with respect to the surrounding correlation noise. For this reason, it is proposed here to apply outlier detection theory to the problem of watermark detection decision.

The literature provides a large range of outlier detection techniques [12], whether they rely on data indicators (*e.g.* samples depth, deviation, distance, density, etc) or make assumptions on the statistical properties of the input data and perform statistical tests. In this study, it is proposed to use Grubbs' test [13] as it is robust, reliable and computationally inexpensive. In some of our earlier experiments, the Extreme Studentized Deviate (ESD) test [14] showed equivalent performances to those of Grubbs'. However, alternatives could be considered, such as Dixon's test [15] or Tietjen-Moore's test [16]. In this paper, we only focus on Grubbs' test.

Typically, Grubbs' test is used to detect the presence of a single outlier in a univariate dataset whose distribution is approximately Gaussian. Both one-sided and two-sided variants have been proposed; here, the proposed watermark detection method is based on the latter. Grubbs' test two-sided statistic writes as

$$\mathcal{T}_G = \frac{\max|s(n) - \mu_s|}{\sigma_s} \tag{1}$$

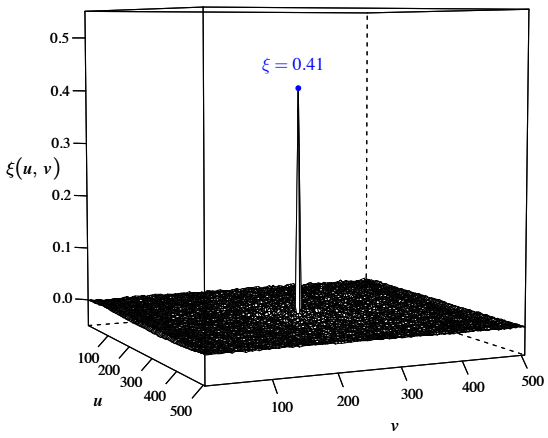
where $s(n)$ denotes the input samples, μ_s and σ_s respectively being the samples mean and standard deviation. The hypothesis that there is an outlier is accepted if $\mathcal{T}_G$ exceeds its critical value, that is

$$\mathcal{T}_G > \frac{N-1}{\sqrt{N}} \sqrt{\frac{(t_{\alpha_G/(2N),\,N-2})^2}{N-2+(t_{\alpha_G/(2N),\,N-2})^2}} \tag{2}$$

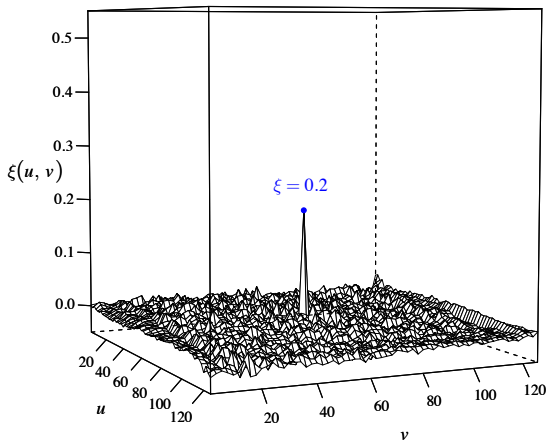
where N denotes the number of samples and $t_{\alpha_G/(2N),\,N-2}$ the critical value of Student's t-distribution with $N-2$ degrees of freedom and a significance level of $\alpha_G/(2N)$.

How does this apply to watermark detection? Let us now consider a typical still image watermark detection scheme: the searched watermark and the possibly marked signal are cross-correlated. The obtained cross-correlation (be it 1D or 2D) is input to Grubbs' test. Figure 1 shows an example of 2D cross-correlation data obtained with a modified version of the algorithm in [10] that was used to watermark the image Lena. Instead of using LC or Rao's detectors, as proposed in the original work, a normalized 2D cross-correlation [17] was computed between the watermarked sub-band and the watermark. Despite variations in correlation amplitudes due to two different sets of watermarking parameters (see section 3.3 for further details), the cross-correlation data features a central peak surrounded by a correlation noise of (much) lower amplitude. Provided that (1) the embedded watermark exhibits good auto-correlation properties and (2) the cross-correlation noise is (approximately) normally distributed, the detection is positive if Grubbs' test cannot reject the hypothesis that the central peak is an outlier.

Grubbs' test only takes two parameters as inputs, the correlation data and the significance level α_G. The latter cor-



(a) Sub-band (1,4) is watermarked, with a dwr of 12



(b) Sub-band (3,4) is watermarked, with a dwr of 20

Figure 1: Cross-correlation 2D data obtained with [10] for image Lena for two sets of embedding parameters.

responds to the maximum probability that the occurrence of the correlation peaks is due to chance alone, in other words the maximum probability that the detected correlation peaks are not outlying values. Therefore, α_G can be used to control the tradeoff between the True Positive (TP) and the FP detection rates: high (resp. low) values bring higher (resp. lower) TP and FP rates. Practical experiments will further detail the role of α_G in section 4. The proposed watermark detection decision technique features an iterated version of Grubbs' test which discards outliers one at a time until no more outliers are found.

3. APPLICATION TO EXISTING WATERMARKING SCHEMES

Three watermarking schemes [9, 10, 3] were selected from the literature in order to evaluate the performances of the proposed Grubbs-based decision technique. All three schemes exhibit various characteristics: different embedding domains and equations, various watermark dimensions and multiple detection algorithms. Table 1 summarizes the main features and characteristics of the three experimented algorithms. Section 3.1 briefly reviews the selected algorithms and Section 3.2 details the modifications which were made to incorporate Grubbs' test. Section 3.3 presents preliminary detection results, and Section 3.4 describes the process by which it was made sure that tested cross-correlation data is approximately Gaussian. More complete and detailed results are available in Section 4.

3.1 Description of experimented algorithms

Lin *et al.*, 2008 [9] – The watermark is embedded within the 3rd Low-High (LH3) sub-band of the DWT decomposition. The targeted sub-band is split into blocks of 7 consecutive coefficients. Each block is used to hold one bit of watermark information, by quantization of its two largest coefficients. At the detection, the difference between the two largest DWT coefficients of every block is compared to a threshold to retrieve the binary watermark information. The detection score is then given by the NC coefficient, which is eventually compared to a fixed detection threshold of 0.23 in order to ensure a FP probability of 1.03×10^{-7}. In a later publication [18], it was shown that this technique may present security issues; this is of no concern for the current study as it purely focuses on detection performances.

Kwitt *et al.*, 2009 [10] – Here, additive spread spectrum embedding is performed in the Dual Tree Complex Wavelet Transform (DT-CWT) domain. A perceptual mask is used to weight the watermark prior its embedding. Their study mostly focuses on various detection methods, and concludes that the Rao detector performs better than the simple LC.

Solachidis and Pitas, 2001 [3] – Multiplicative embedding is performed within the modulus of the Discrete Fourier Transform (DFT) domain, at mid-frequencies of the Fourier spectrum. The watermark is shaped as concentric rings, arranged into angular segments that each carry a watermark coefficient, for a total of 2300 coefficients arranged into 115 concentric rings and 20 sectors. The detection computes a CC between the searched watermark and the information retrieved in the Fourier magnitude.

3.2 Adaptation to Grubbs

As explained in section 2, Grubbs' test takes cross-correlation data as an input. However, the selected watermarking schemes do not feature such an information. Therefore, all three schemes were modified accordingly for Grubbs' test to be successfully incorporated.

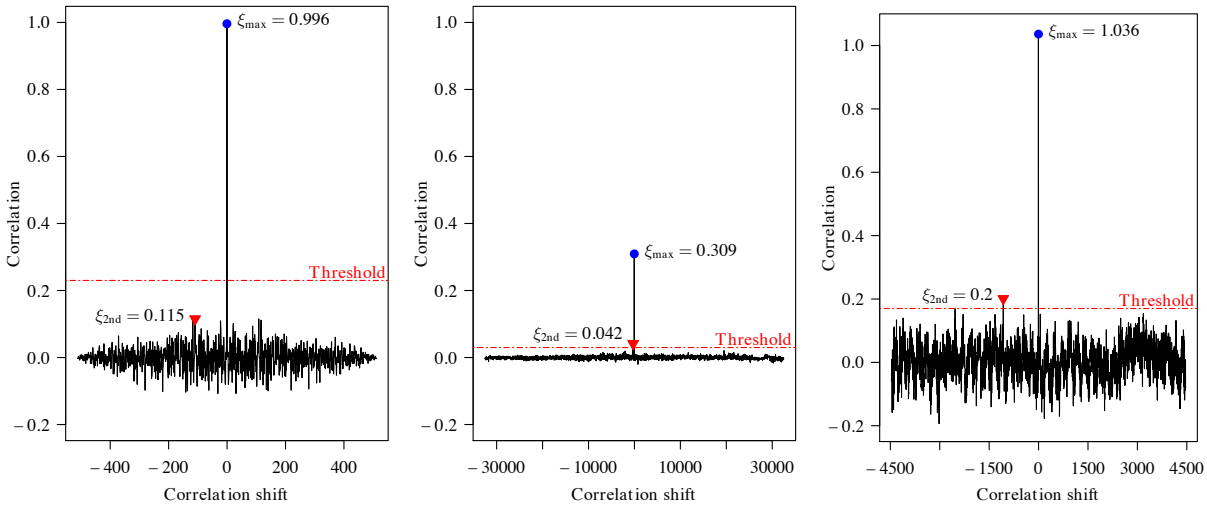
Lin *et al.*, 2008 [9] – Instead of computing the CC between the searched watermark and the retrieved information from the DWT domain, the 1D normalized CC is used to obtain a detection vector whose length is $2 \times 512 - 1$, that is 1023. Grubbs' test is then conducted on the resulting cross-correlation vector.

Kwitt *et al.*, 2009 [10] – Instead of Rao's test, the 2D normalized CC (from [17]) is computed between the extracted sub-band and the searched watermark (both being 128×128 matrices for a level 1 embedding process applied on a 256×256 image). Grubbs' test is then applied to the resulting 255×255 cross-correlation data.

Solachidis and Pitas, 2001 [3] – The hidden infor-

Table 1: Principal features and characteristics of experimented watermarking techniques.

	Embedding		Detection		Watermark
	Domain	*Method*	*Method*	*Correlation*	dimensions
[9]	DWT	Quantization	Blind	CC	1×512
[10]	DT-CWT	Additive	Blind	LC, Rao's test	128×128
[3]	DFT	Multiplicative	Blind	CC	20×115



(a) Case I: [9], max $\xi = 0.996$ (b) Case II: [10], max $\xi = 0.309$ (c) Case III: [3], max $\xi = 1.04$

Figure 2: Variability of the correlation matrices: distribution of the (vectorized) inter-correlation values obtained in three algorithms from the literature. The highest and second highest correlation peaks are respectively depicted by a round symbol (●) and a triangular symbol (▼).

mation is first retrieved by averaging, for each frequency-angular segment, the modulus of the corresponding Fourier coefficients. The 2D normalized CC (from [17]) is then performed between the retrieved matrix and the searched watermark, both of which featuring 20×115 coefficients. The resulting 39×229 CC data is then input to Grubbs' test.

3.3 Preliminary results and remarks

Before delving into large experiments and substantial results, it is proposed to first experiment all three modified watermarking schemes [3, 9, 10] on the Lena 512×512 image. The detection is applied to the watermarked image, without any attacks. Fig. 2 plots the (vectorized) cross-correlation data obtained; in each case, the highest ($\xi_{\max}$) and second highest ($\xi_{2\text{nd}}$) correlation peaks are respectively depicted by a round and a triangular symbol. From now on, the sliding position of the cross-correlation is called *correlation shift* (see x-axes in Fig. 2).

In all three algorithms, the detection peak occurs at the center of the cross-correlation data, that is when the searched watermark and the retrieved information fully overlap and therefore best match. This peak, and its amplitude $\xi_{\max}$, correspond to the correlation value of the original scheme; in [10], this corresponds to the output of the simple LC detector, not Rao's test whose output cannot be compared to a correlation value. In each plot of Fig. 2, the dashed line represent the detection threshold τ of the original algorithm. The detection is thus positive in all three techniques as the cross-correlation peak is higher than the threshold. Moreover, the amplitudes of the cross-correlation noise – in sliding positions other than the center – are mostly below the threshold, thus indicating that the searched watermark is only detected at its exact location.

As mentioned earlier, the length of the vectorized cross-correlation data depends on the experimented algorithm, respectively 1023, $255 \times 255 = 65025$ and $39 \times 229 = 8931$ in [9], [10] and [3]. As can be seen in Fig. 2, the amplitude of the cross-correlation peak significantly varies amongst experimented algorithms. Similarly, the average amplitude of the cross-correlation noise shows strong variations as well. For instance, the second highest cross-correlation peak in [3] (Fig. 2c) is equal to the cross-correlation peak in [10] (Fig. 1b). Another striking observation is that the second highest cross-correlation peak in [3] is higher than the advised detection threshold, thus indicating that that the advocated threshold (0.17) may not be optimal.

Table 2 lists the values of the CC's highest and second highest peaks in all three experimented algorithms on image Lena, as well as correlation gaps between the peaks and the threshold. Here, it can be seen that the cross-correlation gap $\Delta^{\tau}_{2\text{nd}}$ between the original detection threshold and the second highest peak shows a large variability amongst algo-

rithms. The same observation can be made on the gap Δ_{2nd}^{max} between the highest and second highest detection peaks.

Table 2: Variability of the correlation peaks obtained on image Lena.

Alg.	τ[a]	Peaks[b]		Corr. gaps[c]	
		ξ_{max}	ξ_{2nd}	Δ_{2nd}^{τ}	Δ_{2nd}^{max}
[3]	0.170	1.036	0.201	-0.031	0.835
[9]	0.230	0.996	0.115	0.766	0.881
[10]	0.029	0.309	0.042	-0.013	0.267

[a] τ is the detection threshold.
[b] ξ_{max} and ξ_{2nd} are respectively the highest and second highest correlation peaks.
[c] Δ_{2nd}^{τ} is the correlation gap between τ and ξ_{2nd}, and Δ_{2nd}^{max} is the gap between ξ_{max} and ξ_{2nd}.

Besides, even the same embedding algorithm can produce very different cross-correlation data with varying parameters. Such a scenario is illustrated in Fig. 1 for the modified algorithm [10] – see previous Section 2 –. In Fig. 1a, the DT-CWT sub-band (1,4) was watermarked with a dwr of 12, whereas in Fig. 1b the sub-band (3,4) was watermarked with a dwr of 20. The cross-correlation noise (that is the correlation values located around the central peak) exhibit a standard deviation of 1.66×10^{-3} in the first case, and 1.11×10^{-2} in the second case, that is a variation close to one order of magnitude.

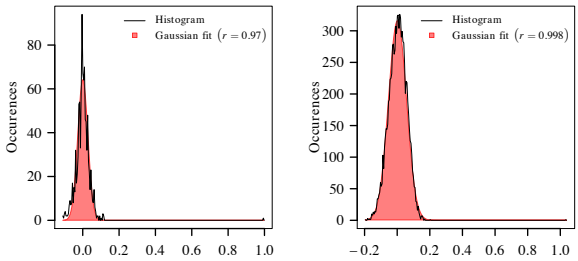
Naturally, these variations in cross-correlation are to be expected with such different embedding techniques and parameters. This emphasizes the fact that the detection threshold must be derived for each embedding technique and parameter set. Moreover, the methodology used to derive the threshold also needs to be adapted to the considered embedding scheme, which makes the matter even more difficult.

When applying Grubbs' test to the cross-correlation data shown in Fig. 2 ($\alpha_G = 10^{-6}$), only the highest peak is positively detected as an outlier in each case. Such a detection performance is especially valuable when considering the large differences between the experimented watermarking schemes and the resulting cross-correlation data. Therefore, these preliminary tests tend to show that Grubbs' test provides a universal method for taking the decision whether a watermark is present or absent.

3.4 Gaussianity of the cross-correlation data

Most statistical methods for estimating the detection threshold make several assumptions on the statistical properties of processed signals. As mentioned earlier, Grubbs' test stands in contrast to such approaches as it solely requires for the cross-correlation data to be normally distributed, even approximately [13, 19]. According to [20], the sample size should be greater than 6, another requirement that is obviously met in the studied context where the cross-correlation data is likely to hold thousands of coefficients.

In the preliminary experiments of previous section, Grubbs' test successfully detected the cross-correlation peaks as outlying values. However, it was not ensured whether the cross-correlation data could be reasonably approximated to a Gaussian distribution. It was thus proposed to bin the cross-correlation data into histograms that were fitted to a Gaussian distribution. Fig. 3 shows the obtained histograms for preliminary cases I (see Fig. 2a) and III (see Fig. 2c). Histograms are superimposed onto the fitted Gaussian curves.



(a) Case I: [9] (b) Case III: [3]

Figure 3: Application of [9] and [3] to image Lena: distribution of the cross-correlation data.

In [3] (Fig. 3b), the cross-correlation histogram fits very well a Gaussian distribution – the correlation (Pearson's r) between the histogram and the fitted curve reaches 0.998. In [9] it is best fitted by a Laplacian distribution, yet the Gaussian fit still provides a reasonable approximation. Table 3 provides goodness of fit for the estimated distributions in terms of Pearson's r and Root Mean Square Error (RMSE). Overall, the distribution of the obtained cross-correlation data are close enough to a Gaussian distribution and therefore can be input to Grubbs' test. It is important to note that the distributions of the cross-correlation data in Fig. 3 are not centered due to the outlying correlation peaks (respectively 0.996 and 1.036 in [9] and [3]).

Table 3: Goodness[a] of fit for cross-correlation data in [9] and [3] on image Lena

	Gaussian fit		Laplacian fit	
	Pearson's r	*RMSE*	*Pearson's r*	*RMSE*
[9]	0.970	0.203	**0.986**	**0.168**
[3]	**0.998**	**0.174**	0.979	0.633

[a] Bold values designate best fitting parameters.

Now focusing on preliminary case II (see Fig. 2b), the observation of the signals processed in algorithm in [10] revealed various kinds of distributions. Figure 4 shows the distribution of four successively processed signals: (1) the DT-CWT sub-band (Fig. 4a), (2) the watermark (Fig. 4b), (3) the watermarked sub-band (Fig. 4c) and (4) the cross-correlation data (Fig. 4d). Again, the distribution of the cross-correlation data in Fig. 4d is not centered due to the outlying correlation peak with amplitude 0.309. Each of these distributions were fitted to both Gaussian and Laplacian models; the resulting goodness of fit is listed in Table 4.

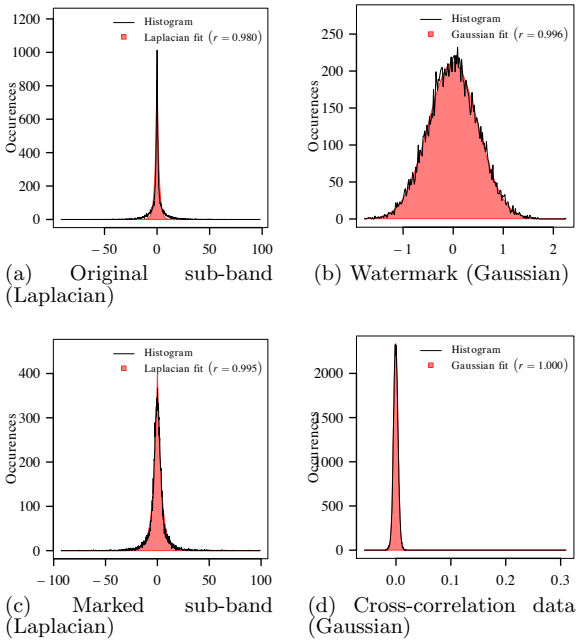
As can be seen, the distributions of both the original and watermarked sub-bands are rather Laplacian, whereas the watermark is normally distributed. In the end, it appears that the cross-correlation data still follows a Gaussian distribution (Pearson's r of 0.997), despite the Laplacian behavior of the sub-band values. Here, typical estimation techniques for the detection threshold – that assume Gaussian distributions of both the host signal and the watermark – may

Table 4: Gaussianity of the signals processed in [10]: goodness[a] of fit for Gaussian and Laplacian model.

	Gaussian fit		Laplacian fit	
	r^{b}	*RMSE*	r^{b}	*RMSE*
Subband	0.934	0.0030	**0.980**	**0.0018**
Watermark	**0.996**	**0.0146**	0.972	0.0541
Marked subb.	0.990	0.0012	**0.995**	**0.0007**
Cross-corr.	**1.000**	**0.1015**	0.987	0.5463

[a] Bold values designate best fitting parameters.
[b] r denotes Pearson's correlation coefficient.

provide erroneous values. On the contrary, the requirements for Grubbs' are all met, therefore ensuring that the detected outlying values indeed correspond to cross-correlation peaks for which the detection is positive.



(a) Original sub-band (Laplacian) (b) Watermark (Gaussian)

(c) Marked sub-band (Laplacian) (d) Cross-correlation data (Gaussian)

Figure 4: Application of [10] to image Lena: distribution of the host signal, the watermark, the marked signal and the cross-correlation.

4. RESULTS

Preliminary results tend to show that Grubbs' test is capable of locating watermark detection peaks in different scenarios. Large scale experiments are now conducted in order to systematically assess Grubbs' test ability to detect the presence or the absence of a watermark from cross-correlation data.

4.1 Experimental apparatus and parameters

Hereon, two main datasets are used for the experiments: dataset $\mathcal{D}_a$ contains 1000 gray level images from the BOWS-2 contest[1]; dataset $\mathcal{D}_b$ contains 10 color images from the Kodak Database[2]. The second dataset $\mathcal{D}_b$ serves in scenarios that involve a large number of detections, so as to keep the experimental scale within reasonable bounds. The datasets are listed and detailed in Table 5.

Three main experimental hypotheses are considered. One of them ($\mathcal{H}1$) should ideally lead to positive detections, contrary to the other two ($\mathcal{H}0$ and $\mathcal{H}2$) in which detection should conclude in the absence of the searched watermark.

- In scenario $\mathcal{H}0$, original – therefore un-watermarked – images from dataset $\mathcal{D}_a$ are input to detection.
- In scenario $\mathcal{H}1$, watermarked images from dataset $\mathcal{D}_a$ are first attacked and then input to detection, which searches for the initially embedded watermark.
- In scenario $\mathcal{H}2$, watermarked images from dataset $\mathcal{D}_b$ are input to detection, which searches for 1000 erroneous watermarks.

When necessary, that is in scenario $\mathcal{H}1$, four types of attacks are simulated: (1) attack *Blur 3* applies Gaussian blur with standard deviation $\sigma = 3$; (2) attack *Pois. 120* inserts Poisson noise with parameter $\lambda = 120$; attack *Pois. 140* inserts Poisson noise with parameter $\lambda = 140$; (4) attack *Med. 5* performs 5×5 median filtering.

Table 5: Datasets used in the experiments

Name	Description	Type	Size	Resolution
$\mathcal{D}_a$	BOWS-2	Gray level	1000	512×512
$\mathcal{D}_b$	Kodak	Color	10	512×768

4.2 A first glance at cross-correlation data

The whole concept of the present study relies on the assumption that cross-correlation data should behave differently depending on the presence or the absence of a watermark. To be more precise, it is expected in scenario $\mathcal{H}1$ that a visible peak emerges at the center of the cross-correlation data, whereas this should not happen in scenario $\mathcal{H}0$. In some cases, however, and notably when geometric distortions are undergone, the correlation peak may move away from the center (provided that the watermarking algorithm is resilient against geometric distortions). Alternatively, the correlation data may contain several peaks (either by design or due to geometric distortions); the iterative implementation of Grubbs' test (see Section 2) is expected to detect them one at a time. To verify this, the modified algorithm [10] was evaluated against the dataset $\mathcal{D}a$ under both $\mathcal{H}0$ and $\mathcal{H}1$ hypotheses.

In each scenario, the 2D cross-correlation data was collected and averaged over the 1000 experimented images. The resulting averages are plotted in Fig. 5. The obtained results show that, over a large number of images, there is indeed a clearly visible cross-correlation peak for the $\mathcal{H}1$ scenario (Fig. 5a). This stands in contrast with scenario $\mathcal{H}0$ (Fig. 5b) in which the central peak of the cross-correlation barely emerges from the surrounding cross-correlation noise.

[1] http://bows2.ec-lille.fr
[2] http://r0k.us/graphics/kodak/

The averaged cross-correlation data of Fig. 5 were input to Grubbs' test for outlier, with $\alpha_G = 10^{-6}$. In Fig. 5a, five outliers are detected, including the central peak and four others in its vicinity. In Fig. 5b, no outliers are detected at all. Similar results were obtained in [9] and [3]. Therefore, this study's underlying assumption holds in practice.

4.3 Gaussianity of the cross-correlation over experimental datasets

Similarly to what was done in section 3.4, it is proposed here to ensure the normality of the collected cross-correlation data over the entire datasets $\mathcal{D}a$ and $\mathcal{D}b$. The modified algorithms [9] and [3] were evaluated under hypotheses $\mathcal{H}0$ (dataset $\mathcal{D}a$), $\mathcal{H}1$ (dataset $\mathcal{D}a$) and $\mathcal{H}2$ (dataset $\mathcal{D}b$) – see section 4.1 for their definitions –. This resulted in 1000 cross-correlation data under hypothesis $\mathcal{H}0$, 4000 under hypothesis $\mathcal{H}1$ and 10000 under hypothesis $\mathcal{H}2$.

The collected cross-correlation data were then binned into histograms, which in turn were fitted to a Gaussian distribution. The goodness of fit was evaluted in terms of Pearson's r and RMSE between the experimental histogram and the fitted curve. For each algorithm and hypothesis, Table 6 lists the average, minimum and maximum figures for both indicators.

Table 6: Gaussian fits goodness of fit

	Pearson's r			RMSE		
	Mean	Min	Max	Mean	Min	Max
	Solachidis *et al.* [3]					
$\mathcal{H}0$	0.9977	0.9555	1.0000	0.0296	0.0043	0.1560
$\mathcal{H}1$	0.9952	0.9367	0.9999	0.0387	0.0068	0.1609
$\mathcal{H}2$	0.9988	0.9650	0.9999	0.0236	0.0072	0.1456
	Lin *et al.* [9]					
$\mathcal{H}0$	0.9883	0.9631	0.9987	0.0856	0.0270	0.1438
$\mathcal{H}1$	0.9897	0.9588	0.9998	0.0814	0.0109	0.1551
$\mathcal{H}2$	0.9916	0.9189	0.9998	0.0720	0.0127	0.2094

In [3], the cross-correlation data fits very well a Gaussian distribution; Pearson's r never falls below 0.9555 and averages at 0.9977. In [9], the goodness of fit is reasonably high as well, high enough to justify the use of Grubbs' test; Pearson's r reaches 0.9189 at its minimum. This can be explained by the fact that, as was already seen in Fig. 3 and Table 3, the cross-correlation data in [9] is best approximated by a Laplacian distribution. Still, such a distribution is well handled by Grubbs' test, provided that its standard deviation is high enough not to present heavy tails – which is the case here.

4.4 True Positive and False Positive rates

Properly setting the value of Grubbs' significance level α_G is crucial as it drives detection performances. High values of α_G (*e.g.* 10^{-2}) are very likely to detect outlying values, but on the other hand are also likely to detect non-outlying values with respect to the considered context. Inversely, on the same input data, low values of α_G (*e.g.* 10^{-8}) ensure not to detect these false outliers, but are also likely to miss truly outlying values. Transposed to watermarking, this means that both TP and FP rates are likely to increase with α_G.

Grubbs' test detection output was collected in modified algorithms [9] and [3] for all three scenarios $\mathcal{H}0$, $\mathcal{H}1$ and $\mathcal{H}2$, for varying α_G (10^{-8} to 10^{-2}). Under $\mathcal{H}0$, none of the 1000 original images of dataset $\mathcal{D}_a$ led to false detections. As for hypotheses $\mathcal{H}1$ (TP) and $\mathcal{H}2$ (FP), the obtained evolution in FP and TP rates as a function of α_G is plotted in Fig. 6. Note that the values for the plotted TP rates correspond to *Pois. 120* attacks only.

As expected, the experimental results show that both TP ($\mathcal{H}1$) and FP ($\mathcal{H}2$) rates increase with α_G. Still, in [3], the TP rate barely increases and remains rather stable over the entire experimented values of α_G. The two modified algorithms present rather opposite behavior. Algorithm [3], on the bright side, displays a high TP rate (around 95%), but on the downside a high FP rate as well (up to 50% for low values of α_G). As for [9], both its TP rates and FP rates are rather low; the first range from 30 to 50% while the second remains below 15%.

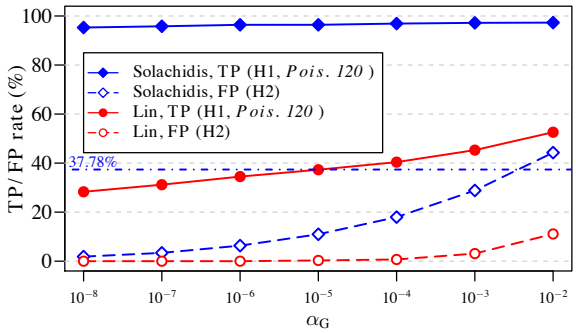
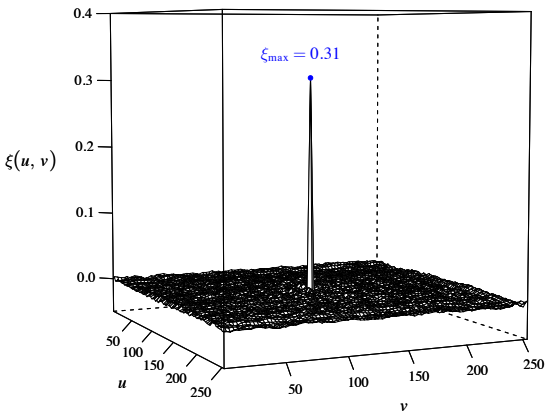


Figure 6: TP and FP rates against Grubbs' significance level α_G in [3] and [9]. TP rates correspond to $\mathcal{H}1$ scenarios featuring the attack *Pois. 120*; FP rates correspond to scenarios $\mathcal{H}2$.

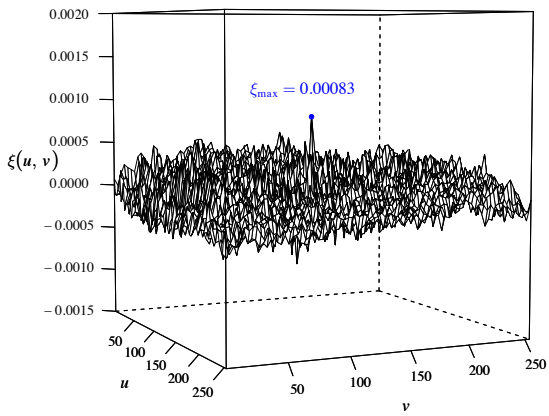
For comparison purposes, we also experimented the original algorithm [3] under hypothesis $\mathcal{H}2$. Interestingly, it features an FP rate of 37.78% (see dotted-dashed line in Fig. 6), meaning that Grubbs' test outperforms the original detection algorithm, in terms of FPs, for significance levels below $\alpha_G = 10^{-3}$. In addition, we will see later that the method in [3] presents abnormally high detection rates. Still, the main objective of this study is to investigate whether the proposed detection decision mechanism is competitive, that is how much better or worse it performs in comparison to the original detection schemes.

4.5 Comparison of original and proposed detection schemes

As was just pointed out, the main objective of our study is to compare the proposed Grubbs-based detection scheme to the initial threshold-based detection scheme. To this end, the modified algorithms [9] and [3] were evaluated under hypothesis $\mathcal{H}1$, a scenario which features four types of attacks (see Section 4.1 for full details). Based on TP and FP experimental rates obtained in previous section, Grubbs' significance level α_G will be set to 10^{-6}, which provides a proper tradeoff between the ability to retrieve a watermark and the ability to distinguish between different watermarks. Corresponding values for TP and FP rates are listed in Table 7.



(a) Average cross-correlation for [10] on 1000 images (from $\mathcal{D}a$) - True positives



(b) Average cross-correlation for [10] on 1000 images (from $\mathcal{D}a$) - False Positives

Figure 5: Average 2D cross-correlation data obtained with [10] against 1000 marked images (a), and against original images (b).

Table 7: FP and TP rates: operating values for a Grubbs significance level of $\alpha_G = 10^{-6}$ under $\mathcal{H}1$ hypothesis and attack *Pois. 120*

	TP rate	FP rate
[9]	96.4%	6.34%
[3]	34.5%	0.0%

For both the original and the modified algorithms, and for each type of attack, the number of watermarked images in which the detection is positive was computed. These values are denoted TP_{Orig} and TP_{Grubbs}, and correspond to the TP rates respectively obtained in the original and the modified algorithm. In addition, we counted the number of images in which the modified detection scheme retrieves the watermark (TP), while the original scheme does not (FN); this number is denoted Δ_{TP}. We also counted the number of images in which the opposite scenario arises: the original scheme retrieves the watermark, but the modified scheme does not; this number is denoted Δ_{FN}. The obtained results are listed in Table 8.

As was observed previously in Fig. 6, results from Table 8 demonstrate weak robustness performances for [9] in comparison to [3]. Nevertheless, the use of Grubbs' test improves the detection performances in most cases, no matter how good is the robustness of the original method. In [3], the introduction of Gaussian blur (attack *Blur 3*) leads to strongly diverging results when using either the original or the modified detection scheme. This suspicious effect is further investigated in the next section.

Fig. 7 plots the detection gain (Δ_{TP}, positive bars) and loss (Δ_{FN}, negative bars) obtained in both experimented methods. The suspicious results obtained in [3] with attack *Blur 3* are omitted for readability purposes. Apart from the attack *Med. 5*, the proposed detection method performs better than the original detection methods and their fixed thresholds. Detection rates significantly improve, by up to 10% in [9] against the attack *Blur 3*.

Table 8: Robustness to attacks (%) for dataset $\mathcal{D}_a$: original and modified schemes.

	Blur 3	Pois. 120	Pois. 140	Med. 5
	Solachidis *et al.* [3]			
TP_{Orig}	78.8*	92.5	87.6	99.9
TP_{Grubbs}	7.5*	96.4	94.5	98.2
Δ_{TP}	0.0*	4.0	7.2	0.0
Δ_{FN}	71.3*	0.0	0.0	1.7
	Lin *et al.* [9]			
TP_{Orig}	20.5	25.1	12.0	95.2
TP_{Grubbs}	30.7	34.5	18.1	98.1
Δ_{TP}	10.2	9.4	6.1	2.9
Δ_{FN}	0.0	0.0	0.0	0.0

* These unexpected results are further discussed in section 4.6

4.6 Investigating suspicious results

Some results obtained with algorithm [3] in previous Section raised suspicions. In particular, under the attack *Blur 3*, it is striking to notice that the original detection scheme retrieves 78.8% of the watermarks, while the modified scheme only detects 7.5% of them. This is especially surprising in regards to other attacks or algorithms where such an effect never occurs. To further investigate this specific scenario, we plotted in Fig. 8 the distribution of the CC returned by the original detection scheme (which is equal to the amplitude of the central peak of the cross-correlation data) in four scenarios: (1) on original un-watermarked images ($\mathcal{H}0$); (2) on watermarked images in the absence of any attack ($\mathcal{H}1'$); (3) on watermarked images after *Blur 3* attack ($\mathcal{H}1_{B3}$); and (4) in a variant of scenario $\mathcal{H}2$ which also features the *Blur 3* attack ($\mathcal{H}2_{B3}$).

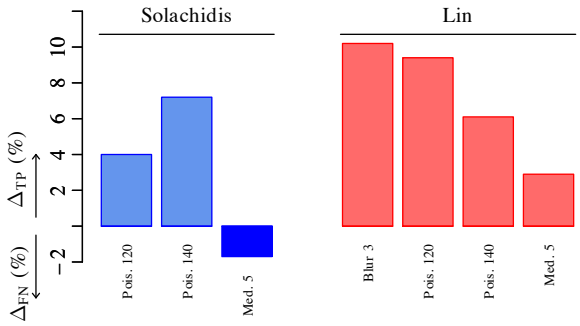


Figure 7: Contribution of Grubbs' test: Δ_{TP} denotes the percentage of images in which Grubbs' test is successfull contrary to the original detector; conversely, Δ_{FN} denotes the percentage of images in which Grubbs' test is unsuccessful contrary to the original detector.

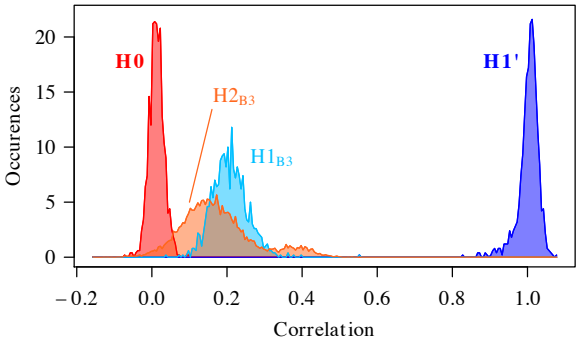


Figure 8: Distribution of the cross-correlation peaks: comparison between scenarios $\mathcal{H}0$, $\mathcal{H}1'$, $\mathcal{H}1_{B3}$ and $\mathcal{H}2_{B3}$.

In the absence of attacks ($\mathcal{H}0$ and $\mathcal{H}1'$), the distribution of the cross-correlation peaks are narrow and clearly disjoint, and the detection threshold (0.17) is located in between the two. The original detection scheme is thus correct for both of these scenarios. When the images are attacked ($\mathcal{H}1_{B3}$ and $\mathcal{H}2_{B3}$), these distributions are widening and shift towards the threshold, so that they strongly overlap. In this case, the advocated threshold (0.17) fails to distinguish between those images in which a watermark should be detected and those in which it should not. For this reason, the technique in [3] features a high FP rate. The use of 0.5 as a threshold may circumvent this issue, but then none of the watermarked and attacked images would be detected at all.

Practical experiments were conducted on dataset $\mathcal{D}_b$ under hypothesis $\mathcal{H}2_{B3}$. Results show that the original detection method in [3] presents a FP rate of 46.1%, while the proposed detection method never detects the wrong watermark. This tends to show that the suspicious results from Table 8 are more likely to be in favor of the proposed detection technique, which in this case takes the better decision to consider that the watermark cannot be retrieved.

4.7 Original versus modified algorithms: investigating the reasons for disagreement

With the exception of the suspicious results obtained under hypothesis $\mathcal{H}1_{B3}$, the other results – see Fig. 7 – show increased detection performances for most attacks when using the proposed detection method. One may wonder, however, under which circumstances exactly Grubbs' test disagrees with the original threshold-based detection method. Experimental results for scenarios $\mathcal{H}1$ and $\mathcal{H}2$ were further examined in two situations: (*i*) when Grubbs' test positively detects an outlier but the peak correlation is below the original detection threshold; (*ii*) when inversely, Grubbs' test does not detect any outlier but the peak correlation is above the original detection threshold.

In both situations (*i*) and (*ii*), the average, minimum and maximum amplitudes of the corresponding cross-correlation data were computed for each value of the correlation shift. The resulting data is plotted in Fig. 9. The dark blue curve represents the average value of the cross-correlation data (**A** in Fig. 9a). The average central peak is represented by a red round symbol (**D** in Fig. 9a). The amplitude range within which the central peak is comprised is delimited by two horizontal light red lines (**E** and **F** in Fig. 9a) that respectively cross the y-axis at the minimum and maximum values of the central peak. Finally, the light blue area delimited by the curves (**B** and **C** in Fig. 9a), represents the amplitude range that comprises the remaining cross-correlation data, which we call cross-correlation noise. In other words, **B** (resp. **C**) plots, in all shifting positions except the center, the minimum (resp. maximum) value amongst collected cross-correlation data. The same representation is used in Fig. 9b, 9c and 9d.

Two scenarios can be envisioned: either (1) the noise range (light blue area) and the peak range (light red area) do not overlap or (2) they do. In the first case, this would tend to show that a watermark can indeed be detected. In the second case, either the hypothesis that a watermark is present should be rejected, or the searched watermark presents bad auto-correlation properties, which is problematic and very likely to induce False Alarms.

Fig. 9a corresponds to 286 $\mathcal{H}1$ scenarios in which the modified algorithm [9] positively detects the watermark while the original algorithm does not. Out of the 1000 images of dataset $\mathcal{D}_a$, this occured 102 times under *Blur 3* attack, 29 times under *Med. 5* attack, 94 times under *Pois. 120* and 61 times under *Pois. 140*; in total, this sums up to 286 cases out of 4000. As can be seen in Fig. 9a, despite the fact that the amplitude of the cross-correlation peak is located below the original detection threshold, its values span a range that is clearly above and disjoint from the cross-correlation noise. The presence of a detection peak cannot be denied; therefore Grubbs' test takes the better decision to consider that a watermark is indeed found.

A similar scenario can be seen in Fig. 9b: the modified algorithm [3] performs better in 86 images out of 4000 (30 against *Pois. 120* attack, 56 against *Pois. 140* attack). Again, although it is less obvious than in Fig. 9a, the values of the cross-correlation peak and those of the cross-correlation noise are most of the time non-overlapping. The presence of a watermark is thus extremely likely, which is consistent with Grubbs' decision.

Conversely, the modified algorithm [3] does not detect any watermark in 772 images (742 against *Blur 3*, 2 against

Pois. 120, 2 against *Pois. 140* and 26 against *Med. 5*) while the original algorithm does. Fig. 9c represents the average, maximum and minimum values of the 770 corresponding cross-correlation data. Here, the dynamic of the cross-correlation noise is much larger relatively to the amplitude of the peak; for some correlation shifts, the maximum amplitude of the noise even exceeds the average amplitude of the peak. Moreover, the cross-correlation noise regularly exceeds the original detection threshold: cross-correlation peaks that should go undetected may therefore be detected which is likely to cause False Alarms. Again, this puts back into question the choice for the original detection threshold, as was already discussed in Section 4.6. In [9], the modified detection scheme never missed a watermark that was detected by the original detector.

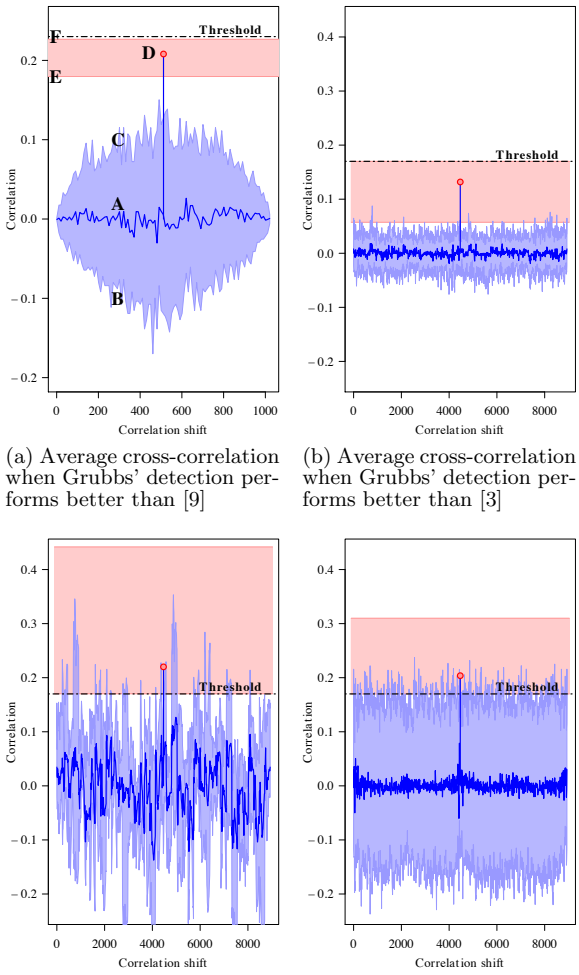
Finally, the modified algorithm [3] was experimented under hypothesis $\mathcal{H}2$. Out of the 10000 detections, the original detection is positive in 3778 cases, that is 37.78% of False Alarms. In contrast, the proposed detection scheme is positive in 634 cases only, that is a reduction of 31.44% in FP rate. Fig. 9d plots a summary of the 3151 images in which the original detection is positive (thus erroneous) and the modified detection is not. Although less blatant here, the maximum amplitude of the cross-correlation noise regularly exceeds the original detection threshold. With Grubbs' test, this limited overlap is significant enough to correctly reject the hypothesis that the correct watermark is present.

4.8 Receiver Operating Characteristics

So far, results suggest that Grubbs' test for outlier is suited to the detection of peaks within cross-correlation data. Moreover, the significance level can be used to tune the accuracy of the detector, that is to (jointly) control the TP and FP rates. Here, it is proposed to plot the Receiver Operating Characteristic (ROC) of the modified detection scheme, for various values of α_G, and to compare it to the ROC curve of the original algorithm [9].

The original detection CC was collected under two hypotheses: $\mathcal{H}0$ and $\mathcal{H}1$ (against *Pois. 120* attack). Their distribution, in either case, are not Gaussian. Rather, $\mathcal{H}0$'s distribution is best fitted to a skew-normal distribution with parameters $\Theta_{\mathcal{H}0} = \{\mu = -8.722 \times 10^{-3}, \sigma = 4.262 \times 10^{-2}, \xi = 1.054\}$, where ξ is the skewness parameter. As for $\mathcal{H}1$ (*Pois. 120*), it is best fitted by a generalized skew-hyperbolic distribution with parameters $\Theta_{\mathcal{H}1} = \{\mu = -0.174, \delta = 0.050, \alpha = 858.3, \beta = 844.4\}$, where β is the skewness parameter. Fig. 10 plots the distributions of the CC under both hypotheses (solid lines) as well as the corresponding fits (colored areas). Here, most detection methods – that often assume normal distribution of the CC – are likely not to estimate a proper detection threshold.

The obtained distributions for $\mathcal{H}0$ and $\mathcal{H}1$ were then used to plot the ROC curve in Fig. 11. For best understanding, the ROC curve is plotted twice: in Fig. 11a, it is plotted over a linear x-axis; in Fig. 11b, it is plotted in its entirety over a logarithmic x-axis. The solid green curve is drawn from experimental data (the solid curves from Fig. 10), while the dashed red curve is drawn from interpolated data (the fits from Fig. 10). The experimental operating point of the original detection scheme is schematized by a green diamond symbol; its interpolation from the fitted curves is schematized by a red square symbol (Fig. 11b). Their positions are matching in Fig. 11a.



(a) Average cross-correlation when Grubbs' detection performs better than [9]

(b) Average cross-correlation when Grubbs' detection performs better than [3]

(c) Average cross-correlation when Grubbs' detection performs worse than [3]

(d) Average cross-correlation for [3] when testing wrong watermarks

Figure 9: Average cross-correlation data in various scenarios. In (a) and (b) (respectively algorithms [9] and [3]), Grubbs' test outperforms the original algorithm. In (c), Grubbs' test fails to detect the watermark but the original algorithm [3] is successful. In (d), Grubbs' test avoids detecting wrong watermarks but the original algorithm [3] does not.

The ROC curve of the modified scheme is plotted in blue with round symbols. Its values are based on experimental results only: it is not possible to interpolate neither FP nor TP rates for other values of α_G with Grubbs' test. As seen earlier, both TP and FP rates in the modified detection scheme increase with α_G.

As can be seen from Fig. 11b, the ROC curve of the proposed method passes over the original curve for values of α_G below 10^{-6}, that is the value which we used in the experiments. The optimal value for α_G is thus 10^{-6} as it provides the best TP rate with a null FP rate. Moreover, the

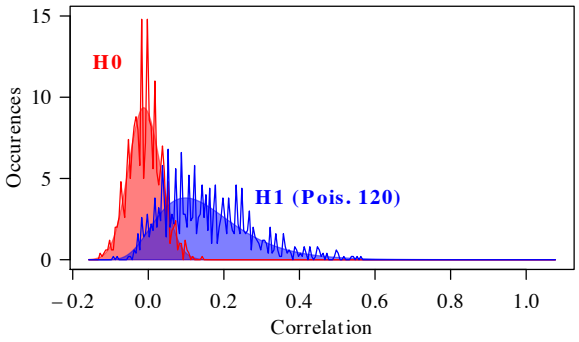
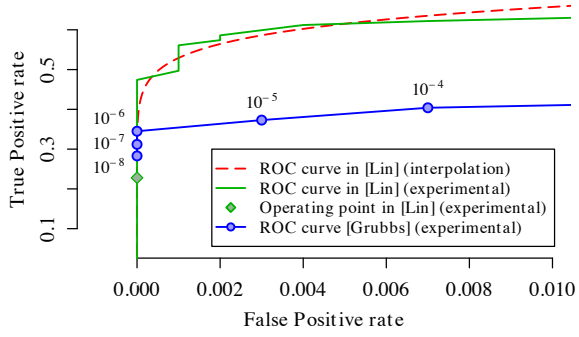
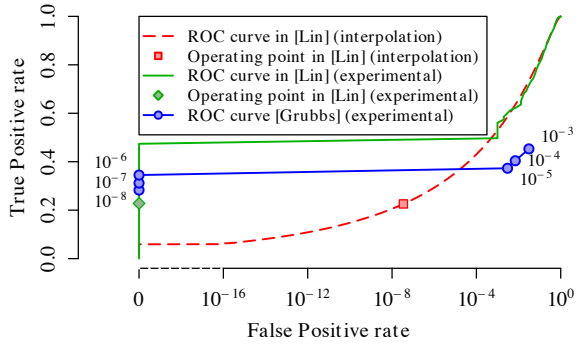


Figure 10: $\mathcal{H}0$ and $\mathcal{H}1$ (Poisson 120) distributions: histograms are superimposed on the corresponding skew-normal and generalized skew hyperbolic fits.

proposed method, α_G increases the TP rate by more than 11% (from 22.8% to 34.5%) in comparison to the original operating point.



(a) Zoomed ROC curve



(b) ROC curve (logarithmic x axis)

Figure 11: ROC curves computed on the initial algorithm in [9]. The experimental data are depicted by the solid green curves, the interpolated data by the dashed red curve, and the Grubbs detections are depicted by the blue symbols.

5. CONCLUSION

This work presents a prospective study on the possible use of outlier detection methods on correlation data for watermark detection without the need to select a threshold. It was shown that Grubbs' test (among other outlier detection methods) can be used to accurately determine whether the watermark is present within the tested content. Grubbs's test is virtually applicable on any correlation-based watermarking method, whatever the embedding domain, the embedding equation, etc. In this work, three algorithms from the literature [3, 10, 9] were adapted to incorporate Grubbs' test. The adaptation is straightforward and brings improvements in most cases. Specifically, our experiments showed that either the FP rate was reduced (in [3]), or the TP rate was increased (in [9]), without one affecting the other. Finally, we showed that varying the significance level in Grubbs' test allows fine tuning of both the False Positive rate and the True Positive rate.

6. REFERENCES

[1] Ingemar J. Cox, Matthew L. Miller, Jeffrey A. Bloom, Jessica Fridrich, and Ton Kalker. Digital Watermarking and Steganography. The Morgan Kaufmann Series in Multimedia Information and Systems, San Francisco, CA, USA, 2nd edition, 2007.

[2] Ante Poljicak, Midija Mandic, and Darko Agic. Discrete Fourier transform–based watermarking method with an optimal implementation radius. Journal of Electronic Imaging, 20(3):033008–1–8, July 2011.

[3] V Solachidis and I Pitas. Circularly symmetric watermark embedding in 2-D DFT domain. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 10(11):1741–53, January 2001.

[4] Mauro Barni, Franco Bartolini, and Alessandro Piva. Improved wavelet-based watermarking through pixel-wise masking. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 10(5):783–91, January 2001.

[5] Wei Liu, Lina Dong, and Wenjun Zeng. Optimum Detection for Spread-Spectrum Watermarking That Employs Self-Masking. IEEE Transactions on Information Forensics and Security, 2(4):645–654, December 2007.

[6] A Piva, M Barni, F Bartolini, and V Cappellini. Threshold Selection for Correlation-Based Watermark Detection. In Proc. COST254 Workshop on Intelligent Communications, pages 5—-6, 1998.

[7] Qiang Cheng and T.S. Huang. Robust optimum detection of transform domain multiplicative watermarks. IEEE Transactions on Signal Processing, 51(4):906–924, April 2003.

[8] Michael Arnold, Peter G. Baum, and Xiao-Ming Chen. Robust detection of audio watermarks after acoustic path transmission. In Proceedings of the 12th ACM workshop on Multimedia and security - MM&Sec '10, page 117, New York, New York, USA, 2010. ACM Press.

[9] WH Lin, SJ Horng, and TW Kao. An Efficient Watermarking Method Based on Significant Difference

of Wavelet Coefficient Quantization. IEEE Transactions on Multimedia, 10(5):746–757, August 2008.

[10] Roland Kwitt, Peter Meerwald, and Andreas Uhl. Blind DT-CWT domain additive spread-spectrum watermark detection. In 2009 16th International Conference on Digital Signal Processing, pages 1–8. IEEE, July 2009.

[11] D. M. Hawkins. Identification of Outliers. Biometrical Journal, 29(2):198–198, 1980.

[12] Victoria Hodge and Jim Austin. A Survey of Outlier Detection Methodologies. Artificial Intelligence Review, 22(2):85–126, October 2004.

[13] Frank E. Grubbs. Procedures for detecting outlying observations in samples. Technometrics, 11(1):1–21, 1969.

[14] B. Rosner. Percentage Points for a Generalized ESD Many-Outlier Procedure. Technometrics, 25(2):165–172, 1983.

[15] W. J. Dixon. Processing data for Outliers. Biometrics, 9(1):74—-89, 1953.

[16] Gary L. Tietjen and Roger H. Moore. Some Grubbs-Type Statistics for the Detection of Outliers. Technometrics, 14(3):583—-597, 1972.

[17] J. P. Lewis. Fast normalized cross-correlation. In Vision interface, pages 120–123, 1995.

[18] P Meerwald, C Koidl, and A Uhl. Attack on "Watermarking Method Based on Significant Difference of Wavelet Coefficient Quantization". IEEE Transactions on Multimedia, 11(5):1037–1041, August 2009.

[19] V. Barnett and T. Lewis. Outliers in Statistical Data. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons; Chichester, 1994.

[20] C Croarkin and P Tobias. NIST/SEMATECH e-handbook of statistical methods. Retrieved January, 1:2014, 2014.