EFFICIENT DETECTION OF ADDITIVE WATERMARKING IN THE DWT-DOMAIN

Roland Kwitt, Peter Meerwald, Andreas Uhl

Dept. of Computer Sciences, University of Salzburg
Jakob-Haringer-Str. 2, A-5020 Salzburg, Austria
phone: + (43) 0662-8044-6347, fax: + (43) 0662-8044-172, email: {rkwitt,pmeerw,uhl}@cosy.sbg.ac.at

ABSTRACT

This paper aims at efficient, blind detection of additive spread-spectrum watermarks in the DWT domain. In our approach, the marginal distributions of the DWT detail subband coefficients are modeled either by the Generalized Gaussian distribution or by the recently proposed one-parameter Cauchy distribution. We investigate the computational demands for parameter estimation, hypothesis testing and threshold selection. Further, we discuss the tradeoff between computation time and detection accuracy.

1. INTRODUCTION

Watermarking has been proposed as a technology to ensure copyright protection by embedding an imperceptible, yet detectable signal in digital multimedia content such as images or video [1]. Many detection approaches for additive watermarks embedded in Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT) coefficients have been proposed in literature so far [2, 4]. For blind watermarking, i.e. when detection is performed without reference to the unwatermarked signal, the host transform coefficients are considered as noise from the viewpoint of signal detection. If we assume Gaussian noise, it is known that the optimal detector is the straightforward linear-correlation (LC) detector [6].

Unfortunately, DCT and DWT coefficients do not obey a Gaussian law in general which renders the LC detector suboptimal and modeling the host signal becomes crucial for detection performance [4]. The authors derive an optimal detector for additive watermarking in DCT transform coefficients following a Generalized Gaussian Distribution (GGD). In [2], it is shown that the mid-frequency DCT coefficients can also be modeled by the family of symmetric alpha-stable distributions and a detector for Cauchy distributed DCT coefficients is derived following the framework of [4]. However, both approaches are based on the strong assumption that the watermark embedding power is known to the detector. In [11], a new watermark detector based on the Rao hypothesis test [12] is proposed for watermark detection in GGD noise. The detector is asymptotically optimal (i.e. for large data records) and does not depend on knowledge about the embedding power any more.

In this work we are not only concerned about the watermark detection performance but also about the computational effort. With efficient watermark detection in mind, we compare the Rao-Cauchy detector first presented in [9] against state-of-the-art detectors. The effort for parameter estimation of the host signal model is often neglected. Clearly, using fixed, pre-determined values (such as proposed in [4]) is suboptimal in terms of detection performance. Alternatively, a number of fast parameter estimation methods [8, 14] provide a tradeoff between computational effort versus detection performance. In this paper we evaluate the impact of fast parameter estimation for the GGD and Cauchy host signal model on watermark detection performance.

Section 2 reviews two statistical models for DWT coefficients and the watermark detection problem. Fast parameter estimation methods are devised in Section 3 before we assess their tradeoff in Section 4. Runtime is compared in Section 5 before we conclude with a discussion of open problems.

2. STATISTICAL MODELS & DETECTION

We assume that a bipolar watermark sequence is embedded in DWT transform coefficients and that the watermark embedding power is unknown at the detection stage. We denote with $\mathbf{H}_j$, $\mathbf{V}_j$ and $\mathbf{D}_j$ the detail subbands with horizontal, vertical and diagonal orientation at scale j of the pyramidal DWT. When it is not necessary to speak of a specific subband, N is the number of subband coefficients and the coefficients are given by $x[1],\ldots,x[N]$ (vector notation). The elements of the bipolar watermark sequence used for marking an arbitrary subband are denoted by $w[t], 1 \leq t \leq N$ with $w[t] \in \{+1,-1\}$. For the rest of the paper, small boldface letters denote vectors, big boldface vectors denote matrices. Additive embedding of the watermark sequence is performed by

$$y[t] = x[t] + \alpha w[t], \quad t \in 1,\ldots,N \tag{1}$$

where $\alpha \in \mathbb{R}$ denotes the watermark embedding power, $y[t]$ denotes a watermarked transform coefficient and $x[t]$ denotes a host image transform coefficient. To derive a hypothesis test, we assume that the transform coefficients $x[t]$ represent a random sample drawn from some underlying probability density function (PDF). For blind detection, the host signal acts as noise and accurate modeling is the key element in deriving a detector.

2.1 Models for Host Signal Noise

It is commonly accepted that the marginal distributions of the detail subband coefficients of natural images are highly non-Gaussian but can be well modeled by the GGD (see [4, 10]). The PDF of the GGD is given by

$$p(x|b,c) = \frac{c}{2b\Gamma(1/c)} \exp\left(-\left|\frac{x}{a}\right|^c\right), \tag{2}$$

with $-\infty < x < \infty$ and $b,c > 0$. In contrast to the Gaussian distribution (which arises as a special case of the GGD for $c = 2$), the GGD is a leptokurtic distribution which allows heavy-tails. A second model is the one-parameter Cauchy distribution which is a member of the family of symmetric alpha-stable (SαS) distributions. This model has already

Supported by Austrian Science Fund project FWF-P19159-N13.

been successfully employed for blind watermarking of low- to mid-frequency DCT coefficients [2] and DWT detail subband coefficients [9]. The PDF of the Cauchy distribution with location parameter $-\infty < \delta < \infty$ and shape parameter $\gamma > 0$ is given by [7]

$$p(x|\gamma,\delta) = \frac{1}{\pi}\frac{\gamma}{\gamma^2+(x-\delta)^2}. \tag{3}$$

In contrast to the Gaussian distribution, the tails of the Cauchy distribution decay at a rate slower than exponential, hence we observe heavy-tails in the PDF.

2.2 Hypothesis Tests for Watermark Detection

Conditioned on one of the noise models from above, we can pursue two ways to derive a watermark detector. First, the most widely-used approach is to construct a Likelihood-Ratio test. Based on the embedding rule of Eq. (1), we can formulate a two-sided parameter test

$$\begin{aligned} &\mathscr{H}_0 : \alpha = 0, \theta \text{ (no/other watermark)} \\ &\mathscr{H}_1 : \alpha \neq 0, \theta \text{ (watermarked)} \end{aligned} \tag{4}$$

with θ denoting a vector of unknown parameters. These parameters are not directly related to the detection problem, but affect the shape of the PDFs under $\mathscr{H}_0$ (null-hypothesis) and $\mathscr{H}_1$ (alternative hypothesis). Depending on the noise model, θ contains either the shape and scale parameter of the GGD or the shape parameter of the Cauchy distribution. Given that α and θ are known a-priori we can formulate a Neyman-Pearson (NP) detector to decide $\mathscr{H}_1$ if

$$L(y) = \frac{p(y;\mathscr{H}_1)}{p(y;\mathscr{H}_0)} > T \tag{5}$$

where T denotes a detection threshold obtained by exploiting the NP criterion [6]. The ratio in Eq. (5) is termed the Likelihood-Ratio and the corresponding test will be referred to as the Likelihood-Ratio test (LRT). Conditioned on the two host noise models, we obtain the LRT statistics for the Generalized Gaussian model (LRT-GG) [4]

$$\rho(\mathbf{y}) = \frac{1}{b^c}\sum_{t=1}^{N}\left(|y[t]|^c - |y[t]-\alpha\cdot w[t]|^c\right) \tag{6}$$

and the Cauchy model (LRT-C) [2]

$$\rho(\mathbf{y}) = \sum_{t=1}^{N}\log\left(\frac{\gamma^2+y[t]^2}{\gamma^2+(y[t]-\alpha\cdot w[t])^2}\right). \tag{7}$$

Obviously, discarding the assumptions of known α and θ, we face a detection problem where we have a nuisance parameter θ and a signal $w[t]$ of unknown amplitude α. In practice, it is assumed that the watermark embedding process does not alter the noise characteristics significantly which allows to estimate θ and α from the watermarked image. To be rigorous, missing a-priori knowledge of α and θ leads to a so called Generalized Likelihood Ratio Test (GLRT) with the disadvantage that we need to determine the Maximum-Likelihood estimate of α under both $\mathscr{H}_0$ and $\mathscr{H}_1$ which is difficult to obtain in case of non-Gaussian noise [5]. The problem can be remedied by using the Rao hypothesis test instead which has the same asymptotic performance as the GLRT but only requires computation of the MLEs under $\mathscr{H}_0$ [5]. We obtain the test statistics of the Rao test conditioned on the Generalized Gaussian model (R-GG) [11]

$$\rho(\mathbf{y}) = \frac{\sum_{t=1}^{N}\operatorname{sgn}(y[t])\cdot w[t]\cdot|y[t]|^c}{\sum_{t=1}^{N}|y[t]|^{2c}} \tag{8}$$

and the Cauchy host noise model (R-C) [9]

$$\rho(\mathbf{y}) = \left[\sum_{t=1}^{N}\frac{y[t]w[t]}{\hat{\gamma}^2+y[t]^2}\right]^2\frac{8\hat{\gamma}^2}{N}. \tag{9}$$

2.3 Threshold Determination

Determining a reasonable detection threshold differs significantly depending on the hypothesis test and noise model. According to [2, 4] the detection statistics Eq. (6) and (7) follow a Gaussian law with mean μ_0 under $\mathscr{H}_0$ and mean μ_1 under the alternative hypothesis. Further $\mu_0 = -\mu_1$ and the variances under both hypothesis are approximately equal, $\sigma_0^2 = \sigma_1^2$. The probability of false alarm (P_f) can now be formulated as

$$P_f = \frac{1}{2}\operatorname{erfc}\left(\frac{T-\mu_0}{\sqrt{2}\sigma}\right) \tag{10}$$

which allows us to set $T = \sqrt{2}\sigma\operatorname{erfc}^{-1}(2P_f)+\mu_0$ according to the NP criterion [1].

According to [6], both Rao test statistics, Eq. (8) and (9), follow a χ_1^2 distribution with one degree of freedom under $\mathscr{H}_0$. Exploiting the relation $Q_{\chi_1^2}(x) = 2Q(\sqrt{x})$ where the functions $Q(\cdot)$ and $Q_{\chi_1^2}(\cdot)$ express the right-tail probabilities of the Gaussian and χ_1^2 distribution, resp., we can write P_f as

$$P_f = \operatorname{erfc}\left(\sqrt{\frac{T}{2}}\right) \quad \text{and express} \quad T = \left[\mathrm{Q}^{-1}\left(\frac{P_f}{2}\right)\right]^2. \tag{11}$$

Under the alternative hypothesis $\mathscr{H}_1$, both Rao test statistics follow a non-central Chi-Square distribution χ_λ^2 with non-centrality parameter λ shown in [11] and [9], resp.

3. PARAMETER ESTIMATION

In this section, we discuss the parameter estimation process of the aforementioned detectors due to the heavy impact on the computational performance. For the GGD, the method of Maximum-Likelihood is widely used. Let $x[1],\ldots,x[N]$ denote i.i.d. random variables following a GGD with shape β and scale α, the MLE of β requires to find the root of the transcendental equation [3]

$$\begin{aligned} 1 + \frac{\Psi(1/\hat{\beta}) + \log\left(\frac{\hat{\beta}}{N}\sum_{i=1}^{N}|x_i|^{\hat{\beta}}\right)}{\hat{\beta}} - \\ \frac{\sum_{i=1}^{N}|x_i|^{\hat{\beta}}\log(|x_i|)}{\sum_{i=1}^{N}|x_i|^{\hat{\beta}}} =: g(\beta). \end{aligned} \tag{12}$$

For Newton-Raphson root-finding, we need the derivative $g'(\beta)$ w.r.t. β which involves computation of the Digamma $\Psi(\cdot)$ and Trigamma $\Psi'(\cdot)$ function. The update step of the Newton-Raphson iteration then follows as $\beta_{n+1} = \hat{\beta}_n - g(\hat{\beta}_n)/g'(\hat{\beta}_n)$. Apart from the computational difficulties in solving Eq. (12), another problem is the search for a good starting value β_1 for the iteration. Usually moment estimates are employed to ensure a starting value close to the MLE for fast (quadratic) convergence. Mallat [10] relates the ratio of the first two sample moments m_1 and m_2 given by

$$m_1 = \frac{1}{N}\sum_{i=1}^{N}|x_i| \quad \text{and} \quad m_2 = \frac{1}{N}\sum_{i=1}^{N}|x_i|^2, \tag{13}$$

to the theoretical moments as

$$\frac{m_1^2}{m_2} = \frac{\Gamma^2(2/\beta)}{\Gamma(1/\beta)\Gamma(3/\beta)}. \quad (14)$$

Lacking analytic expressions, we must employ numerical methods. Do and Vetterli [3] employ a lookup table and linear interpolation to speed up estimation of β_1. However, another computationally fast method is presented in [8]. By defining the right-hand side of Eq. (14) as $F(\beta)$, the idea is to find an invertible approximation of $F(\beta)$. A candidate function proposed in [8] is

$$R(\beta) = \exp(a + b\beta^c) \quad (15)$$

with suitable coefficients a, b and c. $R(\beta)$ can easily be inverted and gives the estimate for β as follows

$$\hat{\beta} = \left(\frac{\log(R) - a}{b}\right)^{1/c} \text{ with } R = \frac{m_1^2}{m_2}. \quad (16)$$

Solving this non-linear curve fitting problem leads to the coefficients $a = -0.2667$, $b = -0.4172$ and $c = -1.1585$. In contrast to [8], we do not cut the range of β into a set of intervals and optimize a, b and c separately for each interval. The estimate can either be used directly or plugged in as a starting value for the Newton-Raphson iteration. We will use it in the latter way. Once we obtain the MLE $\hat{\beta}$, the MLE of $\hat{\alpha}$ has an explicit expression

$$\hat{\alpha} = \left(\frac{\hat{\beta}}{N}\sum_{i=1}^{N}|x_i|^{\hat{\beta}}\right)^{1/\hat{\beta}}. \quad (17)$$

Computationally, we favor the invertible approximation of $F(\beta)$ together with evaluation of Eq. (17) over the iterative approach based on Eq. (12). We further note that another recently proposed estimation method exists [13] based on exploiting a convex shape equation, leading to a computationally appealing estimation procedure. However, this method also requires a numerical root finding procedure which is why we omit further discussion here.

Regarding the estimation of the Cauchy distribution parameter γ, we start with the ML estimation procedure. Given that $x[1], \ldots, x[N]$ denote realizations of N i.i.d. random variables following a Cauchy distribution with $\delta = 0$, the ML estimate of γ is given as the solution (see [7]) to

$$\frac{1}{N}\sum_{t=1}^{N}\frac{2}{1 + (x[t]/\hat{\gamma})^2} - 1 = 0 \quad (18)$$

which has to be solved numerically, e.g. using Newton-Raphson. For that purpose, define the left hand side of Eq. (18) as a function of $\hat{\gamma}$, denoted by $h(\hat{\gamma})$. The estimate of γ in the n-th iteration ($n > 1$) of the Newton-Raphson algorithm is computed by $\hat{\gamma}_{n+1} = \hat{\gamma}_n - h(\hat{\gamma}_n)/h'(\hat{\gamma}_n)$. Here, $h'(\cdot)$ denotes the first derivative of h w.r.t. γ, given by

$$h'(\gamma) = \frac{4\gamma}{N}\sum_{t=1}^{N}\frac{x[t]^2}{(\gamma^2 + x[t]^2)^2}. \quad (19)$$

As the starting value $\hat{\gamma}_1$ for the numerical calculation, an estimation value based on sample quantiles [7] can be used

$$\hat{\gamma}_1 = 0.5(x_p - x_{1-p})\tan(\pi(1-p)), \quad (20)$$

with $0.5 < p < 1$ and x_p, x_{1-p} denoting the sample quantiles. Moment estimation is not possible in case of the Cauchy distribution, since the moments do not exist. As with the GG shape parameter β, it is of course possible to use Eq. (20) alone or even set γ to a fixed value.

Figure 1: Grayscale test images (*Lena*, *Elaine*, *Barbara*, *Boat*, *Bridge*, *Peppers*)

4. IMPACT ON DETECTION PERFORMANCE

To study the impact of different host signal parameter estimation approaches on the detection performance, we plot the detection performance as a function of the Cauchy γ and GGD shape parameter c [4]. Given an estimate of the distribution parameter $\hat{\lambda}$ under $\mathscr{H}_1$ and the threshold T, the estimated probability of missing the watermark ($\hat{P}_m$) in case of the Rao detectors can be directly computed from

$$\hat{P}_m = 1 - Q\left(\sqrt{T} - \sqrt{\hat{\lambda}}\right) + Q\left(\sqrt{T} + \sqrt{\hat{\lambda}}\right) \quad (21)$$

and in case of the LRT detectors we have to resort to

$$\hat{P}_m = 1 - Q\left(\frac{T - \hat{\mu}_1}{\hat{\sigma}}\right). \quad (22)$$

The estimates $\hat{\lambda}, \hat{\mu}_1$ and $\hat{\sigma}$ are calculated from the detection responses under $\mathscr{H}_1$ when embedding and detecting $M = 1000$ random watermarks. The complementary tests for checking whether the detection statistics actually follow the assumed distributions (which allows reliable threshold determination) are explained in Section 5. Since a non-central Chi-Square random variable with one degree of freedom and non-centrality parameter λ is equal to the square of a Gaussian random variable with mean $\sqrt{\lambda}$ and unit variance, we can estimate λ from the responses $\rho_1, \ldots, \rho_M$ as follows

$$\hat{\lambda} = \left(\frac{1}{M}\sum_{i=1}^{M}\sqrt{\rho_i}\right)^2 \quad (23)$$

and $\hat{\mu}_1, \hat{\sigma}^2$ are estimated by the sample mean and variance. In Figure 2, we plot the LRT-Cauchy and Rao-Cauchy detector performance in terms of $\hat{P}_m$ for a probability of false alarm $P_f = 10^{-6}$ as a function of the Cauchy γ parameter for the six test images shown in Figure 1. The γ parameter varies between 0.1 and 45 with a step size of 0.1. The detector performance w.r.t. to the MLE and its fast approximation using quantile estimation, Eq. (20), are denoted by a *circle* and *diamond*, resp. We observe that the fast, approximative γ is very close to the ML estimate and the difference has a negligible impact on watermark detection performance. Further, it is evident that neither the ML nor the quantile estimation approach reaches the peak detection performance. We can see, however, that the parameter estimates are close enough to achieve good performance. The steep ascent of the curve to the left suggests that choosing too small values for γ quickly deteriorates detection performance.

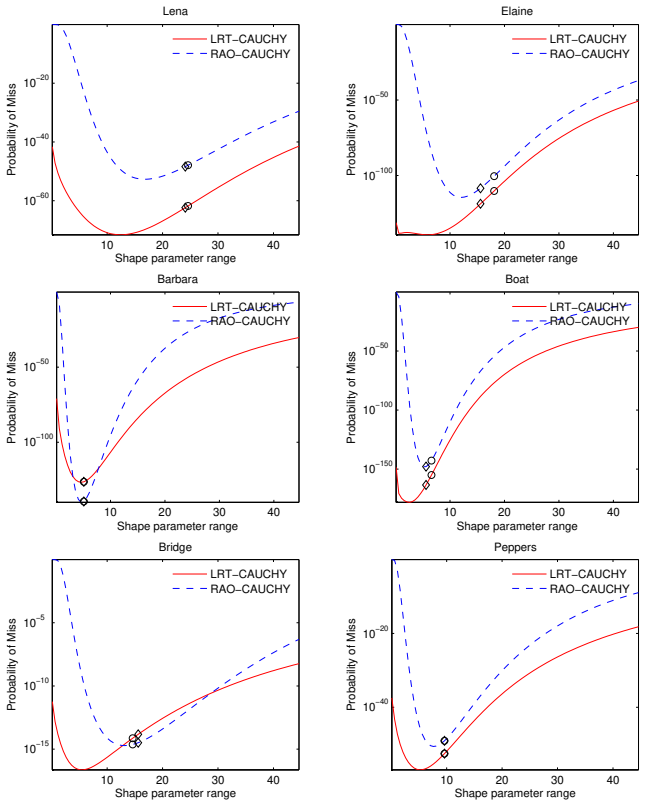

Figure 2: Detection performance of the LRT-Cauchy and Rao-Cauchy as a function of γ at DWR 20 dB; $P_f = 10^{-6}$

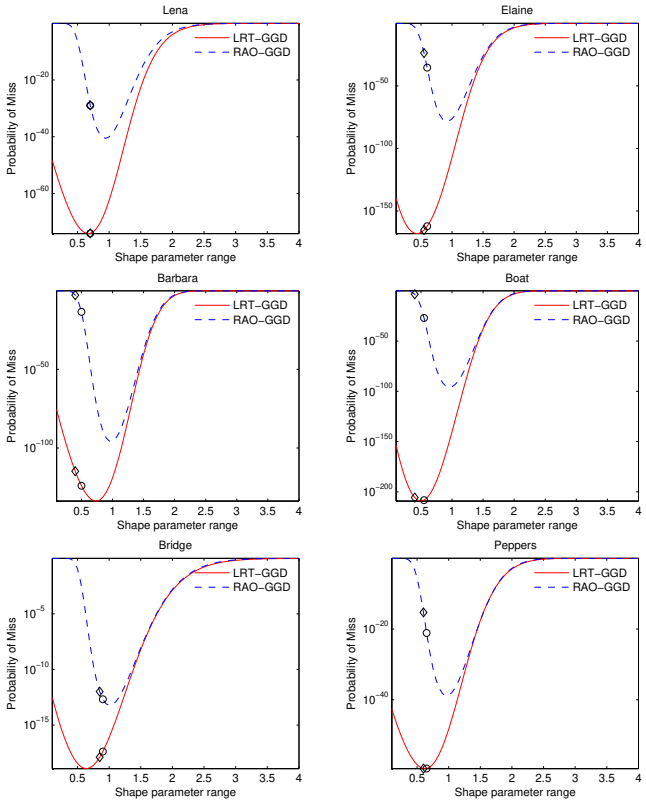

Figure 3: Detection performance of the LRT-GGD and Rao-GGD as a function of c at DWR 20 dB; $P_f = 10^{-6}$

Figure 3 shows the plots for the LRT-GG and Rao-GG detector. The GGD shape parameter c varies from 0.02 to 4. Again, the fast parameter estimate using the invertible approximation to $F(\beta)$, Eq. (16), is reasonably close to the ML estimate. In case of the LRT-GG detector, peak performance is missed by only a small margin. However, in case of the Rao-GG detector, the ML estimate of the shape parameter is far off the value achieving maximum detection performance.

Based on the experiments we conclude that fast, approximate estimation of the shape parameter can replace the ML estimate without significantly reducing detection performance. For the Cauchy detectors, ML estimation of the shape parameter is not equivalent to searching the parameter value that maximizes detection performance, as has been noted earlier for the LRT-GG detectors [4]. For the Rao-GG detector, the parameter's ML estimate does not achieve good detection performance for many images. In Nikolaidis' work [11], the detection results for the Rao-GG detector vary widely across images and the Rao-GG sometimes performs even worse than the LC detector. We speculate that the reason is the same that becomes evident when observing the gap between the peak performance and the estimate parameter values in Figure 3. For the LRT-GG based detector, the use of a fixed, image independent shape parameter has been proposed (e.g. 0.8 for DWT coefficients [4]) to save the estimation effort. We can immediately read off the resulting performance from the plots in Figure 3 and also compare with the performance of the LC detector (setting $c = 2$). For the Cauchy host signal model, no fixed value for the γ parameter has been proposed yet. A good candidate as an image independent parameter is $\gamma = 8$ for the DWT details subband. Comprehensive tests with two large natural image databases confirm our observations.

5. EXPERIMENTAL RESULTS

For the experiments in Section 4 we use the six grayscale test images of size 256×256 shown in Figure 1. The watermark is embedded in the $\mathbf{H}_2$ subband and α is adjusted to reach a Document-to-Watermark Ratio (DWR) of 20 dB for each image. We employ the biorthogonal CDF 9/7 filter. The resulting PSNR is 48.41, 49.28, 47.95, 50.40, 49.51 and 47.88 dB for the *Lena* to *Peppers* image, resp. Due to the high PSNR, no visual degradation can be noticed. The implementation[1] of the detectors and all experiments were conducted in MATLAB. We calculate the detector responses under $\mathscr{H}_0$ and $\mathscr{H}_1$ for 1000 randomly generated watermarks and conduct Chi-Square Goodness-of-Fit (GoF) tests at 5% significance to verify that the detection statistics are either Gaussian (in case of LRT-C, LRT-GG) or Chi-Square with one degree of freedom (in case of R-GG, R-C). Assuming that there is no rejection of the null-hypothesis of the Chi-Square GoF test, we can reliably estimate the detection statistic parameters and check whether the estimates conform to the theoretical values. This is especially important under $\mathscr{H}_0$, since a strong deviation impedes determination of the threshold T according to the probability of false-alarm. In contrast to the LRT detectors, the Rao test does not require computation of detection statistic parameters under $\mathscr{H}_0$ since $\rho \sim \chi^2_1$ and the threshold can easily be calculated from Eq. (11).

Before we examine runtime results for parameter estimation and detection, we compare the number of required

[1] Software available at `http://www.wavelab.at/sources`.

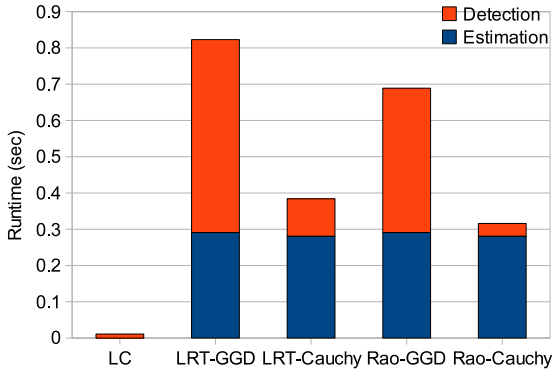

Figure 4: Runtime for signal length $N = 1000000$

Table 1: Number of arithmetic operations

Detector	**Operations**			
	$+,-$	$\times,\div$	pow, log	abs, sgn
LC	N	$N+1$		
LRT-GG [4]	$3N$	2	$2N+1$	$2N$
LRT-C [2]	$4N$	$5N$	N	
R-GG [11]	$2N$	$3N+2$	N	$2N$
R-C [9]	$2N$	$2N+4$		

arithmetic operations to calculate the detection statistics for a signal of length N. Table 1 gives the number of additions and subtractions $(+,-)$, multiplications and divisions $(\times,\div)$, logarithms and exponentiations (log, pow) as well as absolute and signum (abs, sgn) operations. From these numbers it is obvious that the LC detector is by far the simplest in terms of arithmetic operations, since it involves only summations and multiplications of floating point numbers. However, the R-C detector is only slightly more expensive, since the exponentiations in Eq. (9) merely involve integer exponents and additions as well as multiplications which can be very efficiently performed with few CPU cycles. In contrast to that, the LRT-C detector requires N computations of the logarithm and the LRT-GG detector even requires exponentiations with floating point numbers, which is quite expensive in terms of CPU cycles; see Eq. (6) and (7).

In Figure 4 we compare the parameter estimation and watermark detection runtime in seconds for a signal of length $N = 1000000$ on an Intel Core2 2.6 GHz CPU running MATLAB 7.4. Results have been averaged over 10 test runs. Note that the relative detection runtimes are in good agreement with the analytical comparison provided in Table 1. The detectors based on the Cauchy host model are significantly faster than the LRT-GG and R-GG detector. The R-C detector leads by a small margin. For the LC detector, no parameter estimation is required. GGD and Cauchy parameter estimation has been performed with the fast, approximative methods, see Eq. (16) and (20), resp. Note that computation of sample quantiles for the γ estimate requires sorting the data while the moment estimates can be computed with linear effort; so for larger N, the GGD parameter estimation will turn out faster. Our implementation of the described MLE methods is significantly slower (by a factor of five to ten according to experiments; we have to omit details here) due to the iterative Newton-Raphson approach.

6. CONCLUSION

We have compared five blind detectors for additive, spread-spectrum watermarking in terms of runtime efficiency for parameter estimation and hypothesis testing. Further, the impact of fast, approximate parameter estimation for the GGD and Cauchy host noise model was assessed. The Rao-Cauchy detector turned out attractive not only in terms of runtime efficiency but also due to straightforward threshold determination. As further research, we plan to investigate the tradeoff for fixed parameter settings over large test image sets.

REFERENCES

[1] M. Barni and F. Bartolini. *Watermarking Systems Engineering*. Marcel Dekker, 2004.

[2] A. Briassouli, P. Tsakalides, and A. Stouraitis. Hidden Messages in Heavy-Tails: DCT-Domain Watermark Detection Using Alpha-Stable Models. *IEEE Transactions on Multimedia*, 7(4):700–715, Aug. 2005.

[3] M. Do and M. Vetterli. Wavelet-based texture retrieval using Generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing*, 2(2):146–158, Feb. 2002.

[4] J. R. Hernández, M. Amado, and F. Pérez-González. DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure. *IEEE Transactions on Image Processing*, 9(1):55–68, Jan. 2000.

[5] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*, volume 1. Prentice-Hall, 1993.

[6] S. M. Kay. *Fundamentals of Statistical Signal Processing: Detection Theory*, volume 2. Prentice-Hall, 1998.

[7] K. Krishnamoorthy. *Handbook of Statistical Distributions with Applications*. Chapman & Hall, 2006.

[8] R. Krupinski and J. Purczynski. Approximated fast estimator for the shape parameter of Generalized Gaussian distribution. *Signal Processing*, 86(2):205–211, Feb. 2006.

[9] R. Kwitt, P. Meerwald, and A. Uhl. A lightweight Rao-Cauchy detector for additive watermarking in the DWT-domain. In *Proceedings of the ACM Multimedia and Security Workshop, MMSEC '08*, pages 33–41, Oxford, UK, Sept. 2008. ACM.

[10] S. Mallat. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, July 1989.

[11] A. Nikolaidis and I. Pitas. Asymptotically optimal detection for additive watermarking in the DCT and DWT domains. *IEEE Transactions on Image Processing*, 12(5):563–571, May 2003.

[12] C. R. Rao. *Linear Statistical Inference and Its Applications*. Probability and Mathematical Statistics. Wiley, 1973.

[13] K.-S. Song. A globally convergent and consistent method for estimating the shape parameter of a Generalized Gaussian distribution. *IEEE Transaction on Information Theory*, 52(2):510–527, Feb. 2006.

[14] M. Varanasi and B. Aazhang. Parameteric Generalized Gaussian density estimation. *Journal of the Acoustical Society of America*, 86(4):1404–1415, Oct. 1989.