Hidden Messages in Heavy-Tails: DCT-Domain Watermark Detection Using Alpha-Stable Models

Alexia Briassouli, Panagiotis Tsakalides, *Member, IEEE*, and Athanasios Stouraitis, *Senior Member, IEEE*

***Abstract*—This paper addresses issues that arise in copyright protection systems of digital images, which employ blind watermark verification structures in the discrete cosine transform (DCT) domain. First, we observe that statistical distributions with heavy algebraic tails, such as the alpha-stable family, are in many cases more accurate modeling tools for the DCT coefficients of JPEG-analyzed images than families with exponential tails such as the generalized Gaussian. Motivated by our modeling results, we then design a new processor for blind watermark detection using the Cauchy member of the alpha-stable family. The Cauchy distribution is chosen because it is the only non-Gaussian symmetric alpha-stable distribution that exists in closed form and also because it leads to the design of a nearly optimum detector with robust detection performance. We analyze the performance of the new detector in terms of the associated probabilities of detection and false alarm and we compare it to the performance of the generalized Gaussian detector by performing experiments with various test images.**

***Index Terms*—Alpha-stable distributions, discrete cosine transform, image watermarking, Neyman–Pearson detector, statistical modeling.**

I. Introduction

DURING the last decades, imaging technology and applications have advanced greatly as multimedia information is acquired, represented, stored, and distributed in digital format. Digital processing and transmission has major advantages, but problems can arise due to the ease with which digital information is reproduced and distributed. Original multimedia information, such as digital images, music or video, is easily copied without any loss of fidelity and it can be readily distributed to unauthorized users. Traditional methods for the prevention of illegal use of digital multimedia data include the addition of access control headers, the encryption of the original data, and other copy protection techniques. These methods do not provide an effective way of securing intellectual property rights, as the original data remains unprotected once the copyright control mechanisms are surpassed. An access control header, for example, may easily be stripped from the data, leaving it completely unprotected. Cryptography, on the other hand, does not allow the use of digital data in its original form and once the data is decrypted, it is no longer protected. Other security applications, like steganography, aim to effectively hide a message in the cover data without revealing the methods used for its embedding.

A method for the protection of intellectual property rights that has been gaining popularity lately is the embedding of a digital watermark in multimedia data [1]. A signal is embedded directly in the data to be marked without altering it significantly, so, in contrast to encrypted data, watermarked signals can still be used while remaining protected. Watermarking also differs from steganography, as the entire embedding algorithm may be known, its security depending only on a secret key. It does not necessarily prevent the copying of digital data, but should rather be seen as a "last line of defense" in a copyright protection scheme, which identifies the original data source and its intended legitimate destination, so that copyright violations can at least be detected. Thus, watermarking should be considered as part of a general system and not a complete solution by itself. Its basic principle is to embed a signal directly into the data, which serves as a host or a cover for that information. The hidden signal travels with the data, which thus remains "marked" and protected, until the intended receiver removes the watermark. In some applications, such as the digital video disc (DVD), the watermark can also prevent the duplication of its contents, as the DVD player will not play an unwatermarked, illegally copied video [2].

Watermarking is very similar to the communications problem of embedding a weak and usually imperceptible signal in "host" data, i.e., the problem of information hiding. Security applications often use the hiding of information in digital data in order to protect it or to monitor its usage. The watermark is a digital signal, usually containing information about the data origin, destination and owner, as well as additional information concerning transaction dates, serial numbers etc., which can be useful when tracking illegal use of the data [1], [3]. In practical applications, there exist both perceptible and imperceptible watermarks. The former are usually used in preview images in the World Wide Web or in image databases to prevent their commercial use. Digital copyright systems employ imperceptible, i.e., invisible or inaudible watermarks, since their aim is to protect the copyright of the cover data without altering it [3].

Two main types of watermarking schemes exist. A watermarking system may embed a specific piece of information in

Manuscript received March 21, 2001; revised January 12, 2004. The work of P. Tsakalides was supported by the Greek General Secretariat for Research and Technology under Programs ΕΠΕΤ II, Code 97ΕΛ – 152 and ΕΠΑΝ, Code ΗΠΑ-011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tsuhan Chen.

A. Briassouli is with the Beckman Institute Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: briassou@vision.ai.uiuc.edu).

P. Tsakalides is with the Department of Computer Science, University of Crete and Institute of Computer Science, FORTH ICS-FORTH, 711 10 Heraklion, Crete, Greece (e-mail: tsakalid@ics.forth.gr).

A. Stouraitis is with the Department of Electrical and Computer Engineering, University of Patras 261 10 Rio, Greece (e-mail: thanos@ee.upatras.gr).

Digital Object Identifier 10.1109/TMM.2005.850970

the data, such as identification numbers used for image tracking or classification and for video distribution. In this case, the embedded watermark communicates a message which must be extracted with accuracy. The other type of system, often used in copyright-protection applications, embeds a watermark which modifies the original data, but does not necessarily communicate a message. Hence, only the presence of the watermark needs to be verified in order to determine whether or not the data is protected. These two schemes are inherently equivalent, as the detection of a watermark's presence is equivalent to the extraction of a *known* message, while the extraction of a watermark can be seen as an extension of the verification of many 1-bit watermarks.

An important issue concerning watermark recovery—detection or extraction—techniques is the availability of the original data. In general, watermark extraction is facilitated when the host data is available, since one can then detect any distortions in the data and invert them to accurately recover the hidden signal. However, in many practical applications such as data monitoring or tracking, the unwatermarked data is not always available. Moreover, in many multimedia settings, e.g., in video watermarking, the use of the host data is impractical because of its large volume. As a result, it is necessary to design so-called blind watermarking techniques that do not use the original data in the extraction process and treat it as noise superimposed on the hidden signal [1]. Recent advances in the information and game theoretic approach to the watermarking problem suggest that more information can be hidden in the host signal if the latter is not treated as additive noise, but as side information at the encoder, i.e., the embedder [4], [5]. Nevertheless, we consider the popular spread spectrum watermarking system, where the original cover signal is still treated as noise, since these schemes are very often encountered in the literature [6], [7]. This allows us to implement a system whose performance can be compared to that of other systems found in the literature [8].

In this paper, we address the problem of watermark detection in DCT transformed images, i.e., the problem of verifying whether a DCT-domain image has been watermarked. Various methods have been proposed for the detection of a watermark in this domain, but most are based on correlator structures and do not take into account the actual statistical properties of the data [4], [9]. A more systematic approach is to treat watermark verification as a statistical detection problem and to formulate a binary hypothesis test, concerning whether an image was watermarked or not. This approach leads to a detection scheme which is more appropriate for DCT-domain images than a simple Gaussian correlator, since it is based on the data non-Gaussian statistical properties. The detector performance can be analyzed theoretically [10]–[12], by measuring and relating the probability of detection to the probability of false alarm. It is essential to predetermine the false alarm probability and to keep it at low levels, especially in commercial systems, as it directly affects their credibility.

The commonly used correlation detectors are optimal only in the case of Gaussian data [11]. However, it is well known that important DCT coefficients, which correspond to the low and mid frequencies, do not follow a Gaussian distribution [13]. Hernandez *et al.* proposed the design of a statistical watermark detector that models the DCT coefficients by the commonly used generalized Gaussian distribution [14]. The resulting detector structure based on this model led to considerable improvements when compared to the correlation receiver [8]. In this paper, we claim that the alpha-stable family of distributions is sufficiently flexible and rich to characterize the DCT coefficients at least as accurately as the generalized Gaussian distribution. Motivated by our modeling results, we design a novel statistical watermark detector based on the Cauchy member of the alpha-stable family and we analyze its theoretical performance. We use the Cauchy model as it is the only non-Gaussian alpha-stable distribution with a closed form probability density function (pdf), but also because it leads to improved, robust detection schemes [15], [16]. Finally, we compare the performance of the Cauchy processor with that of the generalized Gaussian detector and we show that it achieves better detection than the generalized Gaussian scheme in a wide range of watermark-to-document power levels for real images.

The paper is organized as follows. In Section II, we provide some necessary preliminaries on watermark generation, embedding, and detection structures. In Section III, we present our results on the modeling of DCT coefficients of test images by means of alpha-stable distributions. In Section IV, we design a blind watermark detector based on the Cauchy distribution and we analyze its performance in terms of the associated probabilities of detection and false alarm. In Section V, we compare the performance of our proposed scheme with the performance of the generalized Gaussian detector in real images. Finally, in Section VI, we discuss some of the issues that need to be addressed in the future.

II. DCT-Domain Watermarking for Still Images

In this section, we briefly describe a watermark generation and embedding procedure for still images in the DCT domain, similar to the one presented in [8]. Many watermarking systems are based on additive spread spectrum ideas [6], [7], which are inspired by the spread spectrum modulation schemes used in digital communications in jamming environments [4], [17]. The role of the jammer in the watermarking problem is assumed by the host image (additive noise), as well as a malicious attacker who intentionally tries to destroy or extract the embedded signal [18]. Spread-spectrum techniques have gained popularity for watermarking because they exhibit low probability of intercept (LPI), which in the case of watermarking translates to high imperceptibility of the hidden signal [19]. The anti-jamming properties of these systems increase the robustness of the watermark, while the pseudorandom modulation of the hidden information signal increases the achieved security, since knowledge of this sequence is required for successful watermark retrieval [18].

A. Watermark Generation and Embedding

In the sections that follow, we shall represent an image in the spatial domain as a discrete 2-D sequence $x[\mathbf{n}]$ of its luminance component with $N_1 \times N_2$ pixels and its DCT transform as $X[\mathbf{k}]$. The watermark can be considered as a 2-D DCT signal

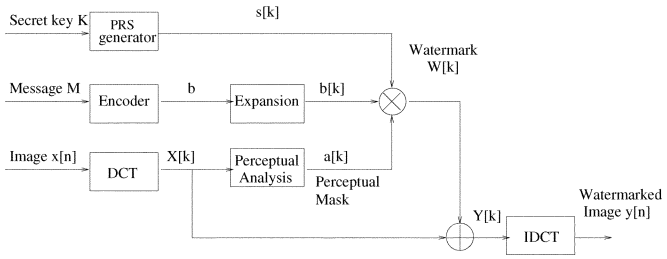

Fig. 1. Watermark embedding block diagram.

$W[\mathbf{k}]$, which is added to the DCT coefficients giving the watermarked image $Y[\mathbf{k}]$ (see Fig. 1). Boldface typesetting is used to represent 2-D indexes like $\mathbf{k} = [k_1, k_2]$ and $\mathbf{n} = [n_1, n_2]$, and vector quantities like the vector of message bits, $\mathbf{b}$. The discrete cosine transform can be applied either to the entire image or in blocks as in the JPEG standard [20], [21]. We used the block-wise DCT not only because of the JPEG standard, but also because its blockwise application is faster than the full-image DCT and, most importantly in our case, because it allows the statistical characterization of the DCT coefficients. Specifically, the application of the DCT in 8×8 blocks of the image leads to 64 DCT coefficients which can be zig zag scanned and thus arranged in order of decreasing importance. Using this ordering, the first coefficients are of the main interest as they correspond to the low and mid frequencies that carry the most information about the image. By applying the blockwise DCT, we can extract these coefficients from each block and estimate their statistical distribution, which is essential in the design of our detection system.

The watermark may or may not contain a message M, which is mapped by an encoder to a N-dimensional codeword vector $\mathbf{b}$ (Fig. 1). In the frequently considered case where the watermark does not contain a message, the codeword vector $\mathbf{b}$ reduces to $b = 1$ [8]. An expansion process follows, during which every element b_i of the codeword vector is repeated over a set S_i of the image pixels so as to cover the entire image. This introduces redundancy in the watermark, since it is repeated over more pixels than necessary for the coding of the message. In this manner it covers the entire image, so it is present in every coefficient and, at the same time, it achieves a certain degree of inherent robustness because of this redundancy.

In order to generate a direct spread spectrum modulated watermark, the resulting signal, consisting of the "spread" message bits, is multiplied by a two-dimensional pseudorandom sequence (PRS) $s[\mathbf{k}]$. The sequence is generated by a pseudorandom noise generator, which is initialized with a seed that depends on the cryptographic key K. This pseudorandom sequence is the *spreading sequence* of the system, which spreads the spectrum of the original signal over many frequencies, making it difficult to detect [19]. To achieve this, it needs to have white noise-like properties, so it takes values $+1$, -1, whose mean is zero and their autocorrelation approximately equal to the Kronecker delta function.

The invisibility of the alterations introduced to the image by the watermark is guaranteed by multiplying the pseudorandom sequence by a perceptual mask $a[\mathbf{k}]$. This mask is obtained by using a psychovisual model in the DCT domain which was proposed by Ahumada *et al.* in [22], [23] and takes into account the properties of the human visual system (HVS). As in [8], one must estimate the visibility threshold $T(i,j)$ for every (i,j) DCT coefficient of each 8×8 block. It is approximated in logarithmic units by the following function:

$$\log T(i,j) = \log\left(\frac{T_{\min}(f_{i,0}^2 + f_{0,j}^2)^2}{(f_{i,0}^2 + f_{0,j}^2)^2 - 4(1-r)f_{i,0}^2 f_{0,j}^2}\right) + K\left(\log\sqrt{f_{i,0}^2 + f_{0,j}^2} - \log f_{\min}\right)^2 \quad (1)$$

where $f_{i,0}$ and $f_{0,j}$ are the vertical and horizontal spatial frequencies, respectively, (in cycles/degree) of the DCT basis functions and $T_{\min}$ is the minimum value of the quadratic function $T(i,j)$ associated with $f_{\min}$. It must be noted that this model is only valid for the AC frequencies and not for the DC coefficient. Since the value of every AC DCT coefficient changes from block to block, the threshold $T(i,j)$ is corrected for every block by the formula

$$T'(i,j) = T(i,j)\left(\frac{X_{0,0}}{\overline{X}_{0,0}}\right)^{\alpha_T} \quad (2)$$

where $X_{0,0}$ is the DC coefficient for each block and $\overline{X}_{0,0}$ is the average screen luminance. These two functions contain parameters which are set to $r = 0.7$, $T_{\min} = 1.1548$, $K = 1.728$, $f_{\min} = 3.68$ cycles/degree and $\alpha_T = 0.649$, following [23]. The perceptual mask is then obtained from the visibility threshold according to the following equation:

$$a[\mathbf{k}] = 4 \cdot (1 + (\sqrt{2} - 1)\delta(l_1)) \cdot (1 + (\sqrt{2} - 1)\delta(l_2)) \cdot \gamma \cdot T'(l_1, l_2) \quad (3)$$

where $l_1 = k_1 \bmod 8$, $l_2 = k_2 \bmod 8$, $\delta(\cdot)$ is the Kronecker function, and $\gamma < 1$ is a scaling factor. Equation (3) gives the final expression for the values of the perceptual mask $a[\mathbf{k}]$ that corresponds to every DCT coefficient of each block. This mask ensures that the watermark will remain imperceptible to the human eye, but at the same time it allows the maximum possible alteration of the pixel values to achieve increased robustness and detectability.

Finally, the resulting watermark $W[\mathbf{k}] = a[\mathbf{k}]s[\mathbf{k}]$ is added to the original image DCT coefficients $X[\mathbf{k}]$, giving the watermarked image $Y[\mathbf{k}] = X[\mathbf{k}] + W[\mathbf{k}]$ (Fig. 1). The hidden signal can be retrieved if one uses a copy of the pseudorandom sequence $s[\mathbf{k}]$ that can be reproduced only if one knows the entire procedure through which it is generated and the cryptographic key that was used as its seed. Consequently, an attacker will not be able to extract the watermark without knowledge of the secret key, even if the entire watermark generation and embedding process is known. Thus, the dependence of the generated watermark on the cryptographic key and the high degree of redundancy achieved through the expansion process result in a particularly robust but invisible watermark.

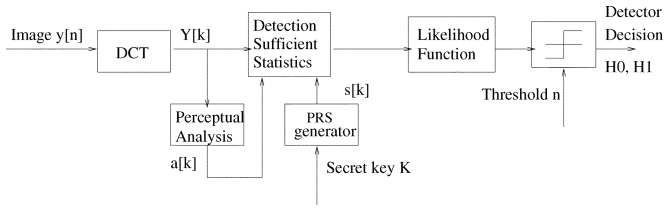

Fig. 2. Watermark detection block diagram.

B. Watermark Detection as a Binary Hypothesis Testing Problem

As already mentioned in Section I, although the watermark detection and extraction problems are often treated separately in the literature, they are inherently equivalent [1], [12]. Most copyright-protection applications contain a known watermark, so the verification of its existence is sufficient. Since we are dealing with the problem of watermark verification, in the following we assume that the image can contain only one possible watermark.

In order to verify the existence of a watermark based on the statistical properties of the given data, one must design a binary hypothesis test where the two hypotheses concern the existence or not of a known signal in a given image. The given data for this test is a DCT transformed image which possibly contains a watermark, but the test is performed without knowledge of the original, unwatermarked coefficients $X[\mathbf{k}]$, so this is actually a blind watermark detection problem (cf. Fig. 2). The two hypotheses can be formulated as follows:

$$\begin{aligned} H_0 &: Y[\mathbf{k}] = X[\mathbf{k}] \\ H_1 &: Y[\mathbf{k}] = X[\mathbf{k}] + W[\mathbf{k}]. \end{aligned} \tag{4}$$

In this detection problem, the watermark $W[\mathbf{k}]$ is the desired signal or information, while the DCT domain image $X[\mathbf{k}]$ plays the role of unknown additive noise. The goal of the optimal watermark detector is to find whether or not there is a watermark in the received image $Y[\mathbf{k}]$, based on the statistical properties of the image $X[\mathbf{k}]$ and its watermark $W[\mathbf{k}]$. A basic assumption is that the PDF of the original coefficients does not change with the embedding of the watermark [8]. This assumption is reasonable, since the watermarked image coefficients should have a distribution similar to that of the host in order to remain (perceptually) similar to it.

We assume that the data has been modeled by an appropriate statistical distribution, so the PDFs of the image and its watermark are known or can be accurately approximated from the given data. In that case, the decision rule is formulated as follows:

$$\Lambda(Y) \underset{H_0}{\overset{H_1}{\gtrless}} \eta' \tag{5}$$

with the likelihood ratio $\Lambda(Y)$, which is compared to the threshold η' is defined as

$$\Lambda(Y) = \frac{f(Y \mid H_1)}{f(Y \mid H_0)} \tag{6}$$

and the PDFs under each hypothesis are known. In practice, the log-likelihood ratio is usually preferred to perform hypothesis testing [11]. The log-likelihood ratio is simply defined as the natural logarithm of the likelihood ratio, $l(Y) = \ln(\Lambda(Y))$, so the decision rule becomes

$$l(Y) = \ln\left(\frac{f(Y \mid H_1)}{f(Y \mid H_0)}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{7}$$

The detection is performed by comparing the quantity $l(Y)$ to an appropriate threshold η. For Neyman–Pearson testing, which is examined here, the threshold η is determined by the probability of false alarm P_{fa}, so a reliable detector is designed for predefined false alarm rates [11].

III. Data Modeling

In the previous section, we presented a setup for treating watermark detection as a binary hypothesis test, based on the statistical properties of the data. Obviously, such a structure will lead to reliable detection results if the data, i.e., the low and mid frequency DCT coefficients, are modeled as accurately as possible. It is often mentioned in the literature that the Gaussian model is not appropriate for these DCT coefficients [8], [13], as they exhibit heavier tails than the normal distribution. The generalized Gaussian density (GGD) is frequently used to describe these coefficients, as its tails decay at a slower rate than the Gaussian distribution.

However, even the GGD does not always sufficiently characterize the low- and mid-frequency DCT coefficients. In [8], it is observed that some samples in the tails of the empirical distributions have relatively high amplitudes and consequently cannot be adequately modeled by the exponentially decaying tails of the GGD. This fact leads to a loss in detection performance, so the authors in [8] propose to use what they call *"point elimination techniques,"* which essentially discard the samples whose magnitude is above a certain threshold. This threshold cannot be determined analytically and it varies depending on the image, so point elimination approaches are heuristic and they cannot be reliably applied to a wide range of image data. Also, their performance cannot be measured objectively since it heavily relies on the arbitrary deletion of high-magnitude data samples. A more systematic and efficient approach to this problem is to consider a better model for the DCT coefficients, which can capture their heavy tailed nature, i.e., the high-magnitude samples in the tails. The symmetric alpha-stable ($S\alpha S$) distribution provides a rich and flexible modeling tool for heavy tailed non-Gaussian data [24], so it is expected to adequately describe the low- and mid-frequency DCT coefficients.

A. Laplacian and Generalized Gaussian Modeling of DCT Coefficients

The Laplacian and generalized Gaussian distributions have often been used in the literature to characterize transform domain (DCT or wavelet) image coefficients. The Laplacian PDF is given by

$$f_X(x) = \frac{\beta}{2}\exp(-\beta|x-\mu|) \tag{8}$$

where $\mu = mean(x)$, $\beta^2 = 2/var(x)$. As shown in [25], the Laplacian distribution is inadequate for the heavier tailed samples, because its tails decay at a fast exponential rate. The generalized Gaussian distribution provides a better model for the

low- and mid-frequency DCT coefficients [8], [13], [14], as its tails decay at a slower rate. This distribution, characterized by the parameter c and the standard deviation of the data σ, has the following PDF:

$$f_X(x) = A\exp(-\beta|x-\mu|^c) \tag{9}$$

whose parameters A and β can be expressed in terms of c and σ as follows:

$$\beta = \frac{1}{\sigma}\left(\frac{\Gamma\left(\frac{3}{c}\right)}{\Gamma\left(\frac{1}{c}\right)}\right)^{1/2}, \quad A = \frac{\beta c}{2\Gamma\left(\frac{1}{c}\right)}. \tag{10}$$

For $c = 1$, the generalized Gaussian distribution reduces to the Laplacian, while the Gaussian density can be obtained for $c = 2$. The optimal value of c may be different for each DCT coefficient, but for practical purposes a constant value over all frequencies suffices. In [8], the value $c = 0.5$ is used in the proposed generalized Gaussian detector, as it has been found to provide a reasonable model for DCT coefficients at low and mid frequencies [25].

B. Alpha-Stable Modeling of DCT Coefficients

In this paper, we propose the use of the symmetric alpha-stable ($S\alpha S$) family of distributions for the modeling of the heavy-tailed DCT coefficients. The symmetric alpha-stable family has recently attracted considerable interest due both to empirical and theoretical reasons. It has been successful in modeling heavy-tailed data in various applications [24], such as underwater acoustic signals, clutter returns in radar [26], financial data, internet traffic [27], as well as transform domain image or audio signals [28], [29]. The $S\alpha S$ family derives its name from its *stability property*, according to which a linear combination of jointly stable random variables remains stable, i.e., its shape is unchanged (stable) under this transformation. Also, according to the generalized Central Limit Theorem, the *only* possible nontrivial limit of a normalized sum of independent identically distributed (i.i.d.) random variables is stable [30].

The $S\alpha S$ distribution can be best described by its characteristic function

$$\varphi(\omega) = \exp\left(j\delta\omega - \gamma|\omega|^\alpha\right) \tag{11}$$

which is parameterized by the location parameter δ ($-\infty < \delta < \infty$), the scale parameter γ ($\gamma > 0$), known as the dispersion, and the characteristic exponent α ($0 < \alpha \leq 2$). The location parameter δ gives the distribution mean for $1 < \alpha \leq 2$, while for $0 < \alpha \leq 1$ it gives the median. The dispersion γ can be any positive number and corresponds to the spread around the location parameter δ, in other words it behaves like the variance of the Gaussian distribution.

The characteristic exponent or stability index, α, is the most important parameter, as it determines the shape of the distribution. For smaller values of α, the tails of the distribution are heavier and the corresponding random process displays high impulsiveness, while larger values of α correspond to distributions with more rapidly decaying tails that approach the Gaussian distribution. Closed-form expressions exist for the PDF of $S\alpha S$ random variables only for $\alpha = 2$ and $\alpha = 1$, which correspond to the Gaussian and Cauchy distributions, respectively. The shape of the $S\alpha S$ PDF resembles that of the Gaussian density as it is bell shaped, however it has much heavier tails, which asymptotically follow an algebraic rate of decay. For a $S\alpha S$ random variable X, as $x \to \infty$, the tails satisfy

$$P(X > x) \sim c_\alpha x^{-\alpha} \tag{12}$$

where $c_\alpha = \Gamma(\alpha)\sin(\alpha\pi/2)/\pi$, $\Gamma(\cdot)$ is the Gamma function and we use $h(x) \sim g(x)$ as $x \to \infty$ to indicate $\lim_{x\to\infty} h(x)/g(x) = 1$. Hence, the tail probabilities of the $S\alpha S$ random variables follow inverse power laws, which coincide with the tail probabilities of the Pareto distribution. The slowly decaying tail behavior for $\alpha < 2$ is then said to follow "stable Paretian laws," whereas exponentially decreasing tails follow a Gaussian law.

(a) (b) (c)

Fig. 3. Three standard test images with various degrees of detail, used for modeling: (a)"Moon," (b)"Lenna," and (c)"Boat.".

C. Experimental Modeling of DCT Coefficients

In our experiments, we consider each image in the block DCT domain as it was described in Section II, i.e., as a two-dimensional $N_1 \times N_2$ discrete sequence $X[\mathbf{k}]$ of the DCT coefficients. The low and mid-frequency DCT coefficients are the most important ones [21], so the rest can be discarded without significant image distortion. Effective lossy image compression is achieved in this manner, which is one of the main reasons why the DCT transform has been adopted by the JPEG standard [20]. In our modeling, we consider the low and mid frequency DCT coefficients that are watermarked in each 8×8 block as in [8], i.e., the zig zag scanned coefficients 7 to 28 (we use the same group of coefficients as in [8] for reasons of comparison) and examine their distribution over all blocks in each image. We examine a wide range of natural standard test images that are frequently encountered in the literature (cf. Fig. 3), ranging from simple images without many details (e.g., "Moon") to more detailed ones (e.g., "Lenna" and "Boat"), so our modeling results can be considered quite general and reliable.

We proceed in two steps. First, we verify whether the DCT coefficients deviate from the Gaussian distribution. For this purpose, we examine the normal probability (p-p) plots and compare them with the probability plots for the best-fitting $S\alpha S$, the Cauchy, and the generalized Gaussian distributions. P-P plots belong to a family of basic diagnostic tools in statistics, especially selected to be of high-relevance to the study of extreme values and heavy-tailed distributions [31]. They plot probability versus data values. A strong deviation of the p-p plot from the straight diagonal line indicates that the given model is incorrect. Hence, an accurate statistical modeling results in p-p plots where the data follow closely the straight line assumed by the model.

The normal p-p plot for "Lenna," displayed in Fig. 4(a), clearly shows that the data deviates from the normal distribution and it is

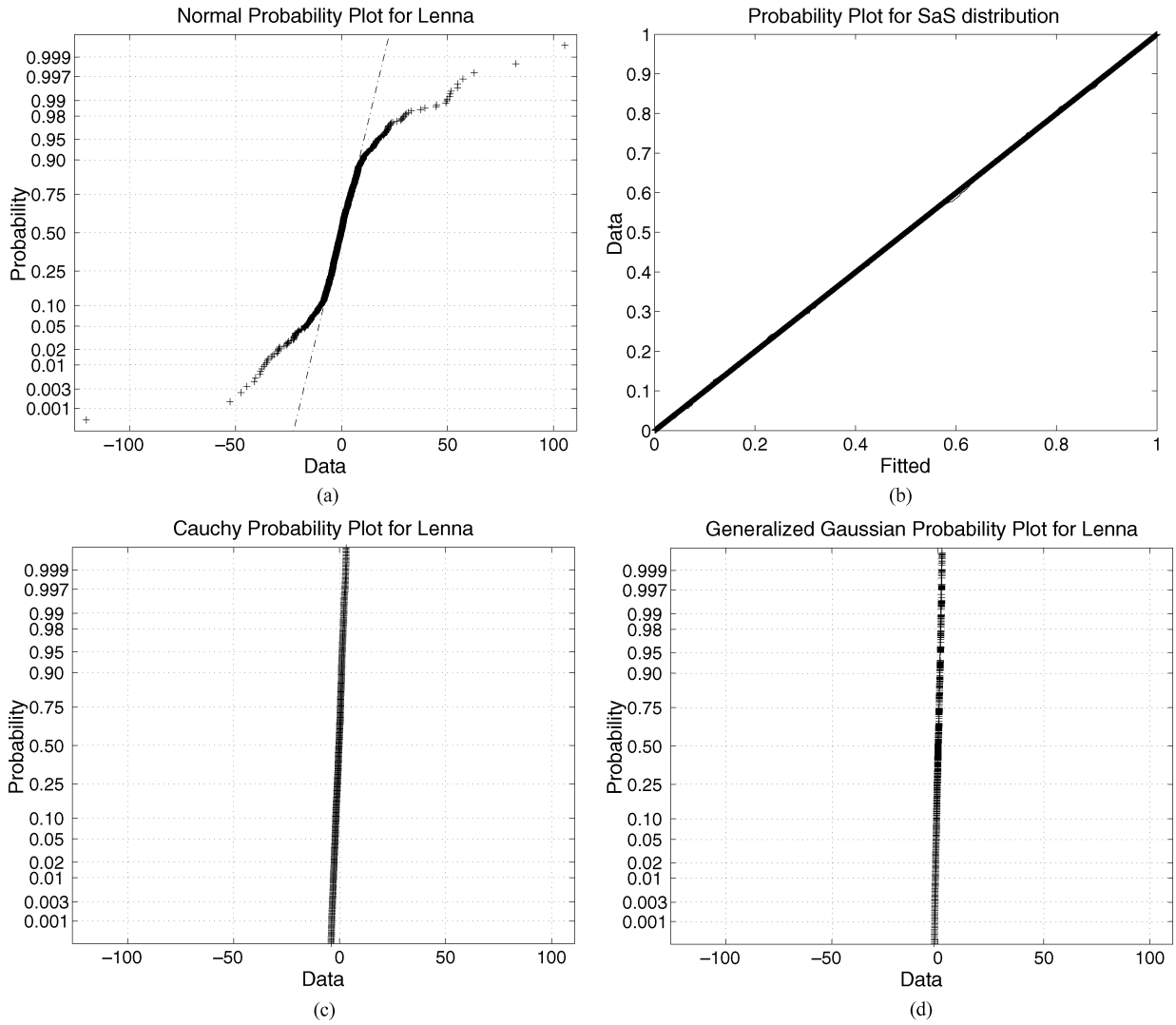

Fig. 4. Probability plots of the "Lenna" image low and mid frequency DCT coefficients for the Gaussian, optimal $S\alpha S$, Cauchy, and generalized Gaussian (GG with $c = 0.5$) distributions. (a) Normal p-p plot. The "+" marks do not follow a straight line so the data is not Gaussian. (b) Stabilized p-p plot for optimal $S\alpha S$ ($\alpha = 1.228$) distribution. The empirical probability density ("+" marks) closely follows the straight $S\alpha S$ line so this model gives a very good fit. (c) Cauchy p-p plot; (d) GG ($c = 0.5$) p-p plot. Note that the "+" marks in (c) and (d) follow a straight line indicating that both the Cauchy and GG distributions are accurate models for the "Lenna" DCT coefficients.

heavy-tailed. The "+" marks in the plot show the empirical probability versus the data value for each point in the sample. The marks are in a curve that does not follow the straight Gaussian line and thus, the normality assumption is violated for this data. The non-Gaussianity of the data does not guarantee, however, that it is stable. Hence, in Fig. 4(b)–(c), we assess the stability of the data. First, the characteristic exponent is estimated and the data sample is fitted with the corresponding $S\alpha S$ distribution. The p-p plots for the optimal $S\alpha S$ and the Cauchy distributions, displayed in Fig. 4(b)–(c), show that the general alpha-stable and even the suboptimal Cauchy models indeed give a very accurate fit. Fig. 4(d) displays the p-p plot for the generalized Gaussian distribution. It is evident that the data can be accurately described by this distribution as well. Experiments performed with the other test images lead to very similar plots, which are not displayed due to space limitations. The p-p plots thus indicate that the alpha-stable family of distributions can describe the low- and mid-frequency DCT coefficients very precisely.

We then estimate the characteristic exponent α to verify that the data is in the stable domain of attraction. For this purpose, we perform maximum likelihood (ML) fitting of the DCT coefficients with $S\alpha S$ distributions and we estimate the corresponding values of α and the associated 95% confidence intervals using the method described by Nolan in [32]. The results are displayed in Table I, where it can be seen that the DCT coef-

TABLE I
CHARACTERIZATION OF ACTUAL DCT COEFFICIENT DATA USING ALPHA-STABLE DISTRIBUTIONS. MAXIMUM LIKELIHOOD PARAMETER ESTIMATES AND 95% CONFIDENCE INTERVALS FOR THE $S\alpha S$ CHARACTERISTIC EXPONENT, α. THE DEVIATIONS OF α FROM THE VALUE $\alpha = 2$ INDICATE THAT THE DATA IS NON-GAUSSIAN

IMAGE	α	IMAGE	α
Baboon	1.4021 $\pm$ 0.0207	Frog	1.6199 $\pm$ 0.0095
Barbara	0.9434 $\pm$ 0.0082	Goldhill	1.3672 $\pm$ 0.0203
Bird	1.1511 $\pm$ 0.0186	Lenna	1.2280 $\pm$ 0.0193
Boat	0.9030 $\pm$ 0.0160	Moon	1.2060 $\pm$ 0.0114
Cameraman	0.7990 $\pm$ 0.0148	Mountain	1.2230 $\pm$ 0.0089
Clown	0.8650 $\pm$ 0.0199	Zelda	1.3285 $\pm$ 0.0100

ficients of the examined images exhibit distinctly non-Gaussian characteristics, with values of α varying between 0.8 and 1.6, away from the Gaussian point of $\alpha = 2$. It is safe to assume that the heavy-tailed nature of the data must definitely be taken into account by the watermark detector in order to ensure reliable detection performance. We should note here that the association of the $S\alpha S$ (or any other) model parameters, namely α and γ, to the underlying physical process that generates the image is a fundamental issue. Such a consideration is beyond the scope of this manuscript as it requires an in-depth study of the mechanism that generates a particular image type, e.g., ultrasound, radar, optical, hyperspectral, etc., followed by the relation of the statistical model parameters to the so-called ground truth of the image.

Naturally, a question of interest is whether the stable fit describes the data more accurately than other density functions proposed in the literature. Here, we compare the $S\alpha S$ fits with those provided by the generalized Gaussian density function. The modeling of the DCT coefficients will be used as a basis for the development of an improved watermark detector, so we consider distributions that will actually be employed for the detector's design. Specifically, we model the samples with the generalized Gaussian distribution having parameter $c = 0.5$ that was picked in [8] to be a good choice for a fixed, image-independent parameter. On the other hand, since the Cauchy ($\alpha = 1$) is the only member of the $S\alpha S$ family (apart from the Gaussian) with a closed-form expression for its PDF, we perform the stable modeling using the Cauchy, as a robust detector can be designed in a straightforward manner for this distribution.

To model the DCT coefficients, we use the amplitude probability density (APD) function, $P(|X| > x)$. The empirical APD can be plotted directly from the data by simply counting the DCT coefficients X for which $|X| > x$. The APD can also be evaluated analytically for each density (e.g., the Cauchy or the generalized Gaussian) after estimating the corresponding distribution parameters from the data. Due to space limitations, in Fig. 5 we present the APD modeling results for six representative test images. It should be noted, however, that similar results were obtained for the other images of Table I as well. It is clear that the Laplacian fit is poor, since the Laplacian tails decay at a fast rate. Comparing the Cauchy and generalized Gaussian curves, we observe that they both provide fairly satisfying goodness of fit for most images. In some images, the Cauchy APD is the better model, in other images the generalized Gaussian seems to prevail. For example, note that the Cauchy density provides a better fit to the mode of the "Cameraman" and "Moon" images (cf. Fig. 5(b) and (e), respectively), the generalized Gaussian PDF is a more accurate fit to the mode of "Bird" [cf. Fig. 5(a)], and the two distributions are equally accurate near the mode, in the other cases.

More interestingly from a detection perspective, Fig. 5 reveals that the actual data tails are almost always heavier than what the generalized Gaussian model assumes, although not as heavy as the tails of the Cauchy APD. In practice, this fact indicates that the alpha-stable family will account for high-magnitude values in the data tails, which the generalized Gaussian distribution misses, as reported in [8]. Thus, one can design an improved detector based on the Cauchy assumption, which will account for high-amplitude data values without resorting to *ad-hoc* techniques such as eliminating high-magnitude samples, as proposed in [8]. Of course, dealing with real images, we see in Fig. 5(f) that for "Boat" the tails of the empirical distribution are closer to the tails of the generalized Gaussian APD than to those of the Cauchy APD. Not surprisingly, this is the only case where the generalized Gaussian processor gives better detection results than the Cauchy, as we will see in Section V.

IV. Statistical Watermark Detection

Blind watermark detectors can be designed based on the binary hypothesis test presented in Section II-B, as long as the data probability distribution is known or can be estimated from the data. Naturally, these detectors are optimal when the data is accurately modeled. In [8], a detector based on the generalized Gaussian model was proposed, as this distribution fairly accurately characterizes DCT or wavelet image coefficients [14]. Motivated by our modeling results in the previous section, we design a watermark detector based on the Cauchy distribution. The Cauchy detector better accounts for the high-amplitude DCT coefficient values in which the watermark is hidden and it is shown to have improved performance for a wide range of watermark-to-document power levels and for a large number of test images.

A. The Cauchy Watermark Detector

A blind watermark detector that considers a $S\alpha S$ model for the data can be easily designed only for the Cauchy distribution (i.e., for $\alpha = 1$), the only member of the $S\alpha S$ family with a closed-form expression for the PDF. Although the Cauchy detector is theoretically suboptimal for non-Cauchy data, we will demonstrate that it performs extremely well in real life applications. In fact, Cauchy-based detectors and estimators were shown to be particularly robust in heavy-tailed environments even under signal mismatch, contrary to traditional detectors [15], [33].

The density function of the Cauchy distribution is given in closed form by the expression

$$f_X(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x - \delta)^2} \tag{13}$$

where γ is the data dispersion and δ is the location parameter. In this case, the likelihood ratio of (7) uses the probability density functions

$$\begin{aligned} f(Y[\mathbf{k}] \mid H_0) &= f_X(Y[\mathbf{k}]) \\ f(Y[\mathbf{k}] \mid H_1) &= f_X(Y[\mathbf{k}] - W[\mathbf{k}]). \end{aligned} \tag{14}$$

For simplicity of notation, in the sequel we will omit the 2-D DCT domain index $\mathbf{k}$ from $X[\mathbf{k}]$ and $Y[\mathbf{k}]$. If the set of DCT coefficients that contain a watermark is denoted as S, the likelihood ratio becomes

$$\begin{aligned} \frac{f(Y \mid H_1)}{f(Y \mid H_0)} &= \prod_S \frac{f_X(Y - W)}{f_X(Y)} \\ &= \prod_S \frac{\gamma^2 + (Y - \delta)^2}{\gamma^2 + (Y - W - \delta)^2}. \end{aligned}$$

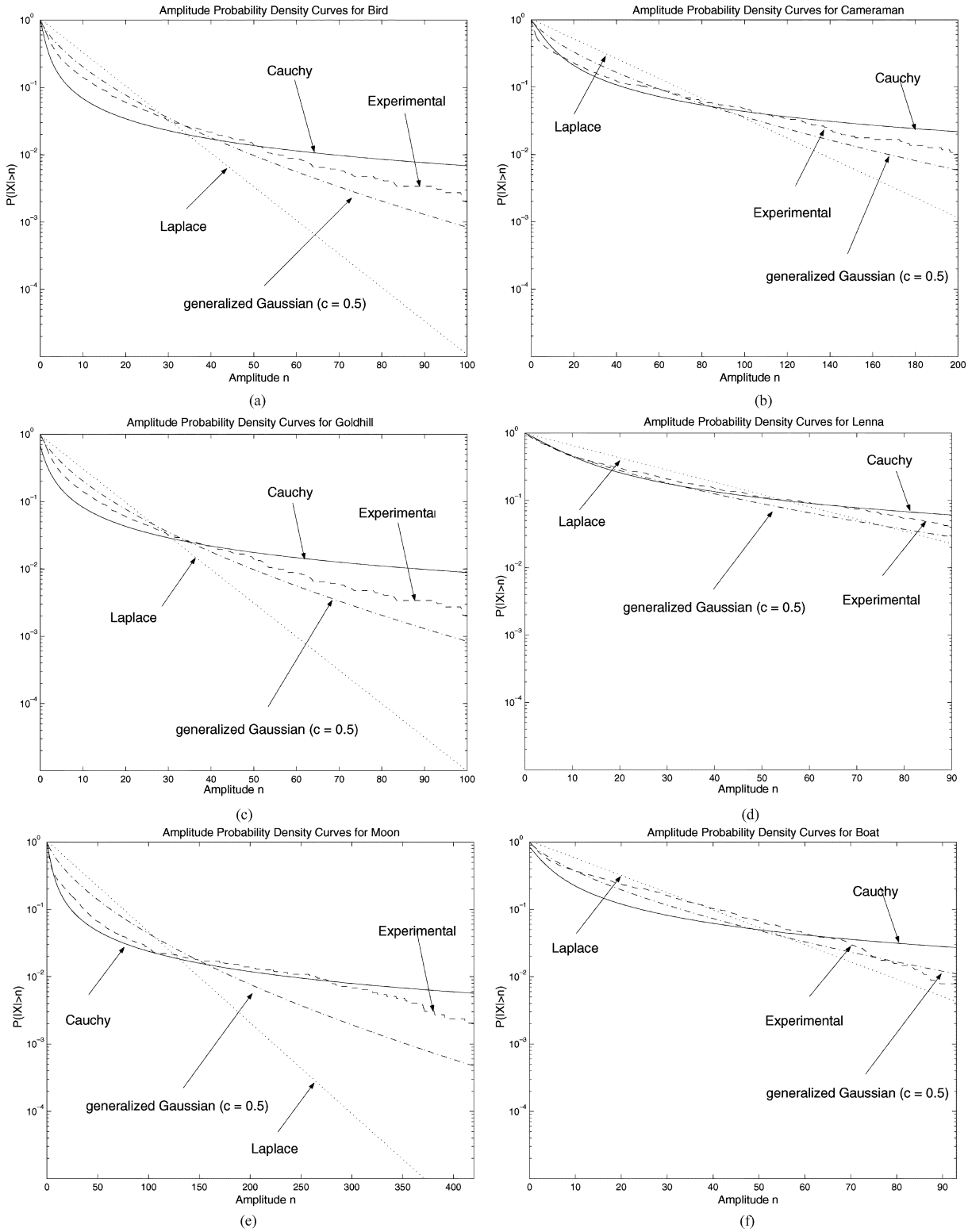

Fig. 5. APD curves $P(|X| > n)$ for the low and mid frequency DCT coefficients of the standard test images: (a) "Bird," (b) "Cameraman," (c) "Goldhill," (d) "Lenna," (e) "Moon," and (f) "Boat." The distributions of interest are the Laplacian, Cauchy, and generalized Gaussian with parameter $c = 0.5$. In most cases, the modes are accurately captured by the Cauchy and the generalized Gaussian model and the empirical tails lie between the tails of the two distributions.

The corresponding log-likelihood ratio is

$$l(Y) = \sum_{S} \ln \frac{\gamma^2 + (Y - \delta)^2}{\gamma^2 + (Y - W - \delta)^2}. \tag{15}$$

This analytical expression of the likelihood ratio permits the theoretical measurement and experimental verification of the Cauchy detector performance, as the receiver operating characteristic (ROC) curves can be derived from it. The ratio $l(Y)$ in (15) is a sum of statistically independent random variables, so

according to the CLT, it should follow a Gaussian distribution for an appropriately large number of samples [15]. Therefore, we consider that it follows an approximately Gaussian distribution with mean and variance under hypotheses H_1 and H_0 that can be estimated from the given data, as shall be done in the sequel.

In order to simplify the theoretical analysis, some realistic assumptions are made concerning the watermark, as in [8]. In the frequently considered case that the watermark does not contain hidden information, the codeword is $b = 1$. In this case, the watermark will be equal to $W[\mathbf{k}] = a[\mathbf{k}]s[\mathbf{k}]$, where $a[\mathbf{k}]$ is the perceptual mask and $s[\mathbf{k}]$ the output of the PRS generator. In our experiments, the sequence $s[\mathbf{k}]$ is assumed to be an i.i.d. two-dimensional random sequence with two equiprobable values $+1$, -1, so the watermark will be equal to $a[\mathbf{k}]$ or $-a[\mathbf{k}]$ at every $\mathbf{k}$-pixel. Thus, the mean value of the log-likelihood function (15) can be computed assuming that under H_0 there is no watermark, while under H_1 the watermark at each pixel can be either $a[\mathbf{k}]$ or $-a[\mathbf{k}]$, each with probability 1/2. Then, the mean of the log-likelihood ratio under H_0, having omitted the index $\mathbf{k}$, becomes as in (16), shown at the bottom of the page. The variance of the log-likelihood function under H_0 is

$$\sigma_0^2 = Var[l(Y) \mid H_0] = E[l(Y) - E[l(Y)] \mid H_0]^2 \tag{17}$$

where under H_0 we have (18), shown at the bottom of the page. Using the above result, the variance becomes as shown in (19), at the bottom of the page. In order to simplify the notation, we temporarily set

$$\begin{aligned} x_1 &= \gamma^2 + (X - a - \delta)^2 \\ x_2 &= \gamma^2 + (X + a - \delta)^2 \\ x_3 &= \gamma^2 + (X - W - \delta)^2. \end{aligned} \tag{20}$$

Assuming that the watermark $W[\mathbf{k}]$ is an i.i.d. random sequence, the variance is

$$\begin{aligned} \sigma_0^2 &= E\left[\sum_S \left(\frac{1}{2}\ln(x_1x_2) - \ln x_3\right)^2\right] \\ &= E\left[\sum_S \left(\frac{1}{4}(\ln(x_1x_2))^2 + (\ln x_3)^2 - \ln(x_1x_2)\ln x_3\right)\right] \\ &= \sum_S \left(\frac{1}{4}(\ln x_1)^2 + \frac{1}{4}(\ln x_2)^2 - \frac{1}{2}\ln x_1 \ln x_2\right) \\ &= \frac{1}{4}\sum_S \left(\ln \frac{x_1}{x_2}\right)^2. \end{aligned} \tag{21}$$

Substituting x_1 and x_2, we get the final expression for the variance of the likelihood ratio

$$\sigma_0^2 = \frac{1}{4}\sum_S \left(\ln \frac{\gamma^2 + (X - a - \delta)^2}{\gamma^2 + (X + a - \delta)^2}\right)^2. \tag{22}$$

$$\begin{aligned} m_0 &= E[l(Y) \mid H_0] \\ &= \sum_S \left(\ln[\gamma^2 + (X - \delta)^2] - \frac{1}{2}\ln[\gamma^2 + (X - a - \delta)^2] - \frac{1}{2}\ln[\gamma^2 + (X + a - \delta)^2]\right) \\ &= \sum_S \ln \frac{\gamma^2 + (X - \delta)^2}{[(\gamma^2 + (X - a - \delta)^2)(\gamma^2 + (X + a - \delta)^2)]^{1/2}}. \end{aligned} \tag{16}$$

$$\begin{aligned} &l(Y) - E[l(Y) \mid H_0] \\ &\quad = \sum_S \left(\ln[\gamma^2 + (X - \delta)^2] - \ln[\gamma^2 + (X - W - \delta)^2] - \ln[\gamma^2 + (X - \delta)^2] + \frac{1}{2}\ln[(\gamma^2 + (X - a - \delta)^2)(\gamma^2 + (X + a - \delta)^2)]\right) \\ &\quad = \sum_S \ln \frac{(\gamma^2 + (X - \alpha - \delta)^2)^{1/2}(\gamma^2 + (X + a - \delta)^2)^{1/2}}{\gamma^2 + (X - W - \delta)^2}. \end{aligned} \tag{18}$$

$$\begin{aligned} \sigma_0^2 &= E\left[\sum_S \ln \frac{(\gamma^2 + (X - a - \delta)^2)^{1/2}(\gamma^2 + (X + a - \delta)^2)^{1/2}}{\gamma^2 + (X - W - \delta)^2}\right]^2 \\ &= E\left[\sum_S \left(\frac{1}{2}\ln[(\gamma^2 + (X - a - \delta)^2)(\gamma^2 + (X + a - \delta)^2)] - \ln[\gamma^2 + (X - W - \delta)^2]\right)\right]^2. \end{aligned} \tag{19}$$

The likelihood ratio mean under H_1 is similarly computed by substituting Y with $X+W$ in (15)

$$m_1 = E[l(Y) \mid H_1] = E\left[\sum_S \ln \frac{\gamma^2 + (X+W-\delta)^2}{\gamma^2 + (X-\delta)^2}\right] = \sum_S \ln \frac{(\gamma^2+(X+a-\delta)^2)^{1/2}(\gamma^2+(X-a-\delta)^2)^{1/2}}{\gamma^2+(X-\delta)^2}. \tag{23}$$

Finally, following the same procedure as before, the variance of the log-likelihood ratio under H_1 is found to be $\sigma_1^2 = \sigma_0^2$. Thus, the log-likelihood ratio $l(Y)$ under H_0 and H_1 can be considered as a Gaussian random variable with mean and variance

$$\begin{aligned} m_0 &= -m_1 \\ \sigma_0^2 &= \sigma_1^2. \end{aligned} \tag{24}$$

These quantities can be computed from the given data, i.e., the DCT coefficients of the image, since they only depend on the data values X and a, and the $S\alpha S$ distribution parameters γ and δ, which are estimated from the data. Thus, with the mean (m_0, m_1) and variance (σ_0^2, σ_1^2) of the log-likelihood ratio known under both hypotheses, the probabilities of false alarm and detection can be estimated from

$$P_{\text{fa}} = Q\left(\frac{t-m_0}{\sigma_1}\right) \tag{25}$$

$$P_{\text{det}} = Q\left(\frac{t-m_1}{\sigma_1}\right) \tag{26}$$

where t is the threshold against which the data are compared and $Q(x)$ is defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-z^2/2} dt. \tag{27}$$

We consider Neyman–Pearson testing [11], where the threshold and consequently the detection probability is found for a predetermined P_{fa}. Then, the threshold is given by the expression [11]

$$t = m_0 + \sigma_1 Q^{-1}(P_{\text{fa}}). \tag{28}$$

By replacing this threshold in (26) and by taking into account that $m_1 = -m_0$, we get

$$P_{\text{det}} = Q\left(Q^{-1}(P_{\text{fa}}) - 2\sqrt{\frac{m_1^2}{\sigma_1^2}}\right). \tag{29}$$

Finally, defining $SNR \triangleq m_1^2/\sigma_1^2$ as in [8], [11] and substituting it in (29), we obtain the following relation between P_{fa} and P_{det}:

$$P_{\text{det}} = Q\left(Q^{-1}(P_{\text{fa}}) - 2\sqrt{\text{SNR}}\right). \tag{30}$$

From the above equation, we conclude that the ROC curve depends only on the value of the SNR for a fixed value of P_{fa}. Obviously, larger values of SNR improve the detector performance, as they will lead to a higher probability of detection for a given P_{fa}.

The proposed Cauchy detector will be compared with the generalized Gaussian detector with $c = 0.5$ whose likelihood ratio was derived in [8]

$$l(Y) = \sum \beta^c (|Y|^c - |Y - a \cdot s|^c) \tag{31}$$

and its mean and variance under H_0 are given by

$$m_0 = -\frac{1}{2}\sum \beta^c \Big(|X+a|^c - |X-a|^c\Big) - \sum \beta^c |X|^c \tag{32}$$

$$\sigma_0^2 = \frac{1}{4}\sum \beta^{2c}\Big(|X+a|^c - |X-a|^c\Big)^2 \tag{33}$$

while under H_1, we have $m_1 = -m_0$ and $\sigma_1^2 = \sigma_0^2$. Using the above expressions, one can form the ROCs for the generalized Gaussian detector and compare them to the ones of the Cauchy processor.

V. Experimental Results

In order to reliably measure the actual performance of the detector proposed in Section IV, we conduct experiments with various test images that were also used for modeling. Due to limitations of space, we present results for the "Bird," "Cameraman," "Goldhill," "Lenna," "Moon," and "Boat" images, which exhibit various degrees of detail.

We initially verify the analysis of Section IV for the statistics of the two tests by calculating the mean and variance of the generalized Gaussian and the Cauchy log-likelihood ratios both theoretically and empirically. Then, the actual watermark detectors are implemented to measure the performance of the two schemes in terms of detection and false alarm probabilities not only theoretically, but also experimentally. Using the expressions of the previous section for the theoretical detection and false alarm probabilities, as well as their experimental values, we plot and compare the theoretical and empirical ROC curves for the two systems. Finally, we examine the performance of each processor for invisible watermarks of varying strength. This comparison provides significant insight into the robustness of the two detectors as it shows how their performance is affected by the embedding of weaker or stronger watermarks.

A. Experimental Values of the Likelihood Ratio Mean and Variance

In our experiments, the watermarks used are generated and embedded following the procedure described in Section II. A perceptual mask $a[\mathbf{k}]$ is obtained for each image according to (3) to ensure that the DCT coefficients suffer the maximum admissible alteration, while the watermark remains invisible to the Human Visual System. This mask is multiplied by 10^4 independently generated pseudorandom sequences $s[\mathbf{k}]$ to generate the random watermarks that are necessary for the Monte Carlo experiments. At every Monte Carlo iteration, we estimate the likelihood ratio $l(Y)$ for the watermarked (H_1) and unwatermarked (H_0) DCT coefficients of each image for both detectors using expressions (15) and (31). The experimental mean and variance of $l(Y)$ is then estimated from the Monte Carlo simulations for each detector and for both hypotheses. The theoretical mean and variance of the likelihood ratio for the Cauchy distribution can

TABLE II
EMPIRICAL AND THEORETICAL VALUES FOR THE MEAN AND VARIANCE OF THE LOG-LIKELIHOOD RATIO UNDER HYPOTHESES H_0 AND H_1 FOR THE CAUCHY AND GENERALIZED GAUSSIAN ($c = 0.5$) DISTRIBUTIONS. NOTE THAT THESE QUANTITIES ARE VERY ACCURATELY COMPUTED ANALYTICALLY

IMAGE	Cauchy m_0	Cauchy σ_0^2	gen. G. m_0	gen. G. σ_0^2
Barbara (th.)	-31.0020	26.3989	-31.0441	30.0209
Barbara (emp.)	-30.7452	29.1460	-30.9135	29.8023
Bird (th.)	-0.6189	0.0623	-0.4599	0.0465
Bird (emp.)	-0.6246	0.0601	-0.4703	0.0458
Boat (th.)	-15.5161	63.1796	-4.9946	4.3438
Boat (emp.)	-15.1175	63.7118	-4.9025	4.2889
Cam. (th.)	-0.5547	0.0669	-0.4124	0.0498
Cam. (emp.)	-0.5412	0.0689	-0.4014	0.0520
Frog (th.)	-19.6874	59.3666	-22.5266	151.9735
Frog (emp.)	-19.5998	59.2362	-22.4808	152.7100
Goldhill (th.)	-5.4702	11.4815	-5.6660	14.9378
Goldhill (emp.)	-5.4230	11.6786	-5.4423	15.2034
Lenna (th.)	-15.5608	36.1617	-18.8175	79.7416
Lenna (emp.)	-15.6835	34.4485	-19.0608	79.5472
Moon (th.)	-15.0867	35.7252	-14.5007	45.0807
Moon (emp.)	-14.7604	35.2401	-14.3597	45.1504

be computed directly from the data via (16) and (22), and for the generalized Gaussian detector via (32) and (33).

The results from these experiments, presented in Table II, justify the assumption that $l(Y)$ follows a Gaussian distribution and show that its mean and variance can be very accurately computed analytically for both detectors. As a consequence, the experimental estimates of the SNR are also exact, since they depend only on m_1 and σ_1^2. Since the ROC curves depend only on the SNR ((30)), the experimental ROC curves are also expected to be very close to the theoretical ones. The performance of the generalized Gaussian and the Cauchy detectors can thus be predicted before actual experiments are conducted, by estimating the SNR experimentally. Table III shows a few characteristic values of the SNR that were derived for the examined test images. In most cases the Cauchy detector leads to a higher SNR, so it is expected to exhibit better performance.

B. Detector Performance in Terms of Receiver Operating Characteristic Curves

The theoretical ROC curves are derived for the generalized Gaussian and the proposed Cauchy detector using (32), (33) and (16), (22), respectively, where $m_1 = -m_0$ and $\sigma_1^2 = \sigma_0^2$. The probability of false alarm is set to the range 10^{-4} to 10^{-1} and the ROCs are derived from (30). The experimental ROC curves are found by measuring the performance of the actual watermark detection systems over all Monte Carlo runs. For each DCT-transformed image, we design a generalized Gaussian and a Cauchy detector as detailed in Section IV. Experiments are then conducted by comparing the likelihood ratio of each processor with the corresponding threshold for each value of the false alarm probability and for 10^5 randomly generated watermarks. If the likelihood ratio is above the threshold under H_1, the watermark is detected, while if it is above the threshold under H_0, a false alarm occurs. In this manner, the experimental false alarm and detection probabilities and the corresponding empirical ROC curves are derived. Note that the SNRs for which we plot the ROC curves are in the range 6 to 9 dB (as shown in Table III). It is possible to embed watermarks that lead to higher values of the SNR, such as 16 dB (for "Bird") or 29 dB (for "Lenna" in [8]), which are still invisible, but for these values the detection probability is equal to 1 for both detection schemes for the whole range of P_{fa}, so it is meaningless to plot the ROCs.

TABLE III
$SNR = 10\log(m_1^2/\sigma_1^2)$(IN DB) FOR THE CAUCHY AND GENERALIZED GAUSSIAN ($c = 0.5$) DISTRIBUTIONS. HIGHER VALUES OF THE SNR CORRESPOND TO IMPROVED DETECTION PERFORMANCE

SNR (in dB)		
IMAGE	$Cauchy$	$Gen.Gaussian(c = 0.5)$
Barbara	15.6113	15.0654
Bird	7.8875	6.5787
Boat	5.8099	7.5914
Cameraman	6.7048	5.3341
Frog	3.3736	5.2362
Goldhill	4.1603	3.3128
Lenna	8.2581	6.4744
Moon	8.0421	6.6869

Fig. 6(a)–(f) clearly show that the empirically obtained ROC curves verify the theoretical ones for both detectors. The ROCs in Fig. 6(a)–(e) indicate that the Cauchy detector outperforms the generalized Gaussian detector. This is in agreement with the modeling results of Section III-C, where it was shown that for these images the $S\alpha S$ model accounts better for the heavy data tails. Fig. 6(f) shows a case, the "Boat" image, where the generalized Gaussian detector performs better than the Cauchy processor. For this image, the generalized Gaussian distribution provided a closer fit to the DCT coefficients than the $S\alpha S$ model, as shown in Fig. 5(f). This kind of behavior is to be expected, since each image has different characteristics, as it was also noted in [8], where the Laplacian detector led to better detection results than the generalized Gaussian scheme for the "Tiger" image.

C. Detector Performance for Watermarks of Varying Strength

The performance of the two detection schemes is also compared when watermarks of varying strength are embedded in a given set of DCT coefficients. For this purpose, we define the "Watermark to Document Ratio" (WDR) as in [34]

$$\text{WDR} = 10\log\left(\frac{\sigma_w^2}{\sigma_x^2}\right) \tag{34}$$

with

$$\sigma_w^2 = \frac{1}{N}\sum_k W[\mathbf{k}]^2, \quad \sigma_x^2 = \frac{1}{N}\sum_k X[\mathbf{k}]^2 \tag{35}$$

where N is the number of data samples and "document" refers to the original host data, i.e., the unwatermarked DCT image coefficients. Watermarks of varying strength $W[\mathbf{k}]$ are embedded in the DCT coefficients, leading to different values of

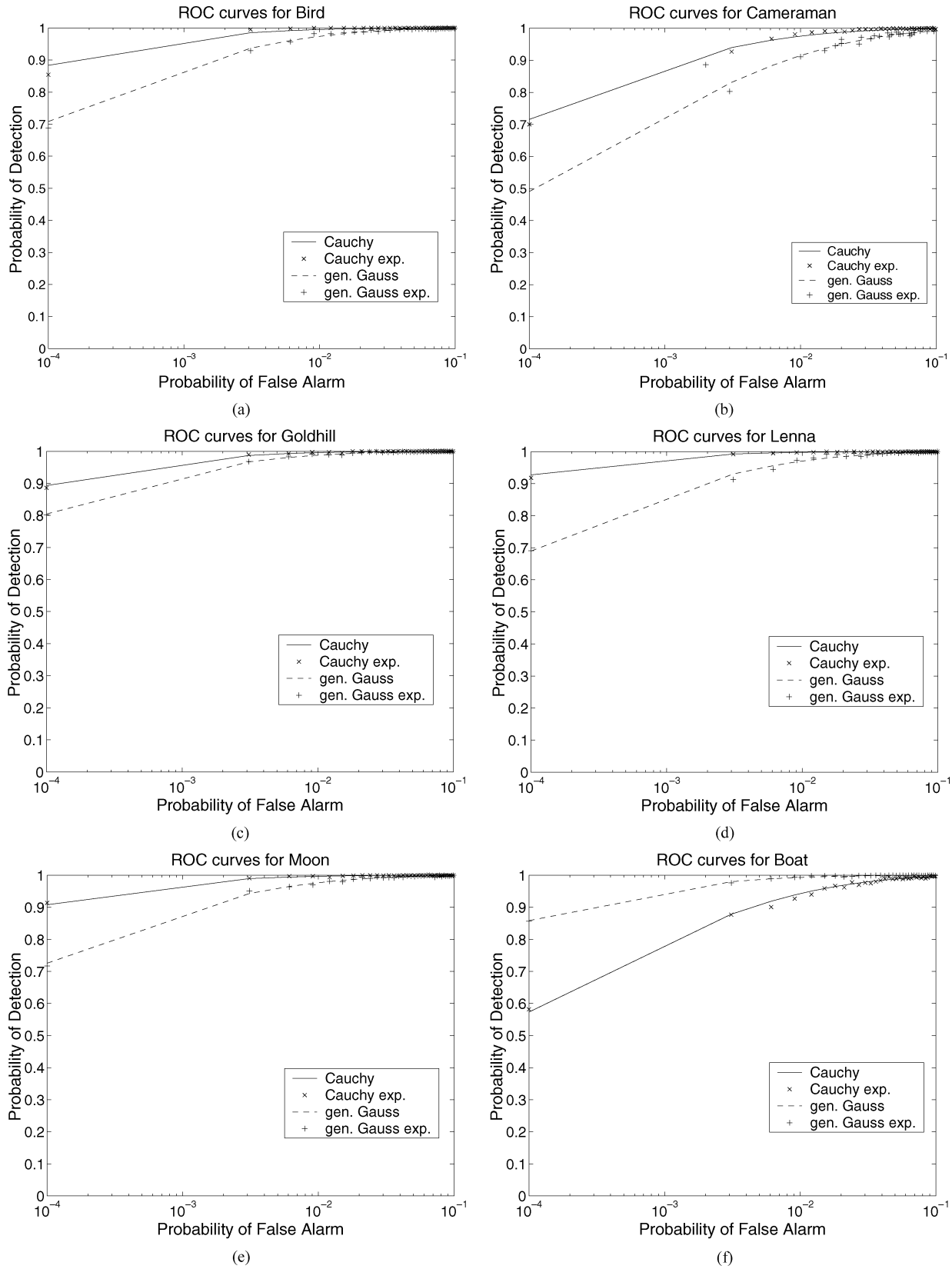

Fig. 6. Empirical and theoretical ROC curves for low- and mid-frequency DCT coefficients of the test images for the Cauchy (solid line) and generalized Gaussian ($c = 0.5$) (dashed line) detectors. The Cauchy detector has better performance in all cases except the "Boat" image (f), as it accounts for the high magnitude samples hidden in the tails, which the generalized Gaussian detector misses.

σ_w^2 and WDR. Naturally, the watermark strength is increased while taking into account the properties of the HVS, in order to remain invisible. Obviously, its power can be decreased to the point where it is still detectable.

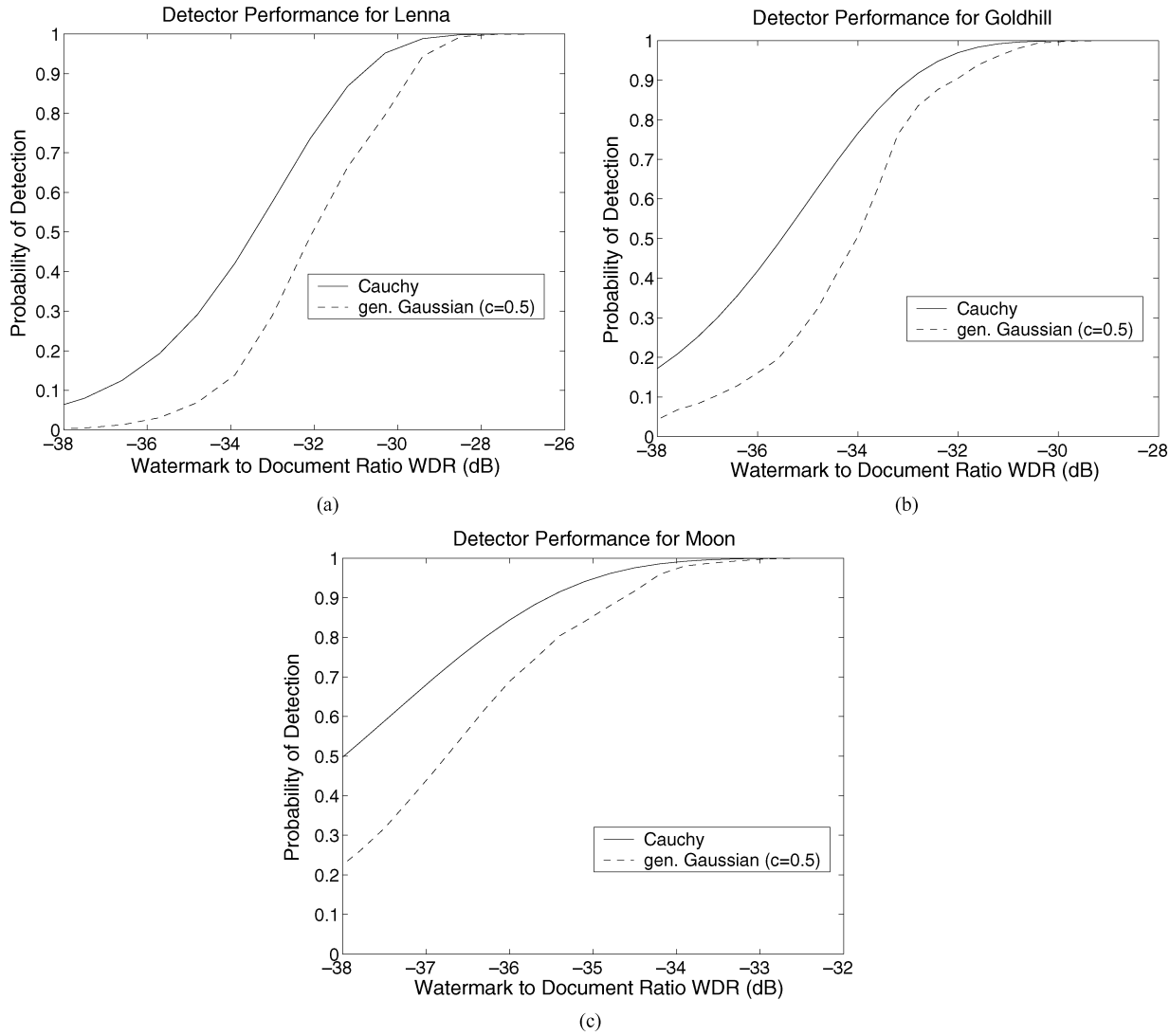

Fig. 7. Probability of detection for watermarks of varying strength parameterized by the Watermark to Document Ratio $\mathrm{WDR} = 10\log(\sigma_w^2/\sigma_x^2)$ (in decibels) for the Cauchy (solid line) and generalized Gaussian ($c = 0.5$) (dashed line) detectors. The false alarm probability is set to $P_{\mathrm{fa}} = 10^{-4}$. (a) "Lenna." (b) "Goldhill." (c) "Moon."

For three of our test images, namely "Goldhill," "Lenna," and "Moon", we embed watermarks for WDR values ranging from $-38\,\mathrm{dB}$ to $-24\,\mathrm{dB}$. For these watermarks, we determine the statistics of the two detectors, as in Section IV. We then consider two cases: first, we fix the probability of false alarm and we estimate the corresponding probabilities of detection; then, we fix the probability of detection and we find the corresponding probabilities of false alarm. In the former case, we set the false alarm probability to the reasonable value $P_{\mathrm{fa}} = 10^{-4}$ and in the latter case we consider $P_{\mathrm{det}} = 0.95$. From the statistics of the likelihood ratios for the two detectors and the predetermined P_{fa} and P_{det}, we then estimate the corresponding thresholds for the likelihood ratio tests. For fixed P_{fa} we use (28) and for a fixed P_{det} the threshold for each detector can be immediately derived from (26)

$$t = m_1 + \sigma_1 Q^{-1}(P_{\mathrm{det}}). \tag{36}$$

Finally, from (26) and (25), we find the probabilities of detection and false alarm for all the WDR values and we plot them in Figs. 7 and 8, respectively.

Note in Fig. 7 that, depending on the particular image, a WDR value greater than $-28\,\mathrm{dB}$ (for "Lenna") or $-33\,\mathrm{dB}$ (for "Moon") corresponds to a relatively "strong" watermark that can be detected with probability equal to 1 by both detectors. Also, note that the performance of each detector depends not only on the WDR, but on the specific characteristics of the image as well. For example, comparing Fig. 7(a) and (c), we observe that a stronger watermark must be embedded in a detailed image such as "Lenna" in order to be reliably detected, whereas in a simple image such as "Moon" a weaker (by about 4 dB) watermark is needed to obtain the same probability of detection. This result makes intuitive sense, since it is easier to detect a signal hidden in a smooth image than in an image with many details. Similarly, in Fig. 8, it can be seen that more false alarms tend to occur in the more detailed images for the same P_{det} and WDR. This can also be explained intuitively, since in a smooth and simple image there are not as many areas with a lot of information that may be mistaken for a watermark. Consequently, the false alarm probability in simple images is expected to be lower than that in more complicated ones.

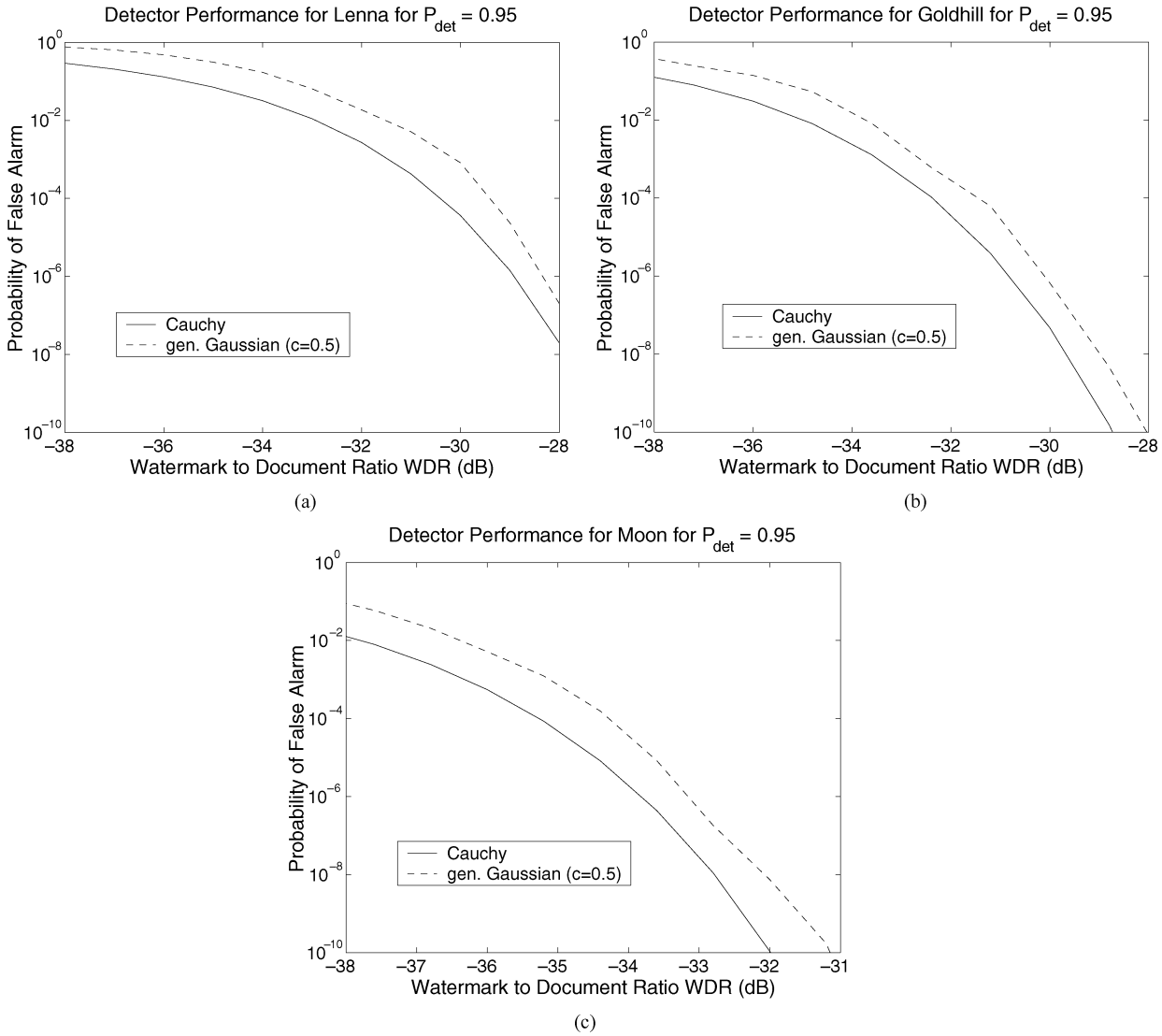

Fig. 8. Probability of false alarm for watermarks of varying strength parameterized by the Watermark to Document Ratio $\mathrm{WDR} = 10\log(\sigma_w^2/\sigma_x^2)$ (in decibels) for the Cauchy (solid line) and generalized Gaussian ($c = 0.5$) (dashed line) detectors. The detection probability is set to $P_{\mathrm{det}} = 0.95$. (a) "Lenna." (b) "Goldhill." (c) "Moon."

Figs. 7 and 8 show that for reasonable values of P_{fa} and P_{det} and for (invisible) watermarks of varying strength, the performance of the Cauchy detector is better than that of the generalized Gaussian. Indeed, as the power of the hidden signal is decreased (WDR decreases), the performance of both schemes deteriorates, but the Cauchy processor outperforms the generalized Gaussian. Thus, the Cauchy detector can reliably detect a watermark in low WDR margins, i.e., even if the watermark has very low strength or it is subject to an attack that partially eliminates it, thus decreasing its power σ_w^2.

VI. Conclusions and Future Work

In this paper, we proposed and analyzed an improved blind watermark detector for DCT-transformed images based on the statistical properties of the DCT coefficients. The new detector considers the Cauchy statistical model, a member of the heavy-tailed $S\alpha S$ family, which accurately describes the DCT coefficients of interest. The accuracy of this model was verified through experimental modeling of the data, which showed that it can better account for the heavy-tailed DCT coefficients than the generalized Gaussian model. Then, the performances of the Cauchy and the generalized Gaussian detectors were compared in terms of their corresponding ROC curves. Experimental results with real images verified the theoretical analysis and provided insight into how the performance of the two detectors is affected by embedding watermarks of varying strength.

Our current work involves the design of a maximum likelihood watermark decoder based on the $S\alpha S$ modeling of the DCT coefficients. This decoder could actually be the second part of a watermark verification and extraction system that also includes a watermark detector. Considering the improved performance of the proposed Cauchy detector, such a decoder is also expected to give good results, thus leading to the design of an integrated watermark verification and extraction system with better performance than the currently existing schemes.

REFERENCES

[1] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proc. IEEE*, vol. 87, no. 7, pp. 1079–1107, Jul. 1999.

[2] J. A. Bloom, I. J. Cox, T. Kalker, J.-P. Linnartz, M. G. Miller, and C. B. S. Traw, "Copy protection for DVD video," *Proc. IEEE*, vol. 87, no. 7, pp. 1267–1276, Jul. 1999.

[3] R. B. Wolfgang, C. I. Podilchuk, and E. J. Delp, "Perceptual watermarks for digital images," *Proc. IEEE*, vol. 87, no. 7, pp. 1108–1126, Jul. 1999.

[4] I. J. Cox, M. L. Miller, and A. McKellips, "Watermarking as communications with side information," *Proc. IEEE*, vol. 87, no. 7, pp. 1127–1141, Jul. 1999.

[5] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," in *Proc. 2000 IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 6, Istanbul, Turkey, 2000, pp. 3630–3633.

[6] V. Capellini, M. Barni, F. Bartolini, and A. Piva, "A DCT-domain system for robust watermarking," *Signal Processing*, vol. 66, pp. 357–372, 1998.

[7] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Robust data hiding for images," in *Proc. IEEE Digital Signal Processing Workshop*, Loen, Norway, 1996, pp. 37–40.

[8] J. R. Hernandez, M. Amado, and F. Perez-Gonzalez, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Trans. Image Process.*, vol. 9, pp. 55–68, 2000.

[9] W. Zeng and B. Liu, "A statistical watermark detection technique without using original images for resolving rightful ownerships of digital images," *IEEE Trans. Image Process.*, vol. 8, pp. 1534–1548, Nov. 1999.

[10] G. Voyatzis and I. Pitas, "The use of watermarks in the protection of digital multimedia products," *Proc. IEEE*, vol. 87, pp. 1197–1207, Jul. 1999.

[11] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.

[12] J. R. Hernandez and F. Perez-Gonzalez, "Statistical analysis of watermarking schemes for copyright protection of still images," *Proc. IEEE*, vol. 87, no. 7, pp. 1142–1166, Jul. 1999.

[13] K. A. Birney and T. R. Fischer, "On the modeling of DCT and subband image data for compression," *IEEE Trans. Image Process.*, vol. 4, pp. 186–193, 1995.

[14] R. C. Reininger and J. D. Gibson, "Distributions of the two-dimensional DCT coefficients for images," *IEEE Trans. Commun.*, vol. COMM-31, pp. 835–839, 1983.

[15] G. A. Tsihrintzis and C. L. Nikias, "Performance of optimum and suboptimum receivers in the presence of impulsive noise modeled as an alpha-stable process," *IEEE Trans. Commun.*, vol. 43, no. 3, pp. 904–914, Mar. 1995.

[16] E. Sayrol, J. Vidal, S. Cabanillas, and S. Santamaria, "Optimum watermark detection in color images," in *Proc. 1999 Int. Conf. Image Processing (ICIP 99)*, vol. 2, 1999, pp. 231–235.

[17] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum perceptual watermarking for images, audio and video," *IEEE Trans. Image Process.*, vol. 6, pp. 1673–1687, 1997.

[18] F. Hartung, J. K. Su, and B. Girod, "Spread spectrum watermarking: Malicious attacks and counterattacks," in *Proc. SPIE Security and Watermarking of Multimedia Contents 99*, San Jose, CA, 1999.

[19] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1995.

[20] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Trans. Consumer Electron.*, vol. 38, no. 1, pp. 18–34, 1992.

[21] M. Nelson, *The Data Compression Book*. San Mateo, CA: M&T Books, 1992.

[22] J. A. Solomon, A. B. Watson, and A. J. Ahumada, "Visibility of DCT basis functions:Effects of contrast masking," in *Proc. Data Compression Conf.*, Snowbird, UT, 1994, pp. 361–370.

[23] A. J. Ahumada and H. A. Peterson, "Luminance-model-based DCT quantization for image compression," in *Proc. SPIE Human Vision, Visual Processing, and Digital Display III*, vol. 1666, 1992, pp. 365–374.

[24] R. Adler, R. Feldman, and M. S. Taqqu, *A Guide to Heavy Tails: Statistical Techniques and Applications*. Boston, MA: Birkhauser, 1998.

[25] R. J. Clarke, *Transform Coding of Images*. New York: Academic, 1985.

[26] G. A. Tsihrintzis and C. L. Nikias, "Advanced Statistical Signal Processing for Radar Clutter Environments," Univ. Southern California, Los Angeles, Tech. Rep. USC-SIPI, 1994.

[27] W. Willinger, M. S. Taqqu, W. E. Leland, and D. V. Wilson, "Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements," *Stat. Sci.*, vol. 10, pp. 67–85, 1995.

[28] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Med. Imag.*, vol. 20, pp. 772–783, Aug. 2001.

[29] P. Georgiou, P. Tsakalides, and C. Kyriakakis, "Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise," *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 291–301, Sep. 1999.

[30] M. Shao and C. L. Nikias, "On Symmetric Stable Models for Impulsive Noise," Univ. Southern California, Los Angeles, Tech. Rep. USC-SIPI-231, 1993.

[31] R. D. Reiss and M. Thomas, *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Berlin, Germany: Birkhauser, 1997.

[32] J. P. Nolan, ""Maximum Likelihood Estimation and Diagnostics for Stable Distributions," Tech. Rep.," Dept. of Math. and Stat., Amer. Univ., 1999.

[33] P. Tsakalides, "Array Signal Processing with Alpha-Stable Distributions," Ph.D. dissertation, Univ. Southern California, Los Angeles, CA, 1995.

[34] J. J. Eggers and B. Girod, "Quantization effects on digital watermarks," *Signal Process.*, vol. 81, no. 3, 2001.

Alexia Briassouli received the Diploma in electrical engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 1999 and the M.Sc. degree in image and signal processing from the University of Patras, Patras, Greece, in 2000. She is currently pursuing the Ph.D. degree in electrical engineering at the University of Illinois at Urbana-Champaign.

From 2000 to 2001, she was a Research Assistant at the Informatics and Telematics Institute, Center of Research and Technology Hellas (CERTH), Thessaloniki, Greece, participating in a European-funded research project. Her research interests lie in the fields of statistical signal and image processing. She has worked on the design of optimal watermark embedding and detection systems for images and video that are robust to various attacks. Her current research interests lie in the areas of statistical image processing and computer vision, and include problems like motion estimation and segmentation for video

Panagiotis Tsakalides (M'95) received the Ph.D. degree in electrical engineering from the University of Southern California (USC) in 1995 and the Diploma in Electrical Engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1990.

He is an Associate Professor of Computer Science at the University of Crete, Crete, Greece, and a Researcher with the Institute of Computer Science, Foundation for Research and Technology-Hellas (ICS-FORTH), Hellas, Greece. From 1999 to 2002, he was with the Department of Electrical Engineering, University of Patras, Patras, Greece. From 1996 to 1998, he was a Research Assistant Professor with the Signal and Image Processing Institute, USC, and he consulted for the U.S. Navy and Air Force. His research interests lie in the field of statistical signal processing with emphasis in non-Gaussian estimation and detection theory, and applications in wireless communications, imaging, and multimedia systems. He has co-authored over 50 technical publications in these areas, including 15 journal papers.

Dr. Tsakalides was awarded IEE's A. H. Reeve Premium in 2002 for the paper (co-authored with P. Reveliotis and C. L. Nikias) "Scalar quantization of heavy-tailed signals," published in the October 2000 issue of the *IEE Proceedings – Vision, Image and Signal Processing*.

Thanos Stouraitis (SM'97) received the B.S. degree in physics from the University of Athens, Athens, Greece, in 1979, the M.S. degree in electronic automation from the University of Athens, Athens, Greece, in 1981, the M.Sc. degree in electrical and computer engineering from the University of Cincinnati, Cincinnati, OH, in 1983, and the Ph.D. degree in electrical engineering from the University of Florida, Gainesville, in 1986 (for which he received the Outstanding Ph.D. Dissertation Award).

He is a Professor of Electrical and Computer Engineering at the University of Patras, Patras, Greece, where he directs the International Graduate Studies Program on Digital Signal Processing Systems. He is also a member of the Administrative Committee of the University of Sterea Hellas. He has served on the faculty of The Ohio State University and has visited the University of Florida and Polytechnic University. His current research interests include signal and image processing systems, application-specific processor technology and design, computer arithmetic, and design and architecture of optimal digital systems. He has led several DSP processor design projects funded by the European Union, American organizations, and the Greek government and industry and has served as a Consultant for various industries. He has authored and co-authored over 140 technical papers. He holds one patent on DSP processor design. He has authored *Digital Signal Processing* (Patras, Greece: University of Patras Press) and co-authored *Digital Filter Design Software for the IBM PC* (New York: Marcel Dekker), in addition to several book chapters.

Dr. Stouraitis serves as the Regional Editor for Europe for the *Journal of Circuits, Systems, and Computers*, has served as Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, Associate Editor for the IEEE TRANSACTIONS ON VERY LARGE SCALE INFORMATION SYSTEMS, Editor for the IEEE Interactive Magazines, an Editor-at-Large for Marcel Dekker Inc., and Associate Editor for the *Journal of Circuits, Systems and Computers*. He regularly reviews for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, IEEE TRANSACTIONS ON COMPUTERS, and IEEE TRANSACTIONS ON EDUCATION, for *IEE Proceedings* (Parts E, F, and G), and for conferences like IEEE ISCAS, ICASSP, VLSI, SiPS, Computer Arithmetic, Euro DAC, etc. He also reviews proposals for the National Science Foundaiton, the European Commission, and other agencies. He has served as the Chair of the VLSI Systems and Applications (VSA) Technical Committee and as a member of the DSP and the Multimedia Technical Committees of the IEEE Circuits and Systems (CAS) Society. He is a founder and the Chair of the IEEE Signal Processing chapter in Greece. He was the General Chair of the 1996 IEEE International Conference on Electronics, Circuits, and Systems (ICECS) and the Technical Program Chair of Eusipco '98 and ICECS '99. He has served as Chair or as a member of the Technical Program Committees of a multitude of IEEE Conferences, including ISCAS (Program Committee Track Chair). He is the General Chair of ISCAS 2006. He received the 2000 IEEE CAS Society Guillemin-Cauer Award (best paper) for the paper "Multi-Function Architectures for RNS Processors."