Optimal adaptive diversity watermarking with channel state estimation

Sviatoslav Voloshynovskiy, Frédéric Deguillaume, Shelby Pereira and Thierry Pun

University of Geneva - CUI, 24 rue du Général Dufour, CH 1211, Geneva 4, Switzerland

ABSTRACT

This work advocates the formulation of digital watermarking as a communication problem. We consider watermarking as communication with side information available for both encoder and decoder. A generalized watermarking channel is considered that includes geometrical attacks, fading and additive non-Gaussian noise. The optimal encoding/decoding scenario is discussed for the generalized watermarking channel.

Keywords: Digital watermarking, diversity, channel state estimation, adaptive receiver, fading

1. INTRODUCTION

Digital watermarking emerged as a tool for copyright protection, document authentication, access control and tamper proofing. As the most cited application, watermarking is used for copyright protection where watermark robustness, invisibility and sufficient informative capacity are simultaneously required. Additionally, watermark decoding should be oblivious. To compromize between these requirements special watermark encoding and perceptual masking are used.

Watermark replication, as encoding, and perceptual masking at embedding are known to drastically increase robustness with respect to various image distortions, but present several disadvantages from the encoding efficiency viewpoint. First, such information encoding can be considered as a simple repetition code, known to be an inefficient form of channel coding. Secondly, fading due to both adaptive perceptual masking, and to possible attacks, may considerably decrease the final system performance. Thirdly, the design of a watermarking system should compromise between resistance to cropping attack and code rate. This practically requires spreading of the watermark over the whole image to allow for its recovery. A compromise has therefore to be found between redundancy and coding efficiency. To solve this compromise in an optimal manner, possible solutions are to use either diversity techniques, or refined encoding approaches such as trellis-coded modulation (TCM). In this paper, we consider the first solution with an interperiod optimal signal encoding based on iterative codes such as turbo codes or low-density parity check codes (LPDC), with the decoder performing watermark channel state estimation based on a reference pilot watermark. At embedding, we use binary phase shift keying (BPSK).

Knowingly or not, almost all replication-based watermarking algorithms rely on standard linear diversity combining techniques. However, approaches presented so far are not effective in the combining of information at the decoder stage. Their inefficiency stems from not accounting for the actual distortions introduced in the watermarking channel. The typical assumption about the additive white Gaussian noise (AWGN) character of the distortion is not valid when the sampling space (the number of replications) is not sufficiently large to satisfy the requirements of the Central Limit Theorem. The optimal decoder for AWGN involves summation of all observations taken from different periods of the watermark repetition. In this case, even a single large impulse noise sample can dominate these sums and defeat the averaging. In fact, increasing the number of watermark repetitions might not increase performance in realistic watermarking scenarios if linear combining is used, but simply increase the probability that an impulse noise is observed in one of the periods. Therefore, the optimal receiver should be designed for the general case of a non-Gaussian channel.

State-of-the-art communication systems in the weak signal case require exact knowledge of the noise distribution to derive closed form expressions for such optimal receiver. Unfortunately, the lack of knowledge of the noise distribution makes it impossible to design an optimal decoder for attacks or distortions that have different probabilistic non-Gaussian models.

Correspondence (S. Voloshynovskiy): Email: svolos@cui.unige.ch; WWW: http://cui.unige.ch/~vision

Therefore, rather than designing a decoder for a particular noise distribution, we propose the following adaptive approach. Estimation of the noise distribution after attack, or channel state estimation, are performed based on a key-dependent reference pilot watermark that is known to the decoder. The generalization of consideration of watermarking as communication with side information is performed in Section 3. The attacking channel and message extraction are considered in Section 4, and Section 5. Two different noise models are analyzed in Section 6 : the stationary Generalized Gaussian model and the nonstationary Gaussian model. Results, shown in Section 7, illustrate various aspects of this adaptive watermarking decoding and confirm its performance.

2. MODERN DIGITAL WATERMARKING PARADIGM

Consider the general model of a watermarking system according to a communications formulation. Its block diagram is shown in Figure 1.

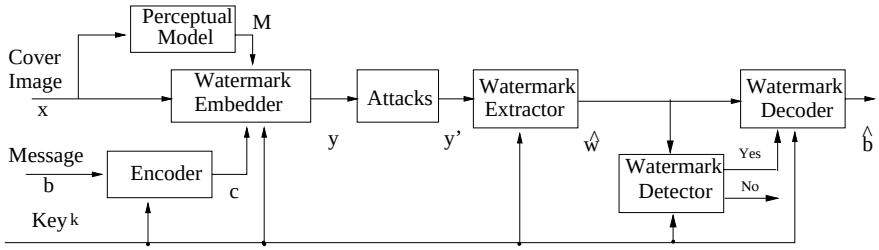


Figure 1. Communication formulation of a watermarking system.

The watermarking system consists of three main parts, i.e. message embedding, attack channel and message extraction. Let us consider in details these main parts.

3. MESSAGE EMBEDDING: COMMUNICATION WITH SIDE INFORMATION

A message $b = (b_1, ..., b_L)$ is to be embedded in the cover image $x = (x_1, ..., x_N)^T$ of size $M_1 \times M_2$, where $N = M_1 \cdot M_2$. The message b contains information about the owner and can be used for authentication purposes. To convert the message into a form efficient for communication, it is encoded using either error correction codes (ECC) or modulated using binary antipodal signaling[1] or M-ary modulation.[2] With respect to ECC, mostly Bose Chaudhuri (BCH) or convolutional codes are used.[3,4] Recent publications[5,6] report successful results using novel Turbo codes and low-density parity-check (LDPC) codes in the DCT and wavelet domains. In the general case, the type of ECC and the set of basis functions for M-ary modulation can be key-dependent. The above conversion is performed in the encoder that produces the codewords $c = Enc(b, Key)$, $c = (c_1, ..., c_K)^T$ which are mapped from {0,1} to {-1,1} using binary phase shift keying (BPSK).

A watermark w is created by some key-dependent function $w = \varepsilon(c, p, M, Key)$ that ensures the necessary spatial allocation of the watermark based on a key-dependent projection function p, and according to HVS features as expressed by a perceptual mask M in order to improve the watermark. The typical choice for the projection function p is a set of two dimensional orthogonal functions used for every codeword bit $\{c_k\}$ such that the empty set is formed by the intersection $P_k \cap P_l, \forall k \neq l$.[2,1] The projection function performs a "spreading" of the data over the image area. It can be also considered as diversity communication problem with parallel channels. Moreover, the projection function can have a particular spatial structure with given correlation properties that can be used for the recovery of affine geometrical transformations.[2,6] The resulting watermark is obtained as the superposition

$$w(j) = \sum_{k=1}^{K} c_k p_k(j) M(j) \tag{1}$$

where $j \in Z$. The watermark embedder performs the insertion of the watermark into the cover image in some transform or coordinate domain, yielding the stego image:

$$y = T^{-1}[h(T[x], w)] \tag{2}$$

where T is any orthogonal transform like block DCT, full-frame FFT and DCT, wavelet or Radon transforms ($T = I$ for the coordinate domain), and $h(.,.)$ denotes the embedding function. The most widely used class of embedding functions conforms to the linear additive model

$$y = h(x, w|M) = x + w(M) \tag{3}$$

that is considered in this paper.

To extend the above model in a more general formulation, one can consider the watermarking as a communication with side information (SI) as is shown in Figure 2.

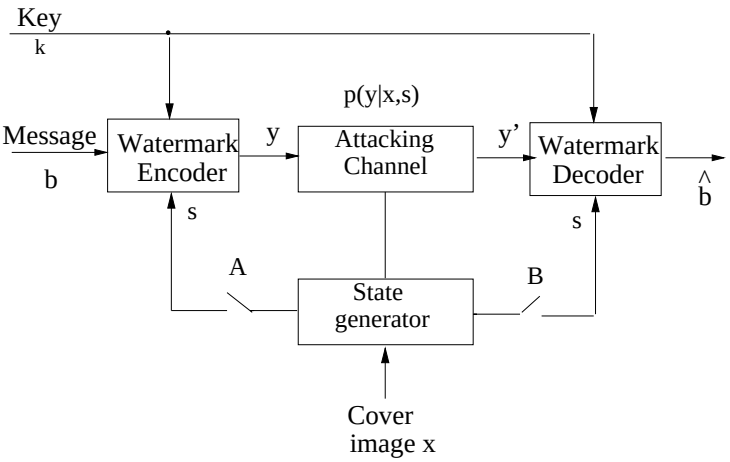


Figure 2. Watermarking as communication with side information.

Let us consider the side information in watermarking applications assuming that the encoder and decoder have access to several items. First, both the encoder and the decoder can access the key used for the watermark embedding. Second, the generalized channel state information can be available. The generalized channel includes the cover data and the attacking channel. The attacking channel includes all possible intentional or unintentional attacks that can be applied during the "life cycle" of the image or videos.

Depending on different combinations of the switches A and B in Figure 2, all watermarking algorithms can be divided into the four classes described below. It is assumed that switch A controls the access of the encoder to the channel state information given by the cover image. Switch B controls the access of the decoder to the attacking channel information given by all possible signal processing and geometrical attacks. The key k is assumed to be available for both encoder and decoder in private watermarking.

Class I: SI is not available (switches A and B open). It is a typical case for all earlier watermarking algorithms that were inspired by the original papers of Cox[7] and Tirkel.[8] It assumes that the watermark is embedded in the cover image and is then decoded without reference to information about channel state. The detection of watermarks is mostly based on the direct correlation of the stego data with the watermark generated based on the key. If the correlation coefficient is above some threshold, then the decision is made of successful detection. As a result, the performance of these schemes is very poor due to two basic assumptions made: all attacks are modeled as additive stationary Gaussian noise that results in the simple correlation detection receiver.[9,10] Secondly, these schemes assume no geometrical attacks.

Class II: SI is available at encoder only (A closed, B open). This scheme has found recently a lot of attention in the watermarking community due to the publication of Cox.[11] The block diagram of this scheme is shown in Figure 3.

The main idea of watermarking as communication with SI at the encoder consists of the fact that the theoretical capacity of oblivious watermarking scheme is equal to that of a decoder with access to the cover data. This conclusion is based on the remarkable paper of Costa.[12] Therefore, there is no more need of the cover data for the decoder, if the cover data is used as SI by the encoder. However, this approach has several drawbacks. First, the complexity of

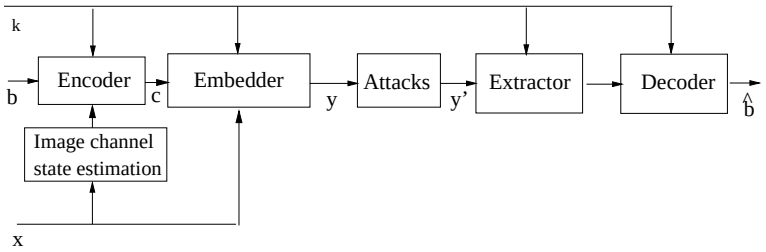


Figure 3. Watermarking as communication with side information regarding cover data available for the encoder.

the encoder is very high. This means that the codebook for every particular image becomes quite large. Therefore, the decoder should also perform a quite complex search. Second, the watermarking channel is only treated as the cover image and attacks are not taken into account that can lead to a missmatch between what was assumed for the design of the codebook and the real situation.

To reduce the complexity of the encoder, different practical algorithms were proposed.[13,14,5] However, geometrical attacks and the attacks whose statistics are different from additive Gaussian remain an open issue.

To relax the lack of decoder adaptivity with respect to the attacking channel state, one can include the worst case attack as information about channel state in the decoder assuming that the decoding will be successful in more favorable conditions. An example of this approach aiming at resisting against low lossy JPEG compression is proposed by Pereira *et al.*[5] The JPEG quantization table for the worst case of Stirmark compression at quality factor QF=10 was included in the design of the encoder.

Class III: SI is available at decoder only (A open, B closed). These schemes are able to estimate the undergone attacks in the attacking channel and are potentially able to resist against geometrical transformations. This relies on the fact the a key-dependent pilot or reference watermark can be used for two purposes. First, the pilot can be considered as the synchronization pattern in some coordinate or transform domain, i.e. mostly in the magnitude spectrum of the DFT due to the known shift and cropping invariant properties, as well as with the simultaneous ability to detect affine transforms.[2,6] Secondly, the pilot embedded in the stego data can be used to estimate fading due to data embedding and attacks, and statistics of noise, if they are different from Gaussian as is the case with a lossy JPEG compression attack. This enables to consider the watermarking as a channel with fading and non-Gaussian noise and leads to diversity reception since the watermark is replicated over the image area. The pilot can be easily regenerated in the decoder based on the key.

Class IV: SI is available at both encoder and decoder (A closed, B closed). This scenario can be considered as the most likely scheme for all future watermarking algorithms that can operate under a wide class of uncertainties with respect to the channel state. The optimality of this scheme is based on the optimal design of the encoder matched with the cover data and adaptivity of the decoder to the attacking channel state assuming fading, non-Gaussian attacks and geometrical transforms, therefore utilizing the advantages of diversity watermarking. The generalized block diagram of this scheme is shown in Figure 4.

4. ATTACKING CHANNEL

An attacking channel produces a distorted version y' of the stego image y. The attacking channel can be modeled in the framework of stochastic formulation using a probability mass function (p.m.f) $Q(y'|y)$ to describe random distortions in the stego image. A successful attack should damage or destroy the watermark while preserving the commercial quality of the image. Therefore, an attacker should introduce distortions that are limited by some upper allowable bound according to the chosen distortion criterion. Although, the MSE is not perfectly matched with the subjective human assessment of image quality, it is commonly used due to the obtained tractable results, and the wide usage of this criteria in the communication community due to the known results for the additive Gaussian channels. Therefore, the aim of the attacker consists in decreasing of the rate of reliable communication subject to the allowable distortion.

However, it is necessary to note that the above consideration will not be complete without geometrical attacks. The geometrical attacks can be mathematically modeled as affine transforms with some random parameters that

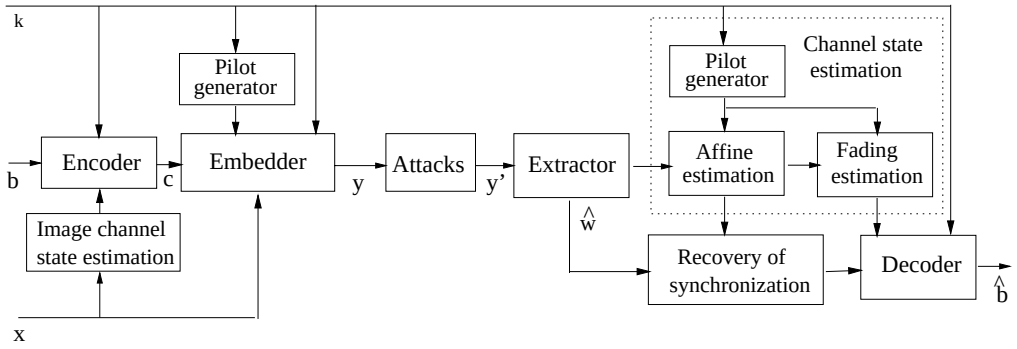


Figure 4. Watermarking as diversity communication with side information about the cover data.

are not known for the decoder. Normally, there are 6 parameters that produce all set of global affine geometrical alterations: scaling, change of aspect ratio, shearing, rotation and shift. More generally, these modifications can be modeled as projective transformations that can occur in the applications such as the "Internet bridge" of Digimark, i.e. reading the watermark in front of a web camera. The random local distortions integrated in the Stirmark benchmark and also known as random bending attack can be modelled by local affine transforms with additive Gaussian noise arising from the interpolation. Therefore, the decoder should access these parameters to have an optimally synchronized watermark decoding. The concept of pilot or reference watermark considered above might be an appropriate solution of this problem both for the affine parameters estimation and for estimation of $Q\left(y'|y\right)$ parameters.

5. MESSAGE EXTRACTION

The recovery process consists of the watermark extractor and decoder which are described below.

5.1. Watermark extractor for oblivious watermarking

The watermark extractor performs an estimate $\widehat{w}$ of the watermark based on the attacked version $\widehat{y}$ of the stego-image:

$$\widehat{w} = Extr\left(T\left[y'\right], Key\right) \tag{4}$$

In the general case, the extraction should be key-dependent. However, the desire to recover data after affine transformation based on the above mentioned self-reference principle, and the opportunity to enhance the decoding performance by reducing the variance of the image considered as noise,[2,15] have motivated the development of key-independent watermark extraction methods. They could represent the main danger to linear additive watermarking technologies, as will be shown below.

Different methods are used for watermark estimation, such as the cross-shaped filter,[2] or MMSE estimates.[1] In the most general case, the problem of watermark estimation can be solved based on a stochastic framework by using Maximum Likelihood (ML), penalized ML, MAP or Minimum Description Length (MDL) estimates.[16] Assuming that both the noise due to the cover image and the noise introduced by an attack can be considered additive with some target distribution $p_X\left(.\right)$, one can determine the ML-estimate:

$$\widehat{w} = \arg\max_{\widetilde{w}\in\Re^N} p_X\left(y' \mid \widetilde{w}\right) \tag{5}$$

which results either in a local average predictor/estimator in the case of a locally stationary independent identically distributed (i.i.d.) Gaussian model of $p_X\left(.\right)$, or a median predictor in case of a corresponding Laplacian p.d.f.. If there is some prior information about watermark statistics, the MAP estimate can be used:

$$\widehat{w} = \arg\max_{\widetilde{w}\in\Re^N}\left\{ p_X\left(y' \mid \widetilde{w}\right) \cdot p_W\left(\widetilde{w}\right)\right\} \tag{6}$$

where $p_W\left(.\right)$ is the p.d.f. of the watermark. To solve problems (5) and (6) it is necessary to develop accurate stochastic models for the cover image $p_X\left(x\right)$ and the watermark $p_W\left(w\right)$.

5.1.1. Stochastic models of cover image: source generation

Stochastic models of cover image applied to content adaptive watermarking were considered in our previous work.[16] We use here the main results of this work and consider either locally i.i.d. non-stationary Gaussian (nG) or globally i.i.d. Generalized Gaussian (sGG) image models. The motivation for these two models are their wide usage in a number of image processing applications including image denoising, restoration and compression, and the existence of tractable closed form solutions of (6) for the particular cases of these models.

The non-stationary Gaussian model is characterized by a distribution:

$$p_x(x) = \frac{1}{(2\pi)^{\frac{N}{2}}} \cdot \frac{1}{|\det R_x|^{\frac{1}{2}}} \cdot \exp\{-\frac{1}{2}(Cx)^T R_x^{-1} Cx\}, \tag{7}$$

where R_x is covariance matrix, $|\det R_x|$ denotes the matrix determinant, and Cx represents a high-pass filtering (decomposition operator) which can be also rewritten as $Cx = (I - A)x = x - Ax = x - \bar{x}$, where I is the unitary matrix, A is a low-pass filter used to compute the non-stationary local mean $\bar{x}$. C could also be considered as a wavelet decomposition operator in which case model (7) is used for every subband.

The stationary GG model has stationary R_x and can be written as:

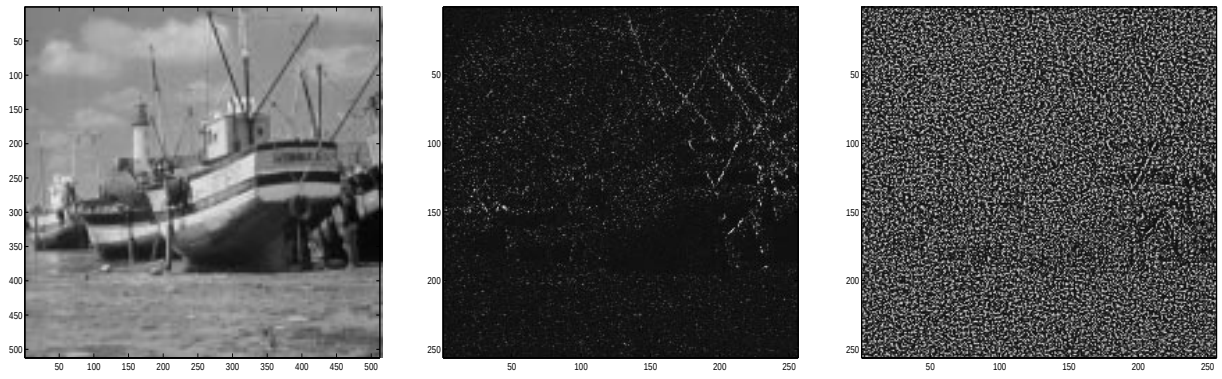
$$p_x(x) = (\frac{\gamma\eta(\gamma)}{2, (\frac{1}{\gamma})})^{\frac{N}{2}} \cdot \frac{1}{|\det R_x|^{\frac{1}{2}}} \cdot \exp\{-\eta(\gamma)(|Cx|^{\frac{\gamma}{2}})^T R_x^{-\frac{\gamma}{2}} |Cx|^{\frac{\gamma}{2}}\}, \tag{8}$$

where $\eta(\gamma) = \sqrt{\frac{, (\frac{3}{\gamma})}{, (\frac{1}{\gamma})}}$ and $, (t) = \int_0^\infty e^{-u} u^{t-1} du$ is the gamma function, and the parameter γ is called the *shape parameter*. Equation (8) includes the Gaussian ($\gamma = 2$) and the Laplacian ($\gamma = 1$) models as special cases. For real images the shape parameter is in the range $0.3 \leq \gamma \leq 1$. The other examples of stationary stochastic models are mixture models that either include two additive Gaussian distributions with different variances, i.e. the increased variance is used to model heavy tails in the distribution, or include Gaussian and Laplacian p.d.fs. Cauchy distributions can be also used to approximate the heavy tail statistics.

There is a strict connection between local non-stationary Gaussian and global stationary Generalized Gaussian models. If we consider the image locally, then it could be accurately modeled by the non-stationary Gaussian model, while treating the same data globally as i.i.d with the same variance one can approximate it using stationary GG model for a particular γ. To show this connection we will consider an example in the wavelet domain; the consideration is also valid for coordinate domain modeling and for multichannel DCT based image representations used in current JPEG compression standard.

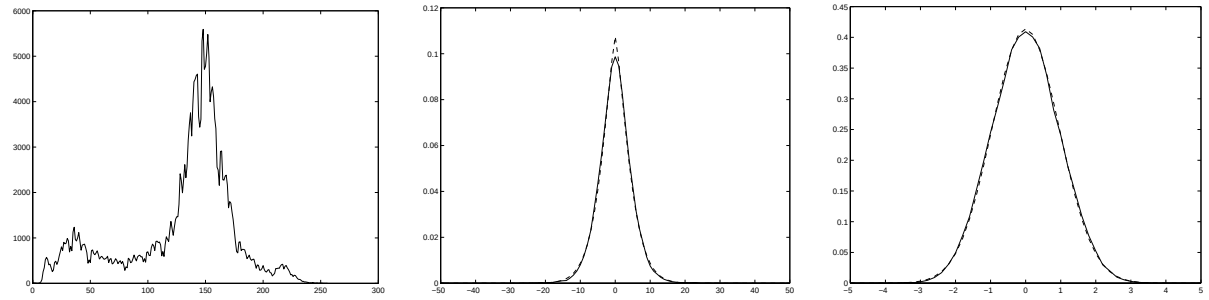
The original Boat image (Figure 5a) is decomposed using a wavelet transform. The first scale coefficients for the diagonal orientation sub-band are shown in Figure 5b. It is necessary to note that the same results can be obtain using a Laplacian image decomposition pyramid or simply by subtracting the local image mean estimated in a window of size 5x5 that will aproximate the Laplacian operator. The above image has non-stationary character, i.e. the regions of edges and textures are more visible and have larger amplitude due to the edge transitions. To normalize the image, i.e. to make its distribution close to normal or Gaussian ($N(0,1)$), we divide it by the estimate of the local standard deviation; this results in the image shown in Figure 5c, which has a more uniform character. Assuming that the image coefficients are stationary, i.e. originate from the same distribution, we plot the corresponding histogram of the images from Figure 5 that are depicted in Figure 6. It is possible to follow the changes in the histograms statistics starting from multimodal (Figure 6a) that can be modeled as a mixture of Gaussian, to unimodals (Figure 6b and c) that are quite accurately approximated using stationary Generalized Gaussian and stationary Gaussian models, respectively.

This simple experiment makes it possible to establish the practically important dependencies between different stochastic models and to formulate an uniform stochastic framework for image modeling. Based on that, the image can be treated as a multichannel stochastic process. Using the inverse order of the above decomposition of the image into zero-mean unit variance Gaussian noise this multichannel model can be presented as in Figure 7. First, each pixel of the image is modeled as a stationary source with $N(0,1)$. Secondly, each pixel is multiplied by the non-stationary standard deviation and biased by the non-stationary mean, resulting in the final observed image. Therefore, considering each pixel locally it can be presented as non-stationary mean non-stationary variance Gaussian



(a) (b) (c)

Figure 5. a) The original Boat image; b) the result of decomposition; c) the normilized decompozed image.



(a) (b) (c)

Figure 6. The histograms of: a) the original Boat image, b) the decomposed and c) the normalized images and their approximation by stationary Generalized Gaussian $sGG(0, 0.6, 17)$ and zero-mean unit variance Gaussian models $N(0, 1)$ respectively.

model. At the same time treating all coefficients globally, i.e. originating from the same i.i.d. source that is represented as a multiplexor in Figure 7, one can use the stationary Generalized Gaussian approximation (Figure 6b). The connection between stationary Generalized Gaussian and non-stationary Gaussian models will be further widely used in the paper for the design of optimal watermark extraction strategy and watermark decoder. Obviously, more complex models can be used that do not have the limitations of i.i.d. models and that take into account local correlation between image pixels.

The important aspect of stochastic image modeling is the estimation of the hyperparameters of the models. In the case of the nG model one must estimate the local mean and the local variance while in the case of the sGG model the local mean, the shape parameter and the global variance should be estimated. To estimate the local image variance the *maximum likelihood* estimate can be used. An example of such estimation is shown in Figure 8. It is necessary to note that the histogram of the local variance can be approximated as Weibull, Rice or gamma distributions or more roughly as exponential or Jeffreys priors. Knowing the statistics of the hyperparameters, as in the case with the local variance, one can model images as doubly stochastic processes.

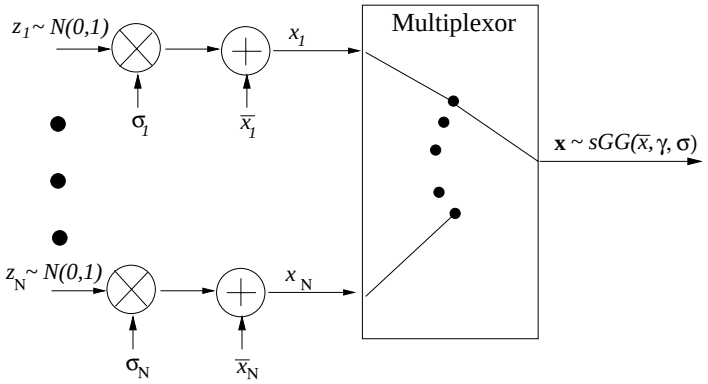
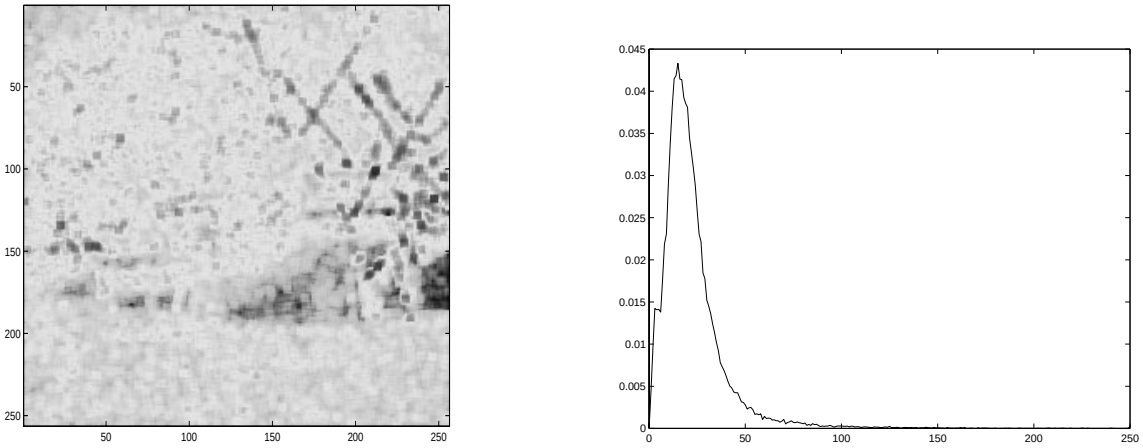


Figure 7. The generalized multichannel stochastic model of image generation.



(a) (b)

Figure 8. a) The local variance of the decomposed image; b) corresponding histogram.

5.1.2. Stochastic model of watermark

In the general case, we can use the same models for perceptually embedded watermark based on equation (1) as for the cover image. If the used perceptual model is known to an attacker and the information about the watermark embedding method is available, one can estimate the watermark directly from the stego image as was discussed above.

Assuming that the image and watermark are conditionally i.i.d. locally Gaussian, i.e. $x \sim N(\overline{x}, R_x)$ and $w \sim N(0, R_w)$ with covariance matrices R_x and R_w, where R_w also includes the effect of perceptual watermark modulation, one can determine:

$$\widehat{w} = \frac{R_w}{R_w + R_x} (y' - \overline{y}') \tag{9}$$

where it is assumed $\overline{y}' \approx \overline{x}$ and $\overline{y}'$ is a local mean of the attacked stego image that can be estimated based on local average, and where $\widehat{R}_x = \max\left(0, \widehat{R}_y - R_w\right)$ is the ML estimate of the local image variance $\left(\widehat{R}_x = \sigma_x^2 I\right)$. It is necessary to note that the local mean of the attacked image can be assumed to be zero, if the above prediction is performed in the wavelet domain. Then, the autocovariance function can be estimated using the ML estimate

for every wavelet sub-band coefficient. The equation (9) is the MAP/MMSE watermark extractor for oblivious watermarking for the considered above stochastic models.

6. WATERMARK DECODING

In the general case the decoder/demodulator design is based on ML or MAP approaches. Since the appearance of b is assumed to be equiprobable and due to the high complexity of the MAP decoders, ML decoders are mostly used in practice. The watermark decoder can be considered to consist of two main parts: a matched filter (detector) that performs a despreading of the data in the way of "coherent accumulation" of the sequence c spread in the watermark w, and the decoder itself that produces the estimate of the message. In most cases the results of attacks and of prediction/extraction errors are assumed to be additive Gaussian. The detector is therefore designed using a ML formulation for the detection of a known signal (projection sets are known due to the key) in Gaussian noise, that results in a correlator detector with reduced dimensionality:

$$r = \langle \widehat{w}, p \rangle \, . \tag{10}$$

Unfortunatly, the above matched filter does not take into account the practically important cases of fading and non-Gaussian noise. The equation (10) is typical for class I watermarking systems considered above. Therefore, to design a more realistic model of an equivalent watermarking channel we consider the transmission of the codeword c containing the encoded watermark and pilot bits through the parallel channel according to the diversity communication (Figure 9). The parameters of the channel are estimated using the pilot. We assume that the parameters of the affine transform $\widehat{A}$ are estimated and recovered prior to the decoding.

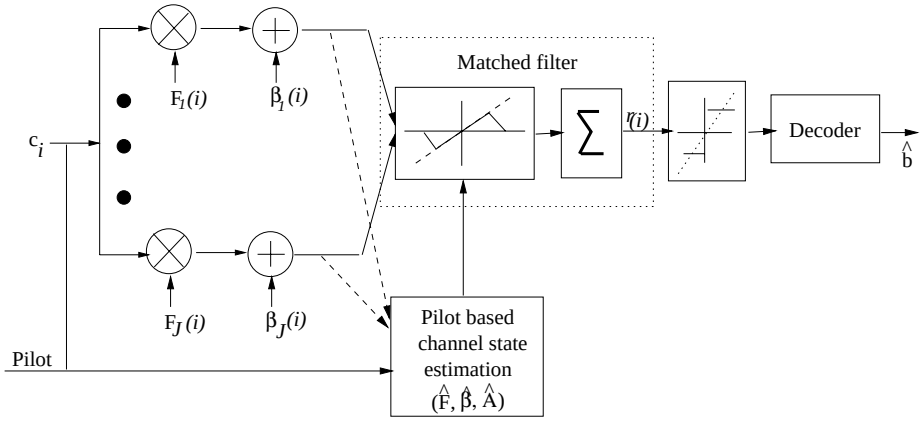


Figure 9. Equivalent parallel channel formulation of digital watermarking.

The equivalent channel model can be presented as:

$$c' = Fc + \beta \tag{11}$$

where F denotes the generalized fading and β the generalized noise in the equivalent parallel channel. The generalized fading includes several factors. First, the watermark is masked by the perceptual mask used for the embedding that has different values for the the different periods of the watermark replication. Secondly, attacks like denoising and lossy compression significantly decrease the strength of the watermark especially in the flat image regions, even reducing it to zero. Thirdly, the watermark extractor also modifies the amplitude of the watermark according to the local image statistics. The generalized noise includes all possible modifications of the watermark after attack that can be described using either non-stationary Gaussian or stationary Generalized Gaussian models. In the more general case a generalized matched filter can be designed that produces the output:

$$r = \langle g(\widehat{w}), p \rangle \tag{12}$$

where $g(.)$ is determined by the statistics of the generalized channel. In the particular case of the non-stationary Gaussian noise model the matched filter will have the following structure:

$$r_{nG} = \left\langle \widehat{R}_{\beta}^{-1} \widehat{F} \widehat{w}, p \right\rangle \tag{13}$$

or equivalently

$$r_{nG}(i) = \sum_{j \in P_i} \frac{\widehat{F}(j)\widehat{w}(j)p(j)}{\sigma_{\beta}(j)^2} \tag{14}$$

for all $i = 1, ..., K$ and where j is the index of diversity or the number of replication of bit c_i, $\widehat{R}_{\beta}$ is an estimate of the covariance matrix of the non-stationary Gaussian noise $N(0, {\sigma_{\beta}}^2 I)$ and $\widehat{F}$ is an estimate of the channel fading. Physically, the estimation of channel parameters is performed based on the assumption that the pilot bits are closely allocated to the corresponding equivalent channel bits of the codeword in the stego image. This does not exactly correpond to the communication channel and this analogy is slightly artificial here. However, assuming some certain degree of correlation between neightborhood pixels in the image we can assume that the pilot bits will have about the same modifications as the bits of the codeword. Therefore, this can be modeled as a slow fading channel with respect to the pilot signal. However, since the codewords are allocated with some random locations over the image local image correlation does not play any significant role, it can be modeled as a fast fading channel.

It is important to note that the matched filter produces a soft output that can be important for further decoding. The scheme with the hard output assuming binary symmetric channels was first proposed by Kundur *et al* in watermarking applications.[17] This scheme is considerably simplified and does not require the estimation of the channel non-stationary variances and parameters of the fading. It is modeled using only error probablities for each channel. As a consequence, weighted coefficients are derived for the diversity summation as in (14).

The assumption about stationary Generalized Gaussian noise distribution $sgg(0, \gamma_{\beta}, \sigma_{\beta})$ leads to the following matched filter:

$$r_{sGG}(i) = \sum_{j \in P_i} \frac{\left|\widehat{w}(j) + \widehat{F}(j)p(j)\right|^{\gamma_{\beta}(i)} - \left|\widehat{w}(j) - \widehat{F}(j)p(j)\right|^{\gamma_{\beta}(i)}}{\sigma_{\beta}(i)^{\gamma_{\beta}(i)}} \tag{15}$$

where $\sigma_{\beta}(i)$ and $\gamma_{\beta}(i)$ are constant for the given codeword bit c_i. The nonlinear structure of the matched filter is similar to a local optimum detector nonlinearity that limits the outliers of the sGG model.[18] This model was considered for the DCT domain watermarking by Hernández *et al*[19] where the sGG model presented the distribution of the DCT coefficients in 64 equivalent channels of JPEG compression. The authors considered this model assuming that all fading in the equivalent channel is only due to the perceptual masking and the used mask was proposed to be estimated directly from the attacked stego image. Therefore, in this formulation the matched filter is not completely adapted to the channel state variations, in contrast with the above considered pilot based technique. It is important to note that both considered matched filters (14) and (15) can be applied for the coordinate, wavelet and DCT domains.

The output of the matched filter is thresholded according to either the hard or the soft decoding (Figure 9) and then decoded. A decoder can be designed based on the MAP:

$$\widehat{b} = \arg\max_{\widetilde{b}} \; p\left(\widetilde{b} \mid r, x, k\right) \tag{16}$$

Assuming that all codewords b are equiprobable, given an observation vector r, the optimum decoder that minimizes the conditional probability of error is given by the ML decoder:

$$\widehat{b} = \arg\max_{\widetilde{b}} \; p\left(r \mid \widetilde{b}, x, k\right). \tag{17}$$

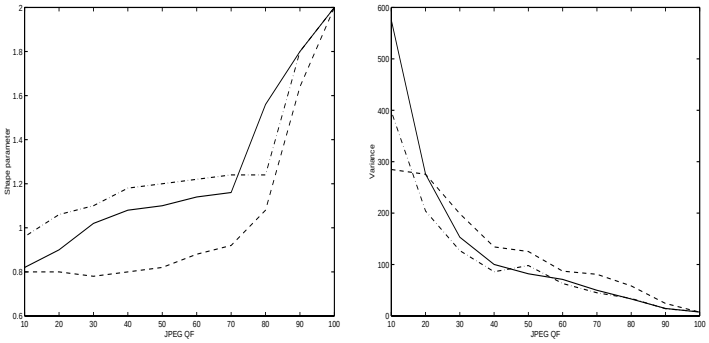
Based on the central limit theorem (CLT) most researchers assume that the observed vector r can be accurately approximated as the output of an additive Gaussian channel noise.[2,15]

7. EXPERIMENTAL RESULTS

In order to verify the adaptivity of the watermark decoder to the channel state variations we performed a number of experiments modeling the channel as lossy JPEG compression, or denoising accompanied by geometrical attacks. The pilot or the reference watermark is used to estimate and to recover from the geometrical transforms to synchronize decoding. Therefore, the final geometrical transform influence is reduced to interpolation errors.

We tested the performance of the proposed approach on the images included in the Stirmark benchmark. Here, we report results obtained for Lena image for lossy JPEG compression. As the basic watermarking algorithm we have chosen the method proposed in our previous work[6] that already contains a reference watermark used for the recovering from geometrical attacks and for reliability estimation. The embedding parameters for watermark were 10 for edges and textures and 2 for flat regions according to a noise visibility function embedding, which resulted in a PSNR of 38 dB.

This work also advocates the possibility to use the reference watermark as pilot for the channel state estimation. A first set of tests was performed to estimate the variation of the equivalent watermarking channel parameters after JPEG compression. The sGG model was used to approximate the pdf of the channel noise. The parameters of the sGG model were estimated using moment matching method and corresponding results for the shape parameter and the variance of the sGG are shown in Figure 10.



(a) (b)

Figure 10. Experimental results for Lena image: (a) the shape parameter estimation of sGG channel noise model for 3 test channels of 64-bit watermark vs JPEG compression; (b) corresponding plot of the estimated variances.

An important observation is a decrease of the shape parameter with a decrease of the QF. This shows that the commonly used linear correlation detector or matched filters designed for Gaussian noise are not anymore optimal for high compression ratio. The sign correlation detector cannot garantee optimality as well, since the shape parameter differs from 1 (Laplacian distribution). Secondly, the variance of channel noise also increases with the increase of compression ratio. That explains the decrease of decoding/detection performance caused by the decrease of the watermark PSNR. These experimental dependencies can be used for the computation of channel capacity and optimal watermak allocation in images.

The results of decoder adaptation to the channel noise variations are shown in Figure 11. The first row (Figure 11) shows the distribution of estimated noise in a channel of 64-bit channel watermark for different compression ratios. The corresponding nonlinearities of the matched filter for the sGG model are depicted in the second row of the same figure. The dashed lines refer to the linear correlator in assumption of Gaussian channel noise. Therefore, the matched filter is adapted to the changes in the equivalent watermarking channel.

The system was tested according to Stirmark3.1 benchmark and Table 1 summarizes the averaged performance of the proposed approach. The obtained results demonstrate the state-of-art performance of the proposed algorithm according to this benchmarking tool.

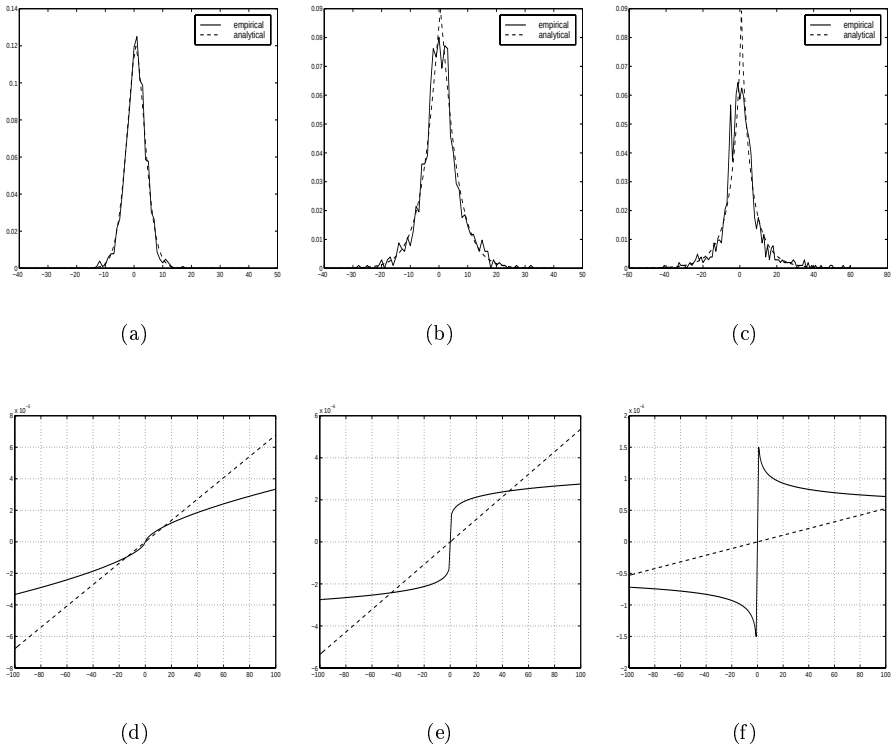


Figure 11. Channel noise estimation and approximation by sGG model (first row) for JPEG lossy compression and corresponding non-linearity of optimal matched filter (second row): (a,d) for JPEG QF=90%, (b,e) QF=40% and (c,f) QF=10%.

Table 1. Averaged results of system perfomance according to Stirmark3.1.

Stirmark attack	*Averaged score*
Signal enhancement	1,00
Compression (JPEG/GIF)	0,99
Scaling	1,00
Cropping	0,99
Shearing	1,00
Rotation (auto-crop, auto-scale)	0,99
Column and line removal	1,00
Flip	1,00

8. CONCLUSION

This paper presents a generalized stochastic approach to the design of oblivious watermarking systems. The approach is based on adequate modeling of the distortions in the equivalent watermarking channel. The optimal decoder is designed for the diversity watermarking and the adaptivity of the decoder to the watermark channel variations is demonstrated. The experiments validate the system and confirm its state-of-art nature.

ACKNOWLEDGMENTS

We thank Alexander Herrigel and Martin Kutter for many fruitful discussions. S. Voloshynovskiy is thankful to Igor Kozintsev (Microprocessor Research Labs, Intel) for clarifying many ideas about joint source-channel coding.

This work has been partially supported by the European project CERTIMARK and by DCT-Digital Copyright Technologies Switzerland.

REFERENCES

1. J. R. Hernández, F. Pérez-González, J. M. Rodríǵez, and G. Nieto, "Performance analysis of a 2-D-multipulse amplitude modulation scheme for data hiding and watermarking of still images," *IEEE Journal on Selected Areas in Communications* **16**, pp. 510–523, May 1998.
2. M. Kutter, "Watermarking resistent to translation, rotation and scaling," in *Proc. SPIE Int. Symp. on voice, Video, and Data Communication*, November 1998.
3. J. Oruanaidh and T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking," *Signal Processing* **66**(3), pp. 303–317, 1998.
4. J. R. Hernández, F. Pérez-González, J. M. Rodríǵez, and G. Nieto, "The impact of channel coding on the performance of spatial watermarking for copyright protection," *in Proc. ICASSP'98* **5**, pp. 2973–2976, May 1998.
5. S. Pereira, S. Voloshynovskiy, and T. Pun, "Effective channel coding for DCT watermarks," in *ICIP 2000*, (Vancouver, Canada), September 2000.
6. S. Voloshynovskiy, F. Deguillaume, and T. Pun, "Content adaptive watermarking based on a stochastic multiresolution image modeling," in *EUSIPCO 2000*, (Tampere, Finland), September 2000.
7. I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing* **6**, pp. 1673–1687, Dec. 1997.
8. T. Tirkel, C. Osborne, and R. van Schyndel, "Image watermarking - a spread spectrum application," in *Proc. IEEE Int. Symposium on Spread Spectrum Techniques and Applications*, vol. 2, pp. 785–789, 1996.
9. C. I. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE Journal on Selected Areas in Communications* **16**, pp. 525–539, May 1998.
10. F. Hartung and B. Girod, "Watermarking of uncompressed and compressed video," *Signal Processing* **66**, pp. 283–301, 1998.
11. I. J. Cox, M. L. Miller, and A. L. McKellips, "Watermarking as communications with side information," *Proceedings of the IEEE* **87**, pp. 1127–1141, July 1999.
12. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory* **29**, pp. 439–441, May 1983.
13. B. Chen and G. Wornell, "Dither modulation: a new approach to digital watermarking and information embedding," in *SPIE: Security and Watermarking of Multimedia Contents*, vol. 3657, (San Jose, California), January 1999.
14. J. Eggers, J. Su, and B. Girod, "A blind watermarking scheme based on structured codebooks," in *Secure images and image authentication, IEE Colloquium*, pp. 4/1–4/6, (London, UK), April 2000.
15. J. R. Hernández and F. Pérez-González, "Statistical analysis of watermarking schemes for copyright protection of images," *Proceedings of the IEEE* **87**, pp. 1142–1166, July 1999.
16. S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T. Pun, "A stochastic approach to content adaptive digital image watermarking," in *Third International Workshop on Information Hiding*, (Dresden, Germany), September 29 - October 1st 1999.
17. D. Kundur and D. Hatzinakos, "Improved robust watermarking through attack characterization," in *Optics Express*, vol. 3, pp. 485–490, December 1998.
18. S. Kassam, *Signal Detection in Non-Gaussian Noise*, Springer Verlag, 1998.
19. J. R. Hernández, M. Amado, and F. Pérez-González, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Trans. on Image Processing* **9**, pp. 55–68, January 2000.