A Stochastic Approach to Content Adaptive Digital Image Watermarking

Sviatoslav Voloshynovskiy[1], Alexander Herrigel[2], Nazanin Baumgaertner[2], and Thierry Pun[1]

[1] CUI - University of Geneva, Department of Computer Science
24 rue General Dufor, CH-1211 Geneva 4, Switzerland
[2] DCT - Digital Copyright Technologies, Research & Development
Stauffacher-Strasse 149, CH-8004 Zurich, Switzerland

Abstract. This paper presents a new stochastic approach which can be applied with different watermark techniques. The approach is based on the computation of a Noise Visibility Function (NVF) that characterizes the local image properties, identifying textured and edge regions where the mark should be more strongly embedded. We present precise formulas for the NVF which enable a fast computation during the watermark encoding and decoding process. In order to determine the optimal NVF, we first consider the watermark as noise. Using a classical MAP image denoising approach, we show how to estimate the "noise". This leads to a general formulation for a texture masking function, that allows us to determine the optimal watermark locations and strength for the watermark embedding stage. We examine two such NVFs, based on either a non-stationary Gaussian model of the image, or a stationary Generalized Gaussian model. We show that the problem of the watermark estimation is equivalent to image denoising and derive content adaptive criteria. Results show that watermark visibility is noticeably decreased, while at the same time enhancing the energy of the watermark.

1 Introduction

Digital image watermarking is applied today as a popular technique for authentication and copyright protection of image data. Based on global information about the image characteristics, many approaches and commercial solutions embed the watermarking signal as random noise in the whole cover image with the same strength regardless of the local properties of the image. This embedding may lead in practice to visible artifacts specially in the flat regions which are characterized by small variability. In order to decrease these distortions the given watermark strength has to be decreased. This, however, reduces drastically the robustness of the watermark against different sorts of attacks, since the image regions which generate the most visible artifacts determine the final maximum strength of the watermark signal to be embedded.
We present for this problem an effective solution[1] which embeds the watermark

[1] This work has been supported by the Swiss National Science Foundation (Grant 5003-45334) and the EC Jedi-Fire project (Grant 25530)

A. Pfitzmann (Ed.): IH'99, LNCS 1768, pp. 211–236, 2000.
© Springer-Verlag Berlin Heidelberg 2000

into the cover image according to the local properties of the image, i.e. applying this technique every watermarking algorithm will be *content adaptive*. In contrast to recently published results we present a new stochastic approach to identify the regions of interest for the embedding. This approach has the advantage that it is applicable for very different types of images and is not constrained with the identification of an adequate set of parameters to be determined before the identification of the local charactersitics as it is the case in [8]. In addition, the approach presented may be applied to different domains, such as coordinate, Fourier and wavelet. We show the interrelationship to the image denoising problem and prove that some of the applied techniques are special cases of our approach. Comparing the derived stochastic models with results recently published we have noticed that some heuristically derived formula are close to our problem solution. The stationary Generalized Gaussian model is superior concerning the strength of the watermark to be embedded as well as the image quality.

2 State-of-the-Art Approaches

Some authors tried to develop content adaptive schemes on the basis of the utilization of luminance sensitivity function of the human visual sytem (HVS) [1]. Since the derived masking function is based on the estimation of the image luminance a soley luminance based embedding is not efficient against wavelet compression or denoising attacks. Another group of watermarking algorithms exploites transfer modulation features of HVS in the transform domain to solve the compromise between the robustness of the watermark and its visibility [3]. This approach embeds the watermark in a predetermined middle band of frequencies in the Fourier domain with the same strength assuming that the image spectra have isotropic character. This assumption leads to some visible artifacts in images specially in the flat regions, because of anisotropic properties of image spectra. A similar method using blocks in DCT (discrete cosine transform) domain was proposed in [4]. In the context of image compression using perceptually based quantizers, this concept was further developed in [5] to a content adaptive scheme, where the watermark is adjusted for each DCT block. However, as the original image is required to extract the watermark, the practical applications of this approach are very limited since it can be shown that the usage of the cover image will results in watermark schemes which can be easily broken. Another DCT based algorithm which uses luminance and texture masking was developed by [6]. The next group of methods is also based on the image compression background [7] and practically exploits 3 basic conclusions of the above paper: (1) all regions of high activity are highly insensitive to distortion; (2) the edges are more sensitive to distortion than highly textured areas; (3) darker and brighter regions of the image are less sensitive to noise. The typical examples of this approach are ([8], [9]). The developed methods consist of a set of empirical procedures aimed to satisfy the above requirements. The computational complexity and the absence of closed form expressions for the perceptual mask

complicate the analysis of the received results. However, experiments performed in these papers show high robustness of these approaches. A very similar method was proposed in [10], where edge detectors are used to overcome the problem of visibility of the watermark around the edges.
The goal of this paper is to develop a coordinate domain content adaptive criterion which may easily be applied to any watermarking technique in coordinate, Fourier, DCT or wavelet domains as perceptual modulation function. The basic idea of our approach consists in the adequate stochastic modelling of the cover image. This allows the estimation of the image as well as the watermark and makes the application of the information theory to the watermarking problem possible. Knowing stochastic models of the watermark and the cover image, one can formulate the problem of watermark estimation/detection according to the classical Bayessian paradigm and estimate the capacity issue of the image watermarking scheme. Developing this concept we show that the problem of watermark estimation is equivalent to image denoising, and derive content adaptive criteria. Finally, we will show the relevance of the developed criterion to the known empirical results.

3 Problem Formulation

Consider the classical problem of non-adaptive watermark embedding, i.e. embedding the watermark regardless of the image content. In the most general case it can be defined according to the next model:

$$y = x + n, \tag{1}$$

where y is the stego image ($y \in \mathcal{R}^N$ and $N = M \times M$), x is the cover (original) image and n is associated with the noise-like watermark image encoded according to the spread spectrum technique [11]. Our goal is to find an estimate $\hat{n}$ of the watermark n either directly, or equivalently, an estimate $\hat{x}$ of the cover image x and then compute an estimation of the watermark as:

$$\hat{n} = y - \hat{x}, \tag{2}$$

where $\hat{n}$ and $\hat{x}$ denote the estimates of the watermark and the cover image respectively. The decision about the presence/absence of the watermark in a given image is then made by a robust detector. This detector must consider the prior statistics of the watermark and the possible errors of its estimation due to the decomposition residual coefficients of the cover image and the possibly applied attack. The generalized scheme of such an approach could be schematically represented according to figure (1). This generalized idea has found practical applications in the watermarking algorithm [1] and steganography method [12]. The key moments of the above approach are the design of the corresponding estimator and the robust detector. The problem of estimation of the cover image from its noisy version is known as image denoising or image smoothing. We will concentrate our consideration on its analysis and its solution in this paper, and

show a way of deriving some stochastic criteria for content adaptive watermarking. To solve this problem we use the *Maximum a Posteriori Probability (MAP)* approach.

Our stochastic approach is based on two models of the image where the image is assumed to be a random process. We consider a stationary and non-stationary process to model the cover image. The stationary process is characterized by the constant parameters for the whole image and the non-statinary has spatially varying parameters. To estimate the parameters a maximum likelihood estimate is used in the specified neighbourhood set. We assume that image is either a non-statinary Gaussian process or a stationary Generalized Gaussian. In contrast to many information theory approaches for watermarking, we don't consider the distribution of an image as purely stationary Gaussian, since the image regions of interest for watermarking have different local features. In addition, the channel capacity is not uniform since the image contents of every window constitute a channel capacity which is closely linked to the local image characteristics which are not uniform over the whole image and dependent from visible artifacts.

4 Watermark Estimation Based on MAP

To integrate the watermarking problem into a statistical framework, a probabilistic model of the watermark and the cover image must be developed. If the watermark has the distribution $p_n(n)$ and the cover image the distribution $p_x(x)$, then according to the MAP criterion, the watermark estimate could be found as:

$$\hat{n} = argmax_{\tilde{n}\in\mathcal{R}^N} L(\tilde{n}|y), \tag{3}$$

where $L(\tilde{n}|y)$ is the log function of the *a posteriori* distribution:

$$L(\tilde{n}|y) = \ln p_x(y|\tilde{n}) + \ln p_n(\tilde{n}). \tag{4}$$

The estimate $\hat{n}$, found according to equation (2), leads to the next formulation:

$$\hat{x} = argmax_{\tilde{x}\in\mathcal{R}^N}\{\ln p_n(y|\tilde{x}) + \ln p_x(\tilde{x})\}. \tag{5}$$

The problems (3), (4) and (5) are according to the formulation (2) equivalent to each other. The formulation (5) is the typical image denoising problem, and will be considered further in the present paper.

To solve this problem it is necessary to develop the accurate stochastic models for the watermark $p_n(n)$ and the cover image $p_x(x)$.

Under the assumption that the watermark is received using spread spectrum technique, it is possible to model it as Gaussian random variable. Let samples $n_{i,j}$ $(1 \le i, j \le M)$ be defined on vertices of an $M \times M$ grid, and let each sample $n_{i,j}$ take a value in $\mathcal{R}$. Let further all the samples be *independent indentically distributed (i.i.d)* . Then:

$$p_n(y|x) = \frac{1}{\sqrt{(2\pi\sigma_n^2)^N}} \cdot \exp\{-\frac{1}{2\sigma_n^2}(y-x)^T(y-x)\}, \tag{6}$$

where σ_n^2 is the variance of the watermark. This assumption is reasonable, because of the fact that the Gaussian distribution has the highest entropy among all other distributions and hence, from the security point of view, any spread spectrum encoding algorithm should approach this distribution in the limit. Thus, the watermark could be modelled as $n \sim N(0, \sigma_n^2)$. The next important question is the development of an adequate *prior* model of the cover image.

5 Stochastic Models of the Cover Image

One of the most popular stochastic image model, which has found wide application in image processing, is the *Markov Radom Field (MRF)* model [13]. The distribution of MRF's is written using a Gibbs distribution:

$$p(x) = \frac{1}{Z} e^{-\sum_{c \in A} V_c(x)}, \tag{7}$$

where Z is a normalization constant called the *partition function*, $V_c(\cdot)$ is a function of a local neighboring group c of points and A denotes the set of all possible such groups or cliques.

In this paper we will consider two particular cases of this model, i.e. the Gaussian and the Generalized Gaussian (GG) models. Assume that the cover image is a random process with non-stationary mean. Then using autoregressive (AR) model notations, one can write the cover image as:

$$x = A \cdot x + \varepsilon = \bar{x} + \varepsilon, \tag{8}$$

where $\bar{x}$ is the non-stationary local mean and ε denotes the residual term due to the error of estimation. The particularities of the above model depend on the assumed stochastic properties of the residual term:

$$\varepsilon = x - \bar{x} = x - A \cdot x = (I - A) \cdot x = C \cdot x, \tag{9}$$

where $C = I - A$ and I is the unitary matrix. If A is a low-pass filter, then C represents a high-pass filter (decomposition operator).

We use here two different models for the residual term ε. The first model is the non-stationary (inhomogeneous) Gaussian model and the second one is the stationary (homogeneous) Generalized Gaussian (GG) model.

The choice of these two models is motivated by the fact that they have found wide application in image restoration and denoising ([14], [15]), and that the best wavelet compression algorithms are based on these models ([16], [17]). Their main advantage is that they take local features of the image into account. In the first case, this is done by introducing the non-stationary variance using a quadratic energy function, and in the second case, by using an energy function, which preserves the image discontinuities under stationary variance. In other words, in the non-stationary Gaussian model, the data is assumed to be *locally i.i.d.* random field with a Gaussian probability density function (pdf), while in the stationary GG model the data is assumed to be *globally i.i.d.*.

The autocovariance function in the non-stationary case can be written as:

$$R_x = \begin{pmatrix} \sigma^2_{x_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{x_2} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma^2_{x_N} \end{pmatrix}, \tag{10}$$

where $\{\sigma^2_{x_i} | 1 \leq i \leq N\}$ are the local variances.
The autocovariance function for the stationary model is equal to:

$$R_x = \begin{pmatrix} \sigma^2_x & 0 & \cdots & 0 \\ 0 & \sigma^2_x & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma^2_x \end{pmatrix}, \tag{11}$$

where σ^2_x is the global image variance.

According to equation (9) the non-stationary Gaussian model is characterized by a distribution with autocovariance function (10):

$$p_x(x) = \frac{1}{(2\pi)^{\frac{N}{2}}} \cdot \frac{1}{|\det R_x|^{\frac{1}{2}}} \cdot \exp\{-\frac{1}{2}(Cx)^T R_x^{-1} Cx\}, \tag{12}$$

where $|\det R_x|$ denotes the matrix determinant, and ^{T} denotes transposition. The stationary GG model can be written as:

$$p_x(x) = (\frac{\gamma\eta(\gamma)}{2\Gamma(\frac{1}{\gamma})})^{\frac{N}{2}} \cdot \frac{1}{|\det R_x|^{\frac{1}{2}}} \cdot \exp\{-\eta(\gamma)(|Cx|^{\frac{\gamma}{2}})^T R_x^{-\frac{\gamma}{2}} |Cx|^{\frac{\gamma}{2}}\}, \tag{13}$$

where $\eta(\gamma) = \sqrt{\frac{\Gamma(\frac{3}{\gamma})}{\Gamma(\frac{1}{\gamma})}}$ and $\Gamma(t) = \int_0^\infty e^{-u} u^{t-1} du$ is the gamma function, R_x is determined according to (11), and the parameter γ is called the *shape parameter*. Equation (13) includes the Gaussian ($\gamma = 2$) and the Laplacian ($\gamma = 1$) models as special cases. For the real images the shape parameter is in the range $0.3 \leq \gamma \leq 1$.

Having defined the above models of the watermark and the cover image, we can now formulate the problem of image estimation according to the MAP approach.

6 Image Denoising Based on the MAP Approach

MAP estimate of the cover image results in the following general optimization problem:

$$\hat{x} = argmin_{\tilde{x} \in \mathcal{R}^N} \{\frac{1}{2\sigma_n^2} \|y - \tilde{x}\|^2 + \rho(r)\}, \tag{14}$$

where $\rho(r) = [\eta(\gamma) \cdot |r|]^{\gamma}$, $r = \frac{x-\bar{x}}{\sigma_x} = \frac{Cx}{\sigma_x}$ and $\|\cdot\|$ denotes the matrix norm. $\rho(r)$ is the energy function for the GG model.

In the case of the non-stationary Gaussian model (12) $\gamma = 2$ (i.e. convex function) and σ_x^2 is spatially varying. The advantage of this model is the existence of a closed form solution of (14) in form of adaptive Wiener or Lee filters [18]. In the case of stationary GG model the general closed form solution does not exist, since the penalty function could be non-convex for $\gamma < 1$. In practice, iterative algorithms are often used to solve this problem. Examples of such algorithms are the stochastic [13] and deterministic annealing (mean-field annealing) ([20]), graduated nonconvexity [19], ARTUR algorithm [21] or its generalization [22]. However, it should be noted that for the particular case of $\gamma = 1$, a closed form solution in wavelet domain exists: it is known as soft-shrinkage ([23], [14]). Of course, it is preferable to obtain the closed form solution for the analysis of the obtained estimate. To generalize the iterative approaches to the minimization of the non-convex function (14) we propose to reformulate it as a *reweighted least squares (RLS)* problem. Then equation (14) is reduced to the following minimization problem:

$$\hat{x}^{k+1} = argmin_{\tilde{x} \in \mathcal{R}^N} \{ \frac{1}{2\sigma_n^2} \|y - \tilde{x}^k\|^2 + w^{k+1} \|r^k\|^2 \}, \tag{15}$$

where

$$w^{k+1} = \frac{1}{r^k} \rho'(r^k), \tag{16}$$

$$r^k = \frac{x^k - \bar{x}^k}{\sigma_x^k}, \tag{17}$$

$$\rho'(r) = \gamma [\eta(\gamma)]^{\gamma} \frac{r}{\|r\|^{2-\gamma}}, \tag{18}$$

and k is the number of iterations. In this case, the penalty function is quadratic for a fixed weighting fuction w.

Assuming w is constant for a particular iteration step, one can write the general RLS solution in the next form:

$$\hat{x} = \frac{w\sigma_n^2}{w\sigma_n^2 + \sigma_x^2} \bar{x} + \frac{\sigma_x^2}{w\sigma_n^2 + \sigma_x^2} y. \tag{19}$$

This solution is similar to the closed form Wiener filter solution [18]. The same RLS solution could also be rewritten in the form of Lee filter [18]:

$$\hat{x} = \bar{x} + \frac{\sigma_x^2}{w\sigma_n^2 + \sigma_x^2} (y - \bar{x}). \tag{20}$$

The principal difference with classical Wiener or Lee filters is the presence of the weighting function w. This weighting function depends on the underlying assumptions about the statistics of the cover image. In the rest of this paper, we will only consider the Lee version of the solution which coincides with the

classical case of Gaussian prior of the cover image ($w = 1$). It is important to note that the shrinkage solution of image denoising problem previously used only in the wavelet domain can easily be obtained from (15) in the next closed form:

$$\hat{x} = \bar{x} + \max(0, |y - \bar{x}| - T) sign(y - \bar{x}),$$

where $T = \frac{\sigma_n^2}{\sigma_x}\sqrt{2}$ is the threshold for practically important case of Laplacian image prior. This coincides with the soft-thresholding solution of the image denoising problem [14].

The properties of the image denoising algorithm are defined by the multiplicative term:

$$b := \frac{\sigma_x^2}{w\sigma_n^2 + \sigma_x^2} \tag{21}$$

in equations (19) and (20). It is commonly known, that the local variance is a good indicator of the local image activity, i.e. when it is small, the image is flat, and a large enough variance indicates the presence of edges or highly textured areas. Therefore, the function b determines the level of image smoothing. For example, for flat regions $\sigma_x^2 \to 0$, and the estimated image equals local mean, while for edges or textured regions $\sigma_x^2 >> \sigma_n^2$, $b \to 1$ and the image is practically left without any changes. Such a philosophy of the adaptive image filtering is very well matched with the texture masking property of the human visual system: the noise is more visible in flat areas and less visible in regions with edges and textures. Following this idea we propose to consider the texture masking property according to the function b based on the stochastic models (12) and (13).

7 Texture Masking Function

In order to be in the framework of the existing terminology, we propose to relate the texture masking function to the *noise visibility function (NVF)* as:

$$NVF = 1 - b = \frac{w\sigma_n^2}{w\sigma_n^2 + \sigma_x^2}, \tag{22}$$

which is just the inverted version of the function b.

In the main application of the proposed NVF in context of watermarking, we assume that the noise (watermark) is an i.i.d. Gaussian process with unit variance, i.e. $N(0, 1)$. This noise excite the perceptual model (22), in analogy with the AR image model. The NVF is the output of the perceptual model (22) to a noise $N(0, 1)$. This model is schematically shown in figure (2). The particularities of the developed perceptual model are determined by the weighting w, which are defined according to eqs. (17) and (18).

7.1 NVF Based on Non-stationary Gaussian Model

In the case of non-stationary Gaussian model the shape parameter γ is equal to 2 and the autocovariance function is defined according to (10). The weighting

function w is then equal to 1 according to equation (17) and NVF can simply be written in the form:

$$NVF(i,j) = \frac{1}{1+\sigma_x^2(i,j)}, \tag{23}$$

where $\sigma_x^2(i,j)$ denotes the local variance of the image in a window centered on the pixel with coordinates (i,j), $1 \le i,j \le M$. Therefore, the NVF is inversely proportional to the local image energy defined by the local variance. In order to estimate the local image variance the *maximum likelihood (ML)* estimate can be used. Assuming that the image is a locally i.i.d. Gaussian distributed random variable, the ML estimate is given by:

$$\sigma_x^2(i,j) = \frac{1}{(2L+1)^2} \sum_{k=-L}^{L} \sum_{l=-L}^{L} (x(i+k,j+l) - \bar{x}(i,j))^2 \tag{24}$$

with

$$\bar{x}(i,j) = \frac{1}{(2L+1)^2} \sum_{k=-L}^{L} \sum_{l=-L}^{L} x(i+k,j+l), \tag{25}$$

where a window of size $(2L+1)\times(2L+1)$ is used for the estimation. This estimate is often used in practice in many applications. However, the above estimate is assymptotically unbiased. To decrease the bias, it is necessary to enlarge the sampling space. From the other side, enlarging the window size violates the requirement of data being locally Gaussian, since the pixels from different regions occur in the same local window. In order to have a more accurate model, it is reasonable to assume that flat regions have a Gaussian distribution while textured areas and regions containing edges have some other highly-peaked, near-zero distribution (for example Laplacian). This assumption requires the analysis of a mixture model with the Huber energy function (so called Huber-Markov Random Fields) and is the subject of our ongoing research. In this paper, we concentrate on the analysis of the eq. (22) and its relation to the stationary GG version of NVF.

7.2 NVF Based on Stationary GG Model

For the stationary GG model we rewrite eq. (22) in the following form, taking eqs. (17) and (18) into account:

$$NVF(i,j) = \frac{w(i,j)}{w(i,j)+\sigma_x^2}, \tag{26}$$

where $w(i,j) = \gamma[\eta(\gamma)]^{\gamma} \frac{1}{\|r(i,j)\|^{2-\gamma}}$ and $r(i,j) = \frac{x(i,j)-\bar{x}(i,j)}{\sigma_x}$.

The particularities of this model are determined by the choice of two parameters of the model, e.g. the shape parameter γ and the global image variance σ_x^2. To estimate the shape parameter, we use a *moment matching* method as

the one used in [17]. The analysis consists of the next stages. First, the image is decomposed according to the equation (9), using equation (25) as an estimate of the local mean. In the second stage, the moment matching method is applied to the residual image and the shape parameter and the variance are estimated. Some typical examples of these estimations are shown in figure (3). The shape parameter for most of real images is in the range $0.3 \leq \gamma \leq 1$. As a comparison with frequently used pdfs such as Gaussian and Laplacian (also known as double-exponential), we have given the plots of these distributions and their corresponding penalty (energy) function in figures 4(a) and 4(b), respectively. The derivatives of these energy functions and their corresponding weighting functions are shown in figures 4(c) and 4(d). Comparing these pdfs, it is clear that smaller shape parameters lead to a distribution, which is more highly-peaked near zero. Another important conclusion regards the convexity of the energy function, which is concave (non-convex) for $\gamma < 1$, convex for $\gamma = 1$ and strictly convex for $\gamma > 1$. This causes the known properties of this non-convex function in the interpolation applications and the corresponding discontinuity preservation feature, when it is used in restoration or denoising applications. Despite of the nice edge preserving properties, the use of the non-convex functions causes the known problems in the minimization problem (15). However, in the scope of this paper we are mostly interested in the analysis of the properties of the term (26), under condition that the cover image is available for designing the content adaptive watermarking scheme.

8 Stochastical NVFs and Empirical Models

In the derivation of the content adaptive function we explicitly used stochastic modelling of images. However, it is also very important to investigate the relevance of these analytical results with the empirically obtained results, which reflect the particularities of the human visual system.

Obviously, the development of a complete model of the HVS that reflects all its particularities is quite a difficult task. Therefore, most of the developed empirical models utilize only the main features which are important for certain applications, and give reasonable approximations for the practical use. Such sort of models have been used in deterministic image denoising and restoration algorithms ([24], [25]). They have also been used in the field of image compression to reduce the visual artifacts of the lossy compression algorithms and to design special quantization schemes (see for applications in DCT domain [26], and in wavelet domain [27] and [28]). This fact explains the appearance of a great amount of watermarking algorithms based on the transform domain embedding ([5]). Another reason that motivates the use of the transform domain is the decrease of the inter-pixel redundancy by performing image decomposition in the transform domain. In our case image decomposition is obtained in a natural way directly in the coordinate domain by decomposing the image into low-frequency and high-frequency fractions. This is similar to the Gaussian and Laplacian pyra-

mids and to the wavelet decomposition, where scale and orientation (3 directions) are additionally exploited.

To take the texture masking properties into account, it was proposed in [24] to use the NVF for image quantization in the coordinate domain in prediction schemes and also to extend it to the problem of image denoising. The most known form of the empirical NVF is widely used in image restoration applications [25]:

$$NVF(i,j) = \frac{1}{1 + \theta\sigma_x^2(i,j)}, \tag{27}$$

where θ is a tuning parameter which must be chosen for every particular image. This version of NVF was the basic prototype for a lot of adaptive regularization algorithms.

Comparing the above function with the stochastically derived NVF based on the non-stationary Gaussian model, it is very easy to establish the similarity between them. The only difference is the tuning parameter θ which plays the role of the contrast adjustment in NVF. To make θ image-dependent, it was proposed to use:

$$\theta = \frac{D}{\sigma_{x_{max}}^2}, \tag{28}$$

where $\sigma_{x_{max}}^2$ is the maximum local variance for a given image and $D \in [50, 100]$ is an experimentally determined parameter.

9 Content Adaptive Watermark Embedding

Using the proposed content adaptive strategy, we can now formulate the final embedding equation:

$$y = x + (1 - NVF) \cdot n \cdot S, \tag{29}$$

where S denotes the watermark strength. The above rule embeds the watermark in highly textured areas and areas containing edges stronger than in the the flat regions.

In very flat regions, where NVF approaches 1, the strength of the embedded watermark approaches zero. As a consequence of this embedding rule, the watermark information is (nearly) lost in these areas. Therefore, to avoid this problem, we propose to modify the above rule, and to increase the watermark strength in these areas to a level below the visibility threshold:

$$y = x + (1 - NVF) \cdot n \cdot S + NVF \cdot n \cdot S_1, \tag{30}$$

with S_1 being about 3 for most of real world and computer generated images. We are now investigating more complex cases, to replace this fixed value by an image-dependent variable, which takes the luminance sensitivity of HVS into account.

Finally, the watermark detection and the message demodulation will be accomplished in accordance with the general scheme presented in (1), where the

estimation $\hat{x}$ of the cover image is done using the image denoising algorithm (15).

The performance of the resulting watermarking algorithm depends strongly on the particular scheme designed to resist geometrical distortions or the basic image processing operations such as filtering, compression and so on. Examples of such distortions are integrated in StirMark watermarking benchmark [2]. For instance, the coordinate domain method given in[1] uses properties of the autocorrelation function of a spatially spreaded watermark, while the Fourier domain approach developed by [3] utilizes a pre-defined template to detect and compensate the undergone geometrical distortions. The adaptive approach proposed here can be integrated into these methods to achieve the best trade-off between the two contradicting goals of increasing the robustness by increasing the watermark strength and at the same time, decreasing the visual artifacts introduced by the watermarking process. This is the subject of our current research.

Here, we address mainly the problem of content adaptive watermark embedding, and investigate its visibility aspects in the next section.

10 Results of Computer Simulation

To illustrate the main features of the proposed content adaptive embedding method, we tested our algorithm on a number of real world and computer generated images. In this paper we will restrict our consideration to several images with most typical properties. The first stage of modeling consists of the calculation of the NVFs according to the developed stochastic models of the cover image. The NVFs based on non-stationary Gaussian (a) and stationary GG (b) models for Barbara and Fish images are shown in Figures (5) and (6), respectively. The NVF calculated according to the non-stationary Gaussian model is smoother and the intensity for the edges is about the same order as that of textured regions. The NVF calculated from stationary GG model looks more *noisy* due to the discontinuity preserving properties of the non-convex energy function of GG prior. From the point of view of noise visibility, both functions reflect the regions of reduced visibility. This corresponds to the desirable features of a content adaptive watermarking approach. To demonstrate the possibility of integrating this adaptive approach into transform domain based techniques, another set of experiments was performed, which demonstrate the relationship between coordinate domain and Fourier domain properties of the cover image and the corresponding NVFs. The magnitude spectrum of Camaraman image (7a) and the corresponding spectra of the NVFs based on the non-stationary Gaussian (7b) and on the stationary GG (7c) model were calculated. It is important to note that the spectra of real images are characterized by a high level of spatial anisotropy, where the principal component directions are determined by the properties of the image. These directions in the spatial spectrum of the original image are determined by the spatial orientation of the edges and textures in the coordinate domain and play a very important role for the high quality image

aqcuisition systems under resolution constraints of imaging aperture [30] . The prospective watermarking techniques have to take these properties into account, to survive denoising or lossy compression attacks. Such a strategy implies that the watermark should be most strongly embedded in these directions. The comparison of the spectrum of the cover image with the spectra of corresponding NVFs, shows that the directions of the principal components coincide. This can be the solution to finding a compromise between making the watermark as robust as possible and avoiding the visible distortions that are introduced by the watermarking process. The next stage of modeling is the comparison of the visual quality of the stego images generated using the different developed NVFs. The test consists in the direct embedding of a Gaussian distributed watermark according to the equation (1) and to the content adaptive schemes described in equation (30), using both NVFs. The strength of the watermark (i.e. its standard deviation) is equal to 5, 10 and 15 for the non-adaptive scheme (1). The corresponding results are shown in Figure (8) in parts (a),(d) and (g). The *peak-signal-to noise ratio (PSNR)* was chosen as the criterion for the comparison of the introduced distortions in the stego image or equivalentlly to estimate the general energy of the watermark

$$PSNR = 10 \log_{10} \frac{\|255\|^2}{\|x - y\|^2}$$

The resulted PSNRs for Barbara image are gathered in table (1).

Table 1. Comparison of the watermark strength for different watermark embedding methods

Embedding method	**Barbara: Watermark strength**		
Non-adaptive	5.0	10.0	15.0
Adaptive non-stationary Gaussian NVF	6.5	15.0	23.0
Adaptive stationary GG NVF	8.0	20.0	31.0
Adaptive Kankanhalli	9.8	19.6	30.38
PSNR (dB)	34.1	28.3	24.7

To receive similar PSNR values in case of adaptive embedding, the watermark strength was increased in these cases to the values given in table (1). The resulting images are depicted in figure (8), parts (b), (e), (h) and figure (8), parts (e), (f), (i) for the non-stationary Gaussian and the stationary GG based NVFs, respectively. The corresponding images in the appendix show that even the PSNR is very similar, the image qualities are quite different. Comparing both visual quality and objective parameters given by the watermark strength, the adaptive scheme based on NVF calculated from stationary GG model seems to be superior to other schemes. This can be explained regarding the above

Table 2. Comparison of the watermark strength for different watermark embedding methods

Embedding method	**Fish: Watermark strength**		
Non-adaptive	5.0	10.0	15.0
Adaptive non-stationary Gaussian NVF	6.5	14.5	23.0
Adaptive stationary GG NVF	9.0	21.0	35.0
Adaptive Kankanhalli	10.5	21.7	32.2
PSNR (dB)	34.64	28.67	25.26

mentioned properties of the local variance estimation in equation (23), given by the ML-estimate (24). In particular, the watermark is very visible on the edges, because here the requirement of data being local stationary is violated. This decreases the performance of the ML-estimate. In the stationary GG model, only the global variance estimation is needed (equation (26)), where the data is considered to be globally i.i.d.. Therefore, distortions due to edges are considerably smaller than in the non-stationary Gaussian model embedding scheme. The same experiments were performed for Fish image and the results are shown in table (2). In some watermarking algorithms [1], the watermark is embedded in the coordinate domain as a pseudo-random spatially spreaded binary sequence -1;1. To investigate the visibility aspects of this watermark in the scope of the proposed approach, a similar test was performed. The results are shown in figure (9). The PSNR values and the watermark strengths practically coincide with the data given in table (1).

Based on the empirical evaluation of different test people, we think that the image quality of our approach is better than the results of the Kankanhalli approach (see the corresponding stego images in the appendix(figures 11, 12).

The stationary GG model shows a superior performance in this case too. The third stage of the modeling consists in the investigation of the watermark robustness against adaptive Wiener denoising. The resulting PSNR of the watermark after the above attack was determined as

$$PSNR_n = 10 \log_{10} \frac{\|255\|^2}{\|n - n_a\|^2}$$

where n_a is the watermark after attack. The results of this experiment are presented in table (2). The watermarks before and after this attack are shown in figure (10) for a PSNR equal to 28.3 dB. Inspection of the results of figure (10) show that the watermark survives mostly in the highly textured regions and areas with edges, where its strength was increased according to the developed content adaptive scheme. Therefore, the obtained results clearly indicate the advantages of the proposed approach for digital image watermarking.

Table 3. The watermark PSNR after adaptive Wiener denoising attack for different watermark embedding methods

embedding method	**watermark PSNR (dB)** (after adaptive Wiener denoising)
non-adaptive	25.99
adaptive non-stationary Gaussian NVF	26.86
adaptive stationary GG NVF	26.97

11 Conclusions and Future Work

In this paper we have presented a new approach for content adaptive watermarking which can be applied to different watermarking algorithms based on the coordinate domain, the Fourier or wavelet domains. In contrast to state-of-the-art heuristic techniques our approach is based on a stochastic modelling framework which allows us to derive the corresponding analytic formula in a closed form solution. It is, therefore, not necessary to investigate in practice the different parameter sets for different image classes but to apply mathematical expressions which can be easily computed. In addition, we have shown the close relationship of the derived optimization problems and the corresponding solutions to image denoising techniques. We have shown that some of the heuristically derived solutions are special cases of our model. Running different tests we have sucessfully compared our approach against other techniques. The test results have also shown that some assumptions in many information theoretic papers for watermarking are not satisfied, namely, the distribution of the image is not stationary Gaussian and the channel capacity is not uniform meaning that depending on the sample, the image contents constitute a channel capacity which is closely linked to the local image characteristics which are not uniform. It is, therefore, necessary to consider the enhanced information theory based approaches which satisfy these constraints. Based on the identified relationship to image denoising problems, we are going to develop the corresponding attacks on commercial watermark schemes to test our models as the basis of a new piracy tool and for further enhancement of the existing watermarking techniques. In the future, we will apply the presented model also for the construction of more robust encoder and decoder schemes.

12 Acknowledgements

We appreciate usefull comments of S. Pereira, G. Csurka and F. Deguillaume during the work on the paper. S. Voloshynovskiy is grateful to I. Kozintsev, K. Mihcak, A. Lanterman and Profs. P. Moulin and Y. Bresler for fruitfull and helpfull discussions during his staying at University of Illinois at Urbana-Champaign.

References

1. M.Kutter: *Watermarking Resisting to Translation, Rotation and Scaling*, Proc. of SPIE, Boston, USA, November 1998. 212, 213, 222, 224
2. M.Kutter, F.Petitcolas: *A fair benchmark for image watermarking systems*, SPIE, Vol.3657, San Jose, January 1999, pp.226-239. 222
3. J.Ruanaidh, T.Pun: *Rotation, Scale and Translation Invariant Spread Spectrum Digital Image Watermarking*, Signal Processing, May 1998, Vol.66, No.3, pp.303-317. 212, 222
4. M.Swanson, B.Zhu, A.Twefik: *Transparent Robust Image Watermarking*, Proc. of 3rd IEEE International Conference on Image Processing ICIP96, 1996, Vol.3, pp.211-214. 212
5. C.Podilchuk, W.Zeng: *Image Adaptive Watermarking Using Visual Models*, IEEE Journal on Selected Areas in Communication, May 1998, Vol.16, No.4, pp.525-539. 212, 220
6. J.Huang, Y.Shi: *Adaptive Image Watermarking Scheme Based on Visual Masking*, Electronic Letters, April 1998, Vol.34, No.8, pp.748-750. 212
7. N.Jayant, J.Johnston, R.Safranek: *Signal Compression Based on Models of Human Perception*, Proc. of the IEEE, 1993, Vol.81, No.10, pp.1385-1422. 212
8. M.Kankanhalli, R.Ramakrishnan: *Content Based Watermarking of Images*, ACM Mulimedia98, Bristol, UK, 1998, pp. 61-70. 212
9. F.Bartolini, M.Barni, V.Cappelini, A.Piva: *Mask Bilding for Perceptually Hiding Frequency Embedded Watermarks*, Proc. of 5th IEEE International Conference on Image Processing ICIP98, Chicago, Illinois, USA, October 4-7, 1998, Vol.1, pp. 450-454. 212
10. J.F.Delaigle, C.De Vleeschouwer, B.Macq: *Watermarking Algorithm Based on a Human Visual Model*, Signal Processing, 1998, Vol.66, pp. 319-335. 213
11. I.Cox, J.Kilian, T.Leighton, T.Shamoon: *Secure Spread Spectrum Watermarking for Multimedia*, NEC Research Institute Tech Rep. 95-10, 1995. 213
12. L.Marvel, C.Retter, C.Boncelet: *Hiding Information in Images*, Proc. of 5th IEEE International Conference on Image Processing ICIP98, Chicago, Illinois, USA, October 4-7, 1998, Vol.1. 213
13. S.Geman and D.Geman: *Stochastic Relaxation, Gibbs Distributions and the Bayesian Restorations of Images*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 1984, Vol.14, No.6, pp.367-383. 215, 217
14. P.Moulin, J. Liu: *Analysis of Multiresolution Image Denoising Schemes Using Generalized-Gaussian Priors*, Proc. IEEE Sig. Proc. Symp. on Time-Frequency and Time-Scale Analysis, , Pittsburgh, PA, October 1998. 215, 217, 218
15. S.Chang, B.Yu, M.Vetterli: *Spatially Adaptive Wavelet Thresholding with Content Modeling for Image Denoising*, Proc. of 5th IEEE International Conference on Image Processing ICIP98, Chicago, Illinois, USA, October 4-7, 1998. 215
16. S.LoPresto, K.Ramchandran, M.Orhard: *Image Coding Based on Mixture Modeling of Wavelet Coefficients and a Fast Estimation-Quantization Framework*, Data Compression Conference 97, Snowbird, Utah, 1997, pp.221-230. 215
17. S.Mallat: *A Theory for Multiresolution Signal Decomposition: The Wavelet Representation*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 1989, Vol.11, No.7, pp.674-693. 215, 220
18. J.S.Lim: *Two-Dimensional Signal and Image Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1990. 217
19. A.Blake, A.Zisserman: *Visual Reconstruction*, MA: The MIT Press, 1987. 217

20. D.Deiger, F.Girosi: *Parallel and Deterministic Algorithms from MRFs Surface Reconstruction*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 1991, Vol.13, No.6, pp.401-412. 217
21. P.Charbonnier, L.Blanc-Feraud, G.Aubert, M.Barlaud: *Deterministic Edge-Preserving Regularization in Computed Images*, IEEE Trans. on Image Processing, 1997, Vol.6, No.2, pp.298-311. 217
22. A.Dalaney, Y.Bresler: *Globally Convergent Edge-Preserving Regularization: An Application to Limited-Angle Tomography*, IEEE Trans. on Image Processing, 1998, Vol.7, No.2, pp.204-221. 217
23. M.Nikolova: *Estimees Locales Forment Homogenes*, Comptes Rendus Ac. Sci. Paris, Serie I, 1997, Vol.325, pp.665-670. 217
24. G.L.Adelson, A.N.Natravali: *Image Restoration Based on Subjective Criterion*, IEEE Trans. Syst., Man, Cyber. SMC-6, 1976, pp.845-853. 220, 221
25. S.Efstratiadis, A.Katsaggelos: *Adaptive Iterative Image Restoration with Reduced Computational Load*, Optical Engineering, Dec. 1990, Vol.29, No.12, pp.1458-1468. 220, 221
26. A.Watson: *DCT Quantization Matrices Visually Optimized for Individual Images*, in Proc. SPIE Conf. Human Vision, Visual Processing and Digital Display IV, 1993, Vol.1913, pp.202-216. 220
27. A.Watson, G.Yang, J.Solomon, J.Villasenor: *Visual Thresholds for Wavelet Quantization Error*, in Proc. SPIE Human Vision and Electronic Imaging, 1996, Vol.2657, pp.381-392. 220
28. A.S.Lewis, G.Knowles: *Image Compression Using 2-D Wavelet Transform*, IEEE Trans. Image Processing, No.4, 1992, pp.244-250. 220
29. S.Pereira, J.Ruanaidh, F.Deguillaume, G.Cscurka, and T.Pun: *Template Based Recovery of Fourier-Based Watermarks Using Log-Polar and Log-Log Maps*, Proc. Int. Conference on Multimedia Computing and Systems, June 1999.
30. I.Prudyus, S.Voloshynovskiy, T.Holotyak: *Adaptive Aperture Formation in Radar Imaging Systems with Nonlinear Robust Image Restoration*, In IX European Signal Processing Conference Eusipco-98, Island of Rhodes, Greece, September 8-11 1998, vol. 3, pp.1365-1368. 223

13 Appendix

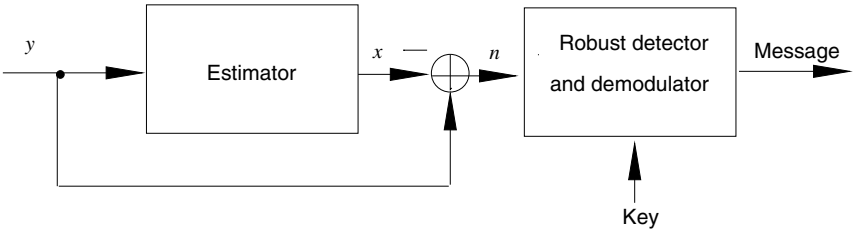


Fig. 1. Block diagram of the watermark detection algorithm according to equation (2).

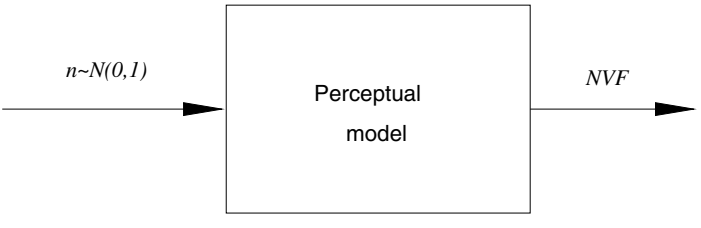


Fig. 2. Generation of the NVF from perceptual model.

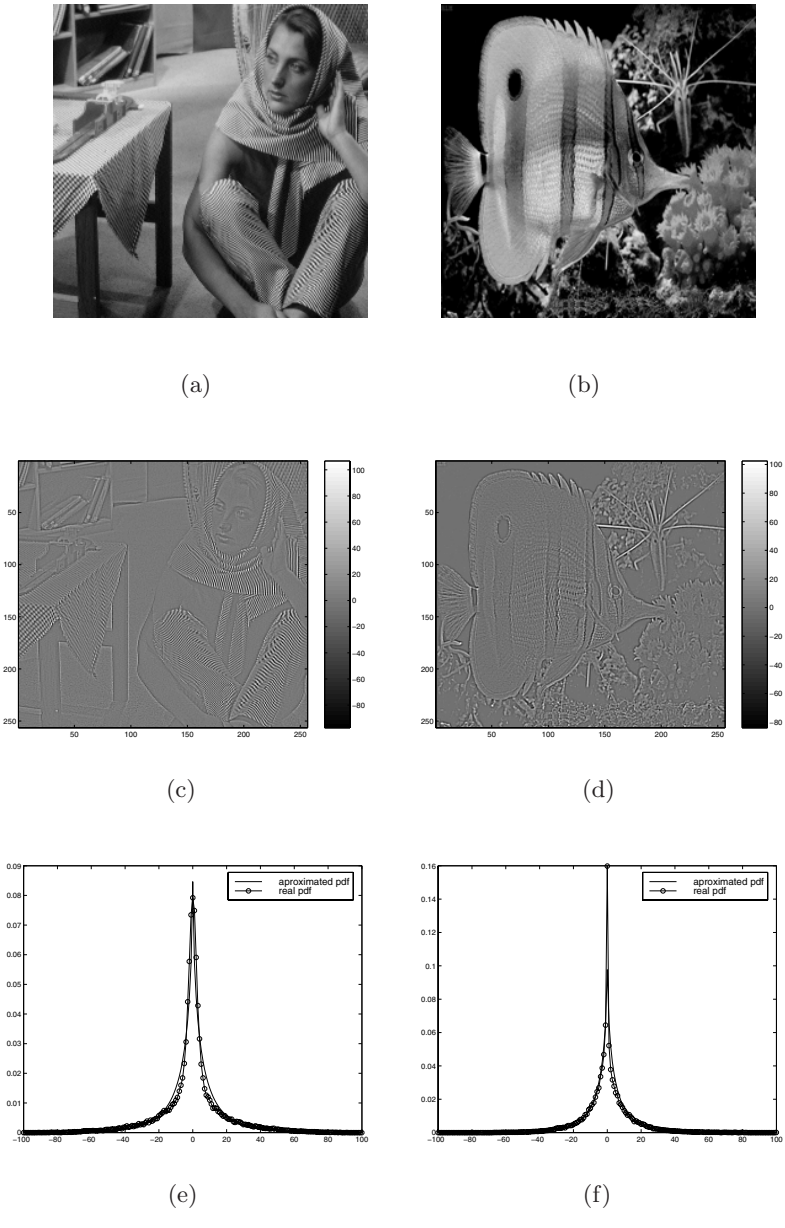


Fig. 3. Results of the GG model parameter estimation: cover images Barbara and Fish (a) and (b); residual decomposed images (c) and (d); histogram plots and their approximation by GG pdfs (e) and (f). The estimated parameters are $\sigma_x = 17.53$ and $\gamma = 0.64$ for Barbara image and $\sigma_x = 13.40$ and $\gamma = 0.68$ for Fish image.

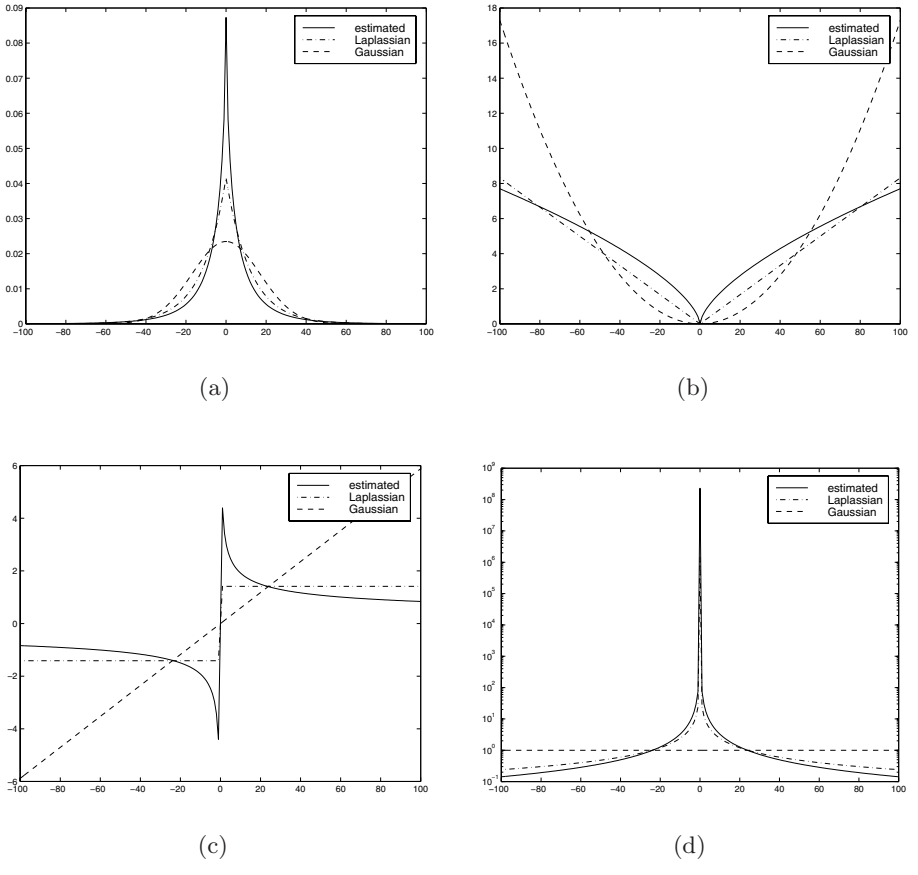
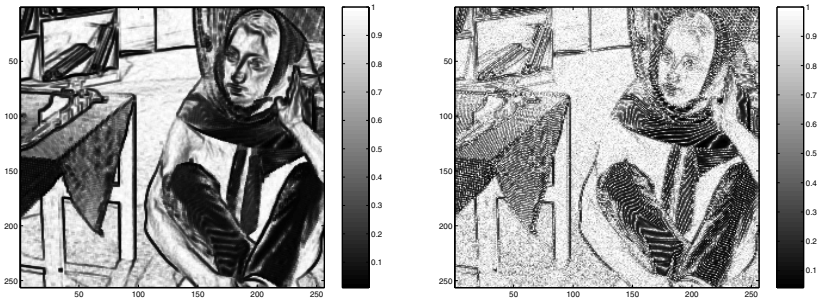
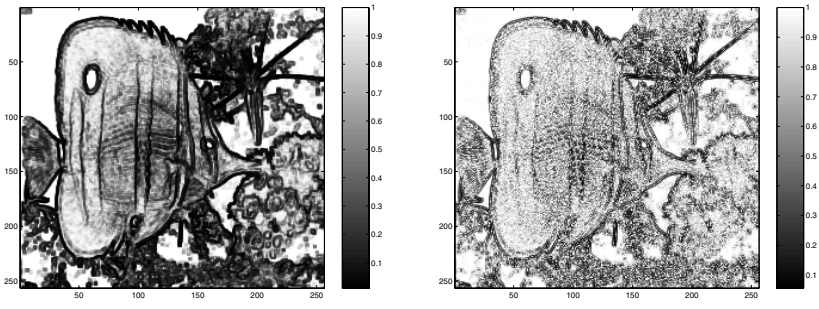


Fig. 4. Comparative analysis of the estimated distribution $\gamma = 0.64$ with Gaussian and Laplacian (a); corresponding energy (penalty) functions ρ (b); the derivatives from the energy functions ρ' (c) and weighting functions w (d). For all distributions $\sigma_x = 17.53$.



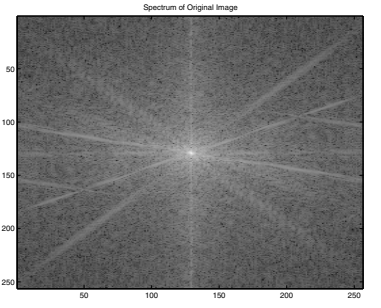
(a) (b)

Fig. 5. NVFs from Barbara image calculated for the non-stationary Gaussian (a) and the stationary GG models (b).

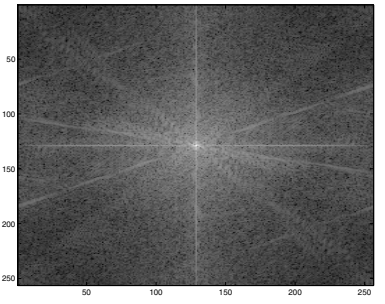


(a) (b)

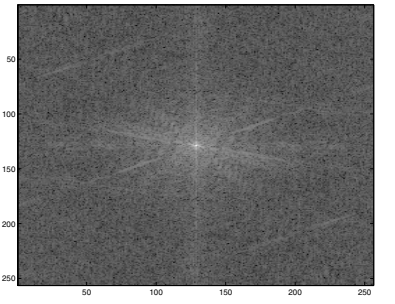
Fig. 6. NVFs from Fish image calculated for the non-stationary Gaussian (a) and the stationary GG models (b).



(a)



(b)



(c)

Fig. 7. Magnitude spectrum of the original image Cameraman (a), and spectra of NVFs calculated according to non-stationary Gaussian (b) and stationary GG (c) models.

Fig. 8. Test on visibility of embedding of Gaussian distributed watermark: direct embedding according to scheme (1) with the watermark strength 5 (a), 10 (d) and 15 (g); and adaptove shemes with NVF (23) (b, e, h) and NVF (26) (c, f, i) with the corresponding watermark strengthes given in Table (1).

(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 9. Visibility for embedding of binary uniformly distributed watermark. For reference see Figure (8).

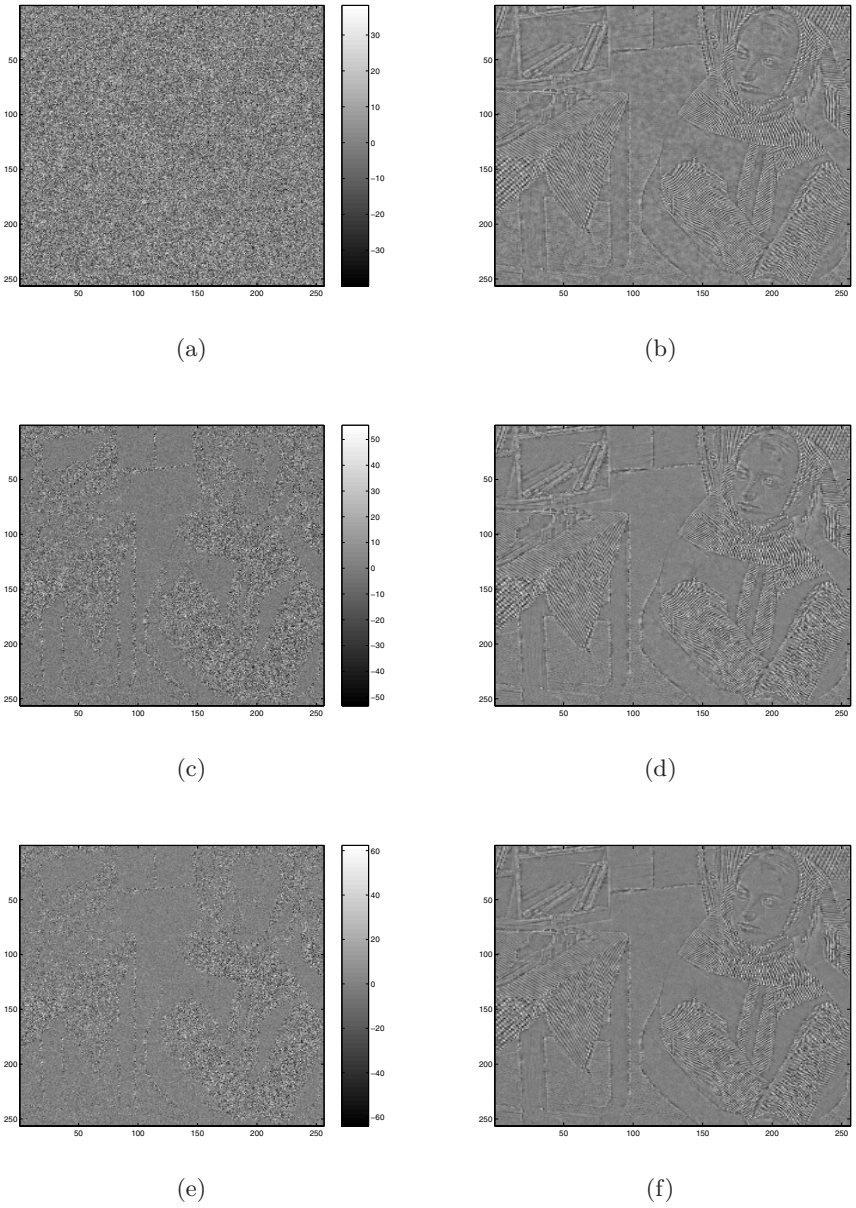
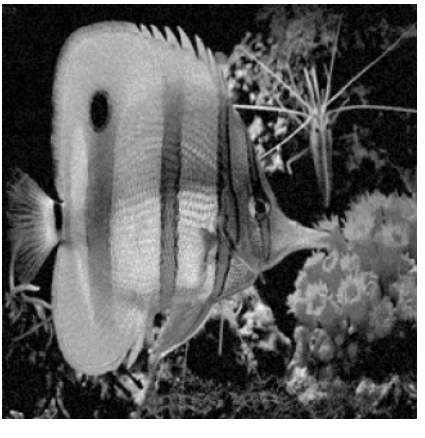


Fig. 10. Results of adaptive Wiener denoising attack: left column represents watermark embedded according to the direct scheme (1) (a), adaptive ones with NVF (23) (c) and NVF (26) (e), and right column shows the remained watermark after denosing attack for the corresponding images from right column.

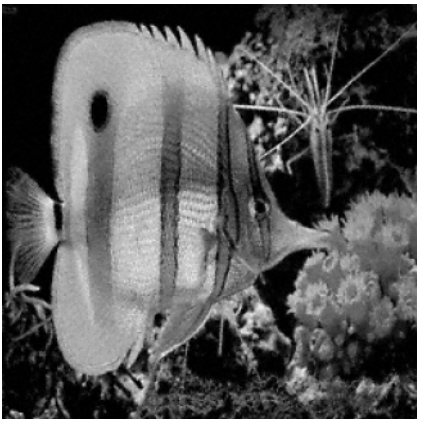
(a)

(b)

Fig. 11. The Barbara stego images calculated for the Kankanhalli (a) and the stationary GG model (b).



(a)



(b)

Fig. 12. The Fish stego images calculated for the Kankanhalli (a) and the stationary GG model (b).