Optimized wavelet domain watermark embedding strategy using linear programming

Shelby Pereira[a] and Sviatoslav Voloshynovskiy[a] and Thierry Pun[a]

[a]University of Geneva - CUI,
24 rue General Dufour, CH 1211
Geneva 4, Switzerland

ABSTRACT

Invisible Digital watermarks have been proposed as a method for discouraging illicit copying and distribution of copyright material. In recent years it has been recognized that embedding information in a transform domain leads to more robust watermarks. In particular, several approaches based on the Wavelet Transform have been proposed to address the problem of image watermarking. The advantage of the wavelet transform relative to the DFT or DCT is that it allows for localized watermarking of the image. A major difficulty, however, in watermarking in any transform domain lies in the fact that constraints on the allowable distortion at any pixel are specified in the spatial domain. In order to insert an invisible watermark, the current trend has been to model the Human Visual System (HVS) and specify a masking function which yields the allowable distortion for any pixel. This complex function combines contrast, luminance, color, texture and edges. The watermark is then inserted in the transform domain and the inverse transform computed. The watermark is finally adjusted to satisfy the constraints on the pixel distortions. However this method is highly suboptimal since it leads to irreversible losses at the embedding stage because the watermark is being adjusted in the spatial domain with no care for the consequences in the transform domain.

The central contribution of the paper is the proposal of an approach which takes into account the spatial domain constraints in an optimal fashion. The main idea is to structure the watermark embedding as a linear programming problem in which we wish to maximize the strength of the watermark subject to a set of linear constraints on the pixel distortions as determined by a masking function. We consider the Haar wavelet and Daubechies 4-tap filter in conjunction with a masking function based on a non-stationary Gaussian model, but the algorithm is applicable to any combination of transform and masking functions. Our results indicate that the proposed approach performs well against lossy compression such as JPEG and other types of filtering which do not change the geometry of the image.

1. INTRODUCTION

The idea of using a robust digital watermark to detect and trace copyright violations has stimulated significant interest among artists and publishers in recent years. Podilchuk[1] gives three important requirements for an effective watermarking scheme: transparency, robustness and capacity. Transparency refers top the fact that we would like the watermark to be invisible. The watermark should also be robust against a variety of possible attacks by pirates. These include robustness against compression such as JPEG, scaling and aspect ratio changes, rotation, cropping, row and column removal, addition of noise, filtering, cryptographic and statistical attacks, as well as insertion of other watermarks.[2] In this publication however, we consider only attacks that do not change the geometry of the image. The other requirement is that the watermark be able to carry a certain amount of information i.e. capacity. In order to attach a unique

E-mail: {Shelby.Pereira,svolos,Thierry.Pun}@cui.unige.ch

identifier to each buyer of an image, a typical watermark should be able to carry at least 60-100 bits of information. However, most of the work in watermarking has involved a one bit watermark. That is, at detection a binary decision is made as to the presence of the watermark most often using hypothesis testing.[3] Barni[4] encodes roughly 10 bits by embedding 1 watermark from a set of 1000 into the DCT domain. The recovered watermark is the one which yields the best detector response. Here we consider the case of an 80 bit watermark.

Watermarking methods can be divided into two broad categories: spatial domain methods such as[5–7] and transform domain methods . Transform domain methods have for the most part focused on DCT,[1,8,9,4,10] DFT[11–13] and most recently wavelet domain methods.[1,14,15] Transform domain methods have several advantages over spatial domain methods. Firstly, it has been observed that in order for watermarks to be robust, they must be inserted into the perceptually significant parts of an image. For images these are the lower frequencies which can be marked directly if a transform domain approach is adopted.[16] Since compression algorithms operate in the frequency domain (for example DCT for JPEG and wavelet for EZW) it is possible to optimize methods against compression algorithms. In[17] it is shown that by carefully matching the embedding strategy with the JPEG compression algorithm it is possible to resist JPEG compression at a quality factor of 10 for small images (64×64). On the other hand, the DFT domain has also been successfully adopted in algorithms which attempt to recover watermarks from images which have undergone affine transformations.[11]

While transform domain watermarking clearly offers benefits, the problem is more challenging since it is more difficult to generate watermarks which are adapted to the human visual system (HVS). The problem arises since typically constraints on the acceptable level of distortion for a given pixel are specified in the spatial domain. In the bulk of the literature on adaptive transform domain watermarks, a watermark is generated in the transform domain and then the inverse transform is applied to generate the spatial domain counterpart. The watermark is then modulated as a function of a spatial domain mask in order to render it invisible. However this spatial domain modulation is suboptimal since it changes the original frequency domain watermark. One approach which has recently appeared is the attempt at specifying the mask in the transform domain.[1]

In this publication we develop an alternate approach where we derive an optimized strategy for embedding a watermark in the wavelet domain when the masking constraints are specified in the spatial domain. The work is an extension of[10] where we propose a flexible framework in which we optimally embed a DCT domain watermark given the constraints imposed in the spatial domain. This framework overcomes the problems with many proposed algorithms which adopt a suboptimal spatial domain truncation and modulation as determined by masking constraints. We begin in section 2 by presenting the spatial domain masking methods we adopt in the rest of the paper. In section 3 the embedding algorithm is described. In section 4 we present our results followed by the conclusion in section 5.

2. SPATIAL DOMAIN MASKING

One of the most popular stochastic image model, which has found wide application in image processing, is the *Markov Random Field (MRF)* model.[18] The distribution of MRF's is written using a Gibbs distribution:

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

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 a special case of this model known as the Generalized Gaussian (GG) model is adopted. Assume that the cover image is a random process with non-stationary mean. Then, using autoregressive (AR) model notation, one can write the cover image as:

$$x = A \cdot x + \varepsilon = \bar{x} + \varepsilon, \tag{2}$$

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:

$$\varepsilon = x - \bar{x} = x - A \cdot x = (I - A) \cdot x = C \cdot x, \tag{3}$$

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

Here we use the stationary Generalized Gaussian (GG) model for the residual term ε. The advantage of this model is that it takes local features of the image into account. This is accomplished by using an energy function, which preserves the image discontinuities under stationary variance.

The auto-covariance function for the stationary model is equal to:

$$R_x = \begin{pmatrix} \sigma_x^2 & 0 & \cdots & 0 \\ 0 & \sigma_x^2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_x^2 \end{pmatrix}, \tag{4}$$

where σ_x^2 is the global image variance. The model can be written as:

$$p_x(x) = \left(\frac{\gamma\eta(\gamma)}{2\Gamma(\frac{1}{\gamma})}\right)^{\frac{N}{2}} \cdot \frac{1}{|\det R_x|^{\frac{1}{2}}} \cdot \exp\{-\eta(\gamma)(|Cx|^{\frac{\gamma}{2}})^T R_x^{-\frac{\gamma}{2}} |Cx|^{\frac{\gamma}{2}}\}, \tag{5}$$

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 (4), and the parameter γ is called the *shape parameter*. Equation (5) 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 $0.3 \leq \gamma \leq 1$.

It has been shown[19] that from the generalized Gaussian model, we can derive a noise visibility (NVF) at each pixel position as:

$$NVF(i,j) = \frac{w(i,j)}{w(i,j) + \sigma_x^2}, \tag{6}$$

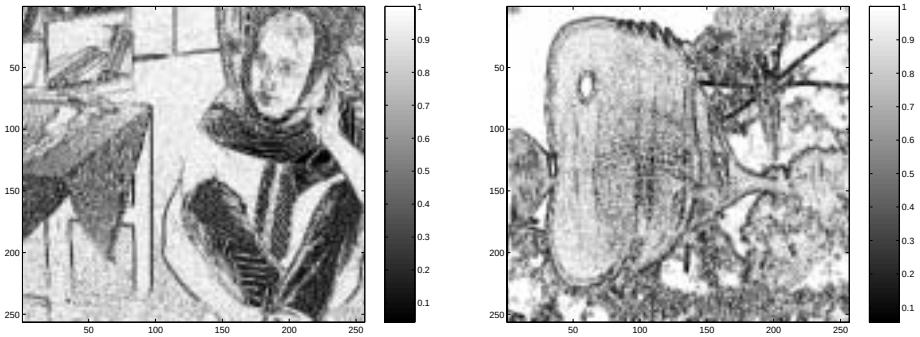
where $w(i,j) = \gamma[\eta(\gamma)]^\gamma \frac{1}{\|r(i,j)\|^{2-\gamma}}$ and $r(i,j) = \frac{x(i,j) - \bar{x}(i,j)}{\sigma_x}$.

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 the *moment matching* method in.[20] The shape parameter for most of real images is in the range $0.3 \leq \gamma \leq 1$. Once we have computed the noise visibility function we can obtain the allowable distortions by computing:

$$\Delta_{pi,j} = (1 - NVF(i,j)) \cdot S + NVF(i,j) \cdot S_1 \tag{7}$$

where S and S_1 are the maximum allowable pixel distortions in textured and flat regions respectively Typically S may be as high as 30 while S_1 is usually about 3. We note that in flat regions the NVF tends to 1 so that the first term tends to 0 and consequently the allowable pixel distortion is at most S_1 which is small. Intuitively this makes sense since we expect that the watermark distortions will be visible in flat regions and less visible in textured regions. Examples of NVFs for two images are given in figure 2. We note that the model correctly identifies textured and flat regions. In particular the NVF is close to 0 in textured regions and close to 1 in flat regions.

(a) (b)



(c) (d)

Figure 1. Original images of Barbara (a) and Fish (b) along with their NVF as determined by a generalized gaussian model (c) and (d).

3. WATERMARK EMBEDDING AS CONSTRAINED OPTIMIZATION

We assume that we are given an image to be watermarked denoted $\boldsymbol{I}$. If it is an RGB image we work with the luminance component. We are also given a masking function $\boldsymbol{V}(\boldsymbol{I})$ which returns 2 matrices of the same size of $\boldsymbol{I}$ containing the values $\Delta_{pi,j}$ and $\Delta_{ni,j}$ corresponding to the amount by which pixel $I_{i,j}$ can be respectively increased and decreased without being noticed. In general, the function $\boldsymbol{V}$ can be a complex function of texture,luminance, contrast, frequency and patterns. Here we adopt the generalized gaussian model of an image and consequently the construction of $\boldsymbol{V}(\boldsymbol{I})$ from equation 7 is straightforward. We note that $\Delta_{pi,j}$ and $\Delta_{ni,j}$ are not necessarily the same since in addition to the NVF we also take into account truncation effects. That is pixels are integers in the range $0-255$; consequently it is possible to have a pixel whose value is 1 which can be increased by a large amount, but can be decreased by at most 1. We wish to embed $\boldsymbol{m} = (m_1, m_2 ... m_M)$ where $m_i \in \{0, 1\}$ and M is the number of bits in the message. Without loss of generality we assume the image $\boldsymbol{I}$ is of size 128×128 corresponding to a very small image. For larger images the same procedure is adopted for each 128×128 large block.

To embed the message, we first divide the image into 16×16 blocks and perform the 1-level wavelet transform. In order to embed a 1 or 0 we adopt a differential encoding strategy in the lowest subband (LL). In particular we choose four neighbouring coefficients and increase two coefficients while decreasing the other two. The choice of which two to increase or decrease is a function of whether we wish to encode

a 1 or a 0 so that at decoding we take the difference between the sums of the two pairs of coefficients and apply the mappings $(+ \rightarrow 1),(- \rightarrow 0)$. We note that it is important to select a 2×2 block of *neighbouring* coefficients since the underlying assumption is that the difference on average is 0. In order to embed the largest possible values while satisfying masking constraints, the problem is formulated for each 16×16 block as a constrained optimization problem. In the case of the Haar wavelet, for a 16×16 block, we have 64 coefficients available in the LL subband. In each block we encode 8 bits by selecting 32 coefficients grouped into 8 2×2 blocks. We then have:

$$\min_{\boldsymbol{x}} \boldsymbol{f}'\boldsymbol{x} \qquad ; \qquad \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \tag{8}$$

where $\boldsymbol{x} = [x_{1,1} \dots x_{16,1} x_{12} \dots x_{16,2} \dots x_{1,16} \dots x_{16,16}]^t$ is the vector of coefficients arranged column by column. $\boldsymbol{f}$ is a vector of zeros except in the positions of the selected coefficients where we insert a (-1) or (1) depending on whether we wish to respectively increase or decrease the value of a coefficient. $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$ contain the constraints which are partitioned as follows.

$$\boldsymbol{A} = \begin{bmatrix} IDWT \\ - - - - \\ -IDWT \end{bmatrix} ; \quad \boldsymbol{b} = \begin{bmatrix} \Delta_p \\ - - - - \\ \Delta_n \end{bmatrix} \tag{9}$$

where IDWT is the matrix which yields the 2D inverse DWT transform of $\boldsymbol{x}$ (with elements of the resulting image arranged column by column in the vector). We also note that we take Δ_p and Δ_n to be column vectors where the elements are taken columnwise from the matrices of allowable distortions. If we let D_{ij} be the coefficients of the 1D inverse DWT (known as the synthesis matrix[22]) then it is easily shown that the matrix IDWT in our 2D case is given by:

$$IDWT = \begin{bmatrix} D_{1,1}D_{1,1} & \dots & D_{1,16}D_{1,1} & D_{1,1}D_{2,1} & \dots & D_{1,16}D_{2,1} & \dots & D_{1,1}D_{16,1} & \dots & D_{1,16}D_{16,1} \\ D_{2,1}D_{1,1} & \dots & D_{2,16}D_{1,1} & D_{2,1}D_{2,1} & \dots & D_{2,16}D_{2,1} & \dots & D_{2,1}D_{16,1} & \dots & D_{2,16}D_{16,1} \\ \vdots & & & & & & & & & \\ D_{16,1}D_{1,1} & \dots & D_{16,16}D_{1,1} & D_{16,1}D_{2,1} & \dots & D_{16,16}D_{2,1} & \dots & D_{16,1}D_{16,1} & \dots & D_{16,16}D_{16,1} \\ D_{1,1}D_{1,2} & \dots & D_{1,16}D_{1,2} & D_{1,1}D_{2,2} & \dots & D_{1,16}D_{2,2} & \dots & D_{1,1}D_{16,2} & \dots & D_{1,16}D_{16,2} \\ \vdots & & & & & & & & & \\ D_{16,1}D_{1,2} & \dots & D_{16,16}D_{1,2} & D_{16,1}D_{2,2} & \dots & D_{16,16}D_{2,2} & \dots & D_{16,1}D_{16,2} & \dots & D_{16,16}D_{16,2} \\ \vdots & & & & & & & & & \\ D_{16,1}D_{1,16} & \dots & D_{16,16}D_{1,16} & D_{16,1}D_{2,16} & \dots & D_{16,16}D_{2,16} & \dots & D_{16,1}D_{16,16} & \dots & D_{16,16}D_{16,16} \end{bmatrix} \tag{10}$$

Stated in this form the problem is easily solved by the well known Simplex method. Stated as such the problem only allows for spatial domain masking, however many authors[21] suggest also using frequency domain masking. This is possible by adding the following constraints:

$$\boldsymbol{L} \leq \boldsymbol{x} \leq \boldsymbol{U} \tag{11}$$

Here $\boldsymbol{L}$ and $\boldsymbol{U}$ are the allowable lower and upper bounds on the amount we by which we can change a given frequency component. The Simplex method can also be used to solve the problem with added frequency domain constraints. We note that by adopting this framework, we in fact allow *all* coefficients to be modified (in a given 16x16 block) even though we are only interested in a subset of coefficients (in the LL subband) at decoding. In words, we are "making space" for the watermark in an optimal fashion by modifying elements from the orthogonal complement of the coefficients we are interested in, while satisfying spatial domain constraints.

4. RESULTS

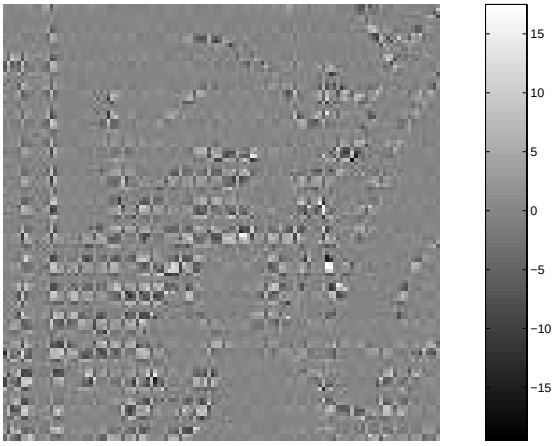
The algorithm was tested on several small images of size 128x128. Prior to embedding the 80 bit message, we first append a 20 bit checksum and then encode the message using turbo codes[23] to yield a binary message of length 512. Turbo codes provide near optimum performance and are consequently superior to other codes used currently in watermarking (mostly BCH and convolution). The 20 bit checksum is essential in determining the presence of the watermark. At detection if the checksum is verified we can safely say (with probability $\frac{1}{2^{20}}$ of error) that a watermark was embedded and successfully decoded.

Both the Haar wavelet and the Daubechies 4-tap filter were tested. In the case of the Haar wavelet, the algorithm was resistant down to a level of 70% quality factor. Better results were obtained for the 4-tap Daubechies filter where the algorithm is robust down to a level of 50% quality factor and is resistant as well to low and high pass filtering. By resistant, we understand that all the bits are correctly decoded and the checksum verified. We note that for the case of the Daubechies 4-tap filter, some minor modifications must be made to the embedding strategy. In particular, when taking the inverse DWT we obtain a block size which is bigger than the original block. These boundary problems are well known in the wavelet literature. The difficulties are easily overcome by imposing that the extra boundary pixels be constrained to be 0. This is done in practice by setting the appropriate values in Δ_p and Δ_n to 0.1 and -0.1 respectively. We do *not* set these all the way to zero since often this leads to an overly constrained problem.

An example is given in figure 4 where the original image (128×128), watermarked image (Daubechies 4-tap filter) and watermark (difference between original and watermarked) are presented. We observe that the watermark is stronger in textured regions as expected. We note that the watermark is slightly visible along the long vertical edge to the left of the image. This is a limitation of the visibility model which does not take into account the high amount of structure to which the eye is particularly sensitive. In order to overcome this problem more sophisticated models are being developed which take into account the presence of lines in the image. In these regions, the allowable distortion must be reduced. Maximizing strength of the watermark while minimizing visibility in an automatic way over a wide range of images is a delicate problem since each image is unique.

On a Pentium 233MhZ computer the algorithm takes 20 minutes to embed the watermark. This is time is non-negligible. The problem arises from the fact that a formidable optimization problem must be solved at embedding. That is at each block we have $2 * 16 * 16 = 512$ constraints. On the other hand the optimization in each block is independent once the global mask has been calculated. Consequently the algorithm can be carried out in parallel and is of theoretic interest.

(a)



(b) (c)

Figure 2. Original image Lena(a) along with watermarked image (b) and watermark in (c)=(a)-(b).

5. CONCLUSION

In this article we have described a new algorithm for the embedding of wavelet domain watermarks in an optimal manner. The algorithm is extremely flexible in that constraints as determined by masking functions can be easily incorporated in the spatial domain and any linear transform domain may be used although here we considered the special cases of the Haar and Daubechies wavelets. Furthermore we show how to handle problems with truncation in an optimal way and propose the novel approach of modifying all wavelet domain coefficients even though we are only interested in a subset in the LL subband. The algorithm allows for the recovery of 80 bits of information in a small image even after a JPEG compression at quality factor 50%. By contrast the same algorithm implemented in the DCT domain in 8×8 blocks is resistant down to a quality factor of 30%.[10] This suggests that there is much to be gained in matching the embedding strategy with the compression algorithm. Consequently, future work will involve attempting to match the embedding with the EZW wavelet compression algorithm. Work is currently also under way to apply the ideas of[11] so as to make the algorithm resistant to geometric changes as well.

REFERENCES

1. 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.
2. F. A. P. Petitcolas and R. J. Anderson, "Attacks on copyright marking systems," in *2nd International Information Hiding Workshop*, pp. 219–239, (Portland, Oregon, USA), April 1998.
3. G. Voyatzis and I. Pitas, "The use of watermarks in the protection of digital multimedia products," *Proceedings of the IEEE* **87**, July 1999.
4. M. Barni, F. Bartolini, V. Cappellini, A. Piva, and F. Rigacci, "A M.A.P. identification criterion for DCT-based watermarking," in *EUSIPCO'98*, (Rhodes, Greece), September 1998.
5. W. Bender, D. Gruhl, and N. Morimoto, "Method and apparatus for data hiding in images," *U.S. Patent # 5689587* , 1996.
6. R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermark," in *International Conference on Image Processing*, vol. 2, pp. 86–90, IEEE, (Austin, Texas, U.S.A.), 1994.
7. I. Pitas, "A method for signature casting on digital images," in *Proceedings of the IEEE Int. Conf. on Image Processing ICIP-96*, pp. 215–218, (Lausanne, Switzerland), Sept.16-19 1996.
8. E. Koch and J. Zhao, "Towards robust and hidden image copyright labeling," in *Proceedings of 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, (Neos Marmaras, Halkidiki, Greece), June 1995.
9. A. Bors and I. Pitas, "Embedding digital parametric signatures in images," in *EUSIPCO-96*, (Trieste, Italy), September 1996.
10. S. Pereira and T. Pun, "A framework for optimal adaptive DCT watermarks," in *EUSIPCO 2000*, (Tampere, Finland), submitted.
11. S. Pereira and T. Pun, "Fast robust template matching for affine resistant watermarks," in *3rd International Information Hiding Workshop*, (Dreseden, Germany), September 1999.
12. J. Oruanaidh and S. Pereira, "A secure robust digital image watermark," *Electronic Imaging: Processing, Printing and Publishing in Colour* , May 1998.
13. M. Barni, F. Bartolini, A. D. Rosa, and A. Piva, "Capacity of the watermark-channel: How many bits can be hidden within a digital image ?," in *SPIE: Security and Watermarking of Multimedia Contents*, vol. 3657, pp. 437–448, (San Jose, California), January 1999.
14. M. Barni, F. Bartolini, V. Cappellini, A. Lippi, and A. Piva, "A DWT-based technique for spatio-frequency masking of digital signatures," in *Security and Watermarking of Multimedia Contents*, P. W. Wong and E. J. Delp, eds., vol. 3657, pp. 31–39, The Society for Imaging Science and Technology (IS&T) and the International Society for Optical Engineering (SPIE), SPIE, (San Jose, California, U.S.A.), 25–27 Jan. 1999.
15. W. Zhu, Z. Xiong, and Y. Q. Zhang, "Multiresolution watermarking for images and video," *IEEE Transactions on Circuits and Systems for Video Technology* **9**, pp. 545–550, June 1999.

16. I. Cox, J. Killian, T. Leighton, and T. Shamoon., "Secure spread spectrum watermarking for images, audio and video.," in *Proceedings of the IEEE Int. Conf. on Image Processing ICIP-96*, pp. 243–246, (Lausanne, Switzerland), 1996.
17. S. Pereira, S. Voloshynovskiy, and T. Pun, "Effective channel coding for DCT watermarks," in *ICIP 2000*, (Vancouver, Canada), submitted.
18. D. Geman and S. Geman, "Stochastic relaxation, gibbs distributions and the bayesian restorations of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **14**(6), pp. 367–383, 1984.
19. 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.
20. S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. PAMI* **11**, pp. 674–693, 1989.
21. M. Swanson, B. Zhu, and A. Tewfik, "Robust data hiding for images," in *7th IEEE Digital Signal Processing Workshop*, pp. 37–40, (Loen, Norway), September 1996. G:WM1-A23.
22. M. Vetterli and J. Kovacević., *Wavelets and Subband Coding*, Prentice Hall, 1995.
23. C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. Comm.* , pp. 1261–1271, October 1996.