A NOVEL WATERMARKING SCHEME BASED ON BILINEAR INTERPOLATION FOR DIGITAL IMAGES

Vincent Martin, Marie Chabert and Bernard Lacaze

ENSEEIHT/IRIT, National Polytechnic Institute of Toulouse
2 Rue Camichel, BP 7122, 31071 Toulouse Cedex 7, France
vincent.martin, marie.chabert, bernard.lacaze@enseeiht.fr

ABSTRACT

Interpolation often acts as a perturbation in watermarking schemes. In an opposite approach, this article proposes a watermarking algorithm based on interpolation in the spatial domain. The perceptual properties of interpolation allow to generate an imperceptible mark and to substitute this mark to the host image. A theoretical study of the detection and decoding performance is provided, as well as the robustness to attacks compared to classical watermarking algorithms. Imperceptibility and security properties of this scheme are also discussed.

1. INTRODUCTION

Digital watermarking is a promising solution to property and integrity protection of digital data. Watermarking embeds a secret message at the content-level under the constraints of imperceptibility, security and robustness to attacks. Most algorithms are either based on additive embedding or substitution by a codebook element.

In Direct Sequence (DS) Spread Spectrum watermarking [1], the additive mark is the secret message modulated by a pseudo-noise. The message can be later detected by correlation with this pseudo-noise. Insertion can be performed either in the spatial domain (luminance) or in invertible transform domains such as the Discrete Fourier Transform, the Discrete Cosine Transform or the Discrete Wavelet Transform [2]. Classical spread spectrum methods are subject to host interference. However, extensions using knowledge of the host image statistics provide improved performance [3], thanks to Wiener prefiltering at the detector or optimal decoding for a given host image statistical model.

Informed watermarking provides better performance by using knowledge upon both the host image and the detection technique at the embedding [4]. Recent advances focus on random binning inspired from Costa's work in information theory [5]. The inserted mark is selected in a random codebook divided into bins. Each bin is associated to a possible secret message. For a given secret message, the inserted mark is the element of the adequate bin which is closest to the host data. In practice, a reasonably large but suboptimal codebook can be constructed using quantization. A popular scalar quantization-based watermarking scheme is called Scalar Costa Scheme (SCS) [6]. In Spread Transform Scalar Costa Scheme (ST-SCS) [6], robustness to noise is improved by quantizing the projection of the data onto a pseudo-random vector. Moreover, several recent algorithms revisit spread spectrum techniques in the framework of informed embedding. Improved Spread Spectrum [7] proposes a new modulation technique that removes the signal as source of interference. Its simplest form, called Linear Improved Spread Spectrum (LISS), provides the same order of robustness to noise as basic quantization schemes.

The problem of constructing a continuously defined function from given dicrete data is called interpolation. Image interpolation techniques include, in range of increasing performance, nearest-neighbor, bilinear [8], cubic-spline and B-spline [9] interpolation. [8] provides a comparison between the different interpolation techniques in terms of approximation error and execution time.

Usually, interpolation acts as a perturbation in watermarking schemes. Interpolation is involved in most geometrical attacks such as rotation. Indeed, such attacks result in the distortion of the original data coordinates. The interpolation is then used to derive the pixel values on the original discrete grid. Interpolation is also necessary to perform watermarking in a continuous transformed domain such as the Fourier-Mellin domain [10]. More specific algorithms also refer to interpolation. A hierarchical and deterministic secret sharing procedure built on polynomial interpolation can be used to construct a mark provided to an additive watermarking scheme [11]. 3D objects are represented by non-uniform rational B-Splines that provide an insertion domain for substitutive algorithms [12]. Surprisingly, no watermarking embedding scheme has been yet specifically designed on interpolation. This article proposes an interpolation-based substitutive watermarking algorithm in the spatial domain. Section 2 presents W-interp, a watermarking algorithm based on bilinear interpolation. Section 3 provides a theoretical study of its performance in the context of additive white Gaussian noise (AWGN) attack. Section 4 studies the perceptual impact of W-interp on the original image, while Section 5 discusses its security level. Section 6 provides an experimental study of the robustness of W-interp to various attacks, as well as a comparison to the classical watermarking schemes DS, ST-SCS and LISS.

Let denote $M = [m(l)]_{l \in \{1,\dots,L\}}$ the binary antipodal message of size L. L is called the payload. Let denote I the original image, W the mark and I_W the watermarked image. These quantities are handled as matrices as follows:

$$I = [i(n_1, n_2)]_{n_1 \in \{1,\dots,N_1\}, n_2 \in \{1,\dots,N_2\}}$$
$$I_W = I + W = [i_W(n_1, n_2)]_{n_1 \in \{1,\dots,N_1\}, n_2 \in \{1,\dots,N_2\}}$$

The watermarked image I_W is transmitted and possibly attacked, leading to the image $I_W^{'}$. A classification of attacks can be found in [2]. A single noise source $B = [b(n_1, n_2)]_{n_1 \in \{1,\dots,N_1\}, n_2 \in \{1,\dots,N_2\}}$ can model the distortions introduced as well by the transmission channel and by the so-called waveform attacks. Under the assumption of mild attacks, the noise model amounts to the widespread AWGN channel model:

$$I_W^{'} = I_W + B \text{ where } B \sim \mathcal{N}(0, \sigma_B^2)$$

When more severe attacks occur, the noise model may be more sophisticated with possibly non-Gaussian distribution. Such attacks may lead to intractable derivation of the watermarking performance. In such case, the performance is studied through simulations only. The simulations provide the averaged performance on the test image set composed of Lena, Baboon and Fishingboat [13].
For a given I and denoting σ_W^2 the variance of W, let define the document to watermark ratio (DWR) and the watermark to noise ratio (WNR):

$$\text{DWR} = \frac{\sum_{n_1=1}^{N_1}\sum_{n_2=1}^{N_2} i(n_1,n_2)^2}{N_1 N_2 \sigma_W^2},\ \text{WNR} = \frac{\sigma_W^2}{\sigma_B^2}$$

The document to noise ratio is DNR=DWR WNR. DWR (resp. DNR) measures W (resp. B) imperceptibility with respect to the host image. WNR measures transmission noise and attack influence.

2. WATERMARKING ALGORITHM BASED ON INTERPOLATION

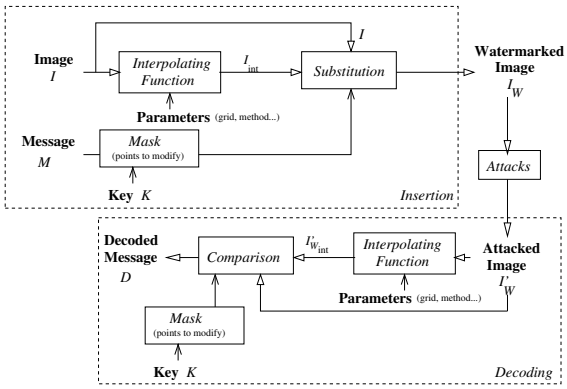

Fig. 1. Watermarking scheme W-interp

W-interp is a substitutive, known-host state [3] watermarking scheme presented in Fig.1. W-interp is a blind watermarking scheme, since I is not used at the decoding. The algorithm is characterized by the choice of the interpolation technique, the interpolation grid and the positions $\mathcal{S}$, outside the interpolation grid, of N_I interpolated pixels. The pixels in $\mathcal{S}$ are such that the resulting interpolation error is large enough (respectively not too large) to guarantee the algorithm robustness (respectively imperceptibility). $P_I = N_I/L$ is the redundancy. $\mathcal{S}$ is divided into L non-overlapping, randomly constructed sets of size P_I: $\mathcal{S} = \mathcal{S}_1 \cup ... \cup \mathcal{S}_L, \mathcal{S}_i \cap \mathcal{S}_j = \emptyset \quad \forall i \neq j$. The values of the pixels in $\mathcal{S}_l$ are modified according to the bit $m(l)$.

At the encoding, the values of these pixels are substituted by an interpolated value (resp. left unchanged) if $m(l) = 1$ (resp. $m(l) = -1$). The interpolation technique is the bilinear interpolation with the chessboard-like grid $(2\mathbb{Z}+1)\times 2\mathbb{Z} \cup 2\mathbb{Z}\times(2\mathbb{Z}+1)$. Bilinear interpolation at the point of continuous coordinates (x, y) is the mean of the 4 nearest neighbors on the grid weighted by their distance from (x, y) (Fig. 2(a)):

$$i_{int}(x,y) = \frac{y-y_1}{y_2-y_1}\left(\frac{x-x_1}{x_2-x_1}i(x_1,y_1) + \frac{x_2-x}{x_2-x_1}i(x_2,y_1)\right) + \frac{y_2-y}{y_2-y_1}\left(\frac{x-x_1}{x_2-x_1}i(x_1,y_2) + \frac{x_2-x}{x_2-x_1}i(x_2,y_2)\right)$$

W-interp substitutes $i(n_1, n_2)$ by

$$\widetilde{i}(n_1,n_2) = i_{int}(n_1+\tau_x(n_1,n_2), n_2+\tau_y(n_1,n_2))$$

where $\tau_x(n_1, n_2)$ and $\tau_y(n_1, n_2)$ are independent random variables uniformly distributed over $]-\frac{1}{2}, +\frac{1}{2}[$ (Fig. 2(b)). The introduction of these random shifts improves the algorithm security level as explained in section 5. The secret key K consists of the interpolated point coordinates $\mathcal{S}$ and the associated random shifts.

The decoding compares I'_W and $\widetilde{I'_W}$. Let denote $R = \widetilde{I'_W} - I'_W$.

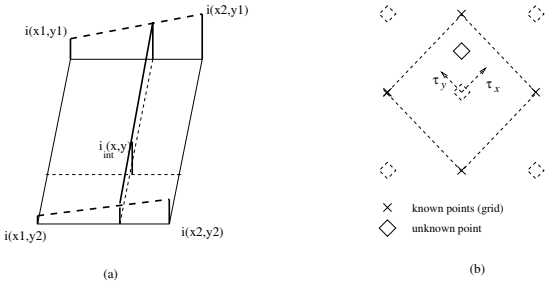

Fig. 2. (a) Bilinear interpolation (b) Random shifts in the coordinates

For a given bit, the mean square error $\rho^2(l) = \sum_{(n_1,n_2)\in\mathcal{S}_l} r(n_1,n_2)^2$ is compared to an image-dependent threshold η. If $\rho^2(l) < \eta$, the decision is $d(l) = +1$, else $d(l) = -1$. η can be chosen empirically as the mean of the decoding results: $\eta = \frac{1}{L}\sum_{l=1}^{L}\rho^2(l)$. However, a theoretical threshold is derived in section 3 under appropriate hypotheses about the interpolation error distribution. W-interp is a host-rejecting watermarking method since in the absence of any attack perfect decoding, thus a rate N_I/N_1N_2, can be achieved. W-interp is not an informed watermarking method since no knowledge about the detection technique is used during the embedding.

3. THEORETICAL PERFORMANCE STUDY

This section theoretically studies the performance in the presence of an AWGN attack. A theoretical detection threshold is derived when the attack parameter σ_B^2 is known. The resulting performance are significantly better than those obtained with the empirical threshold. When the attack parameter σ_B^2 is unknown, the theoretical performance can as well be derived and consistency with simulations is demonstrated.

The histogram of the interpolation error ϵ_I for a given image I suggests a zero-mean generalized Gaussian distribution for this variable. For simplicity, it will be modeled as a zero-mean Gaussian variable of variance $\sigma_{\epsilon_I}^2$ in the following.

3.1. Neyman-Pearson Detector

For simplicity and without any loss of generality, this section considers a single bit mark ($L = 1$) with $m(1) = 1$. The detection problem consists in a binary hypothesis test:
• hypothesis H_0: absence of mark,
• hypothesis H_1: presence of a mark.
Let P_d denote the probability of detection and P_{fa} the probability of false alarm. The optimal Neyman-Pearson detector maximizes P_d for a given P_{fa}. The corresponding test statistics is here $T = \sum_{\mathcal{S}} r(n_1,n_2)^2$ with $r(n_1,n_2) = \epsilon_{I_W}(n_1,n_2) + \epsilon_B(n_1,n_2)$. The variance of ϵ_B can be expressed as $(1+c)\sigma_B^2$, where the constant c depends on the interpolation technique ($c = 4/9$ for the bilinear interpolation and the considered random shifts). As R is zero-mean Gaussian, T follows a χ_P^2 distribution under both hypotheses.
Under hypothesis H_0, $R \sim \mathcal{N}(0, 1.44\sigma_B^2 + \sigma_{\epsilon_I}^2)$.
However, under hypothesis H_1, the substitution at the embedding

leads to $\epsilon_{i_W}(n_1, n_2) = 0$. Consequently, $R \sim \mathcal{N}(0, 1.44\sigma_B^2)$. Let $\mathrm{F}_{\chi^2_P}$ denote the χ^2_P cumulative distribution function. The Neyman-Pearson detector decides H_0 when $T < \eta$ with $\eta = 1.44\sigma_B^2 \mathrm{F}^{-1}_{\chi^2_P}(1-\mathrm{P_{fa}})$ with $P_d = 1 - \mathrm{F}_{\chi^2_P}(\eta/(1.44\sigma_B^2 + \sigma^2_{\epsilon_I}))$. The detection performance is evaluated through the Receiver Operating Characteristic (ROC) curves. These curves display P_d as a function of P_{fa}. W-interp provides good detection results since it is affected by AWGN attacks only for very high values of σ_B^2 (Fig. 3 (a)) or of DWR (Fig. 3 (b)), whereas reasonable values would be DNR=DWR=38 dB.

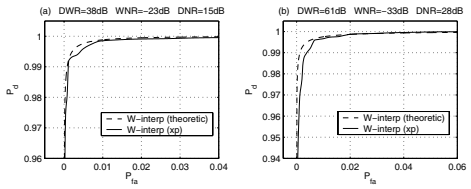

Fig. 3. Receiver Operating Characteristic for W-interp

3.2. Decoding problem

The decoding problem consists in estimating the binary original message from I'_W. The decoding performance is measured experimentally through the bit error rate (BER):

$$\mathrm{BER} = \frac{1 - \sum_{l=1}^{L} \delta(d(l), m(l))}{L}$$

where δ denotes the Kronecker symbol. The optimal decision threshold η_{th} minimizes the BER. Assuming the equiprobability of the binary message symbols, η_{th} is solution of:

$$\frac{1}{1.44\sigma_B^2 + \sigma^2_{\epsilon_I}} \mathrm{f}_{\chi^2_P}\left(\frac{\eta_{\mathrm{id}}}{1.44\sigma_B^2 + \sigma^2_{\epsilon_I}}\right) = \frac{1}{1.44\sigma_B^2} \mathrm{f}_{\chi^2_P}\left(\frac{\eta_{\mathrm{id}}}{1.44\sigma_B^2}\right)$$

Fig. 4 displays the experimental and theoretical BER. The theoretical threshold η_{th} is an improvement to the empirical one η. Fig. 7 demonstrates W-interp decoding robustness to the AWGN attack.

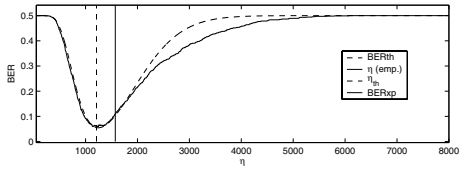

Fig. 4. Choice of η: DWR=38 dB, $L = 1024$, WNR=-10 dB, Fishingboat

4. PERCEPTUAL ANALYSIS

W-interp imperceptibility directly results from the interpolation technique (cf Fig. 5). Imperceptibility is empirically observed for DWR>

Fig. 5. Lena (detail): original, watermarked and watermark, DWR=38 dB

38 dB. A given DWR corresponds to a maximum number N_I of interpolated pixels. This number N_I depends on the variance $\sigma^2_{\epsilon_I}$ and thus on I and on the interpolation technique. For a given DWR, the better the interpolation performance, the greater the redundancy N_I.

$$\mathrm{DWR} = \frac{\sigma_I^2 N}{\sigma^2_{\epsilon_I} N_I}$$

Moreover, note that the interpolation leads to large pixel modifications only in regions of high local variance. Generation of psychovisual masks such as the Noise Visibility Function (NVF) [14] for the DS techniques shows that a modification in these regions is less perceptible (Fig.6). A limitation of the range of shifts $(\tau_x(n_1, n_2), \tau_y(n_1, n_2))$ can also improve the imperceptibility at the expense of the security.

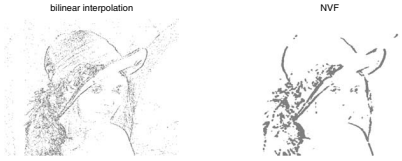

Fig. 6. Greatest modifications for interpolation and NVF (10^4 pts)

5. SECURITY

A lot of attention has been recently paid to the security of watermarking techniques [15]. Attacks on the security aim at uncovering or estimating the secret key K from several observations of data watermarked with K. Without any random shift ($K = \mathcal{S}$), an attacker could have performed a systematic interpolation on the points outside the grid. The interleaved redundant message samples can be estimated. There exist algorithms to uncover $\mathcal{S}$ from the scattered message samples when several observations are available [15]. Thus, the secrecy of the mark location $\mathcal{S}$ is not sufficient.

Suppose now that $\mathcal{S}$ is known to the attacker, but that the interpolation coordinates are shifted by $(\tau_x(n_1, n_2), \tau_y(n_1, n_2))$. The difference between the interpolation results for different shift values can be modeled as zero-mean Gaussian with variance σ_S^2. If $i(n_1, n_2) \in \mathcal{S}_l$ with $m(l) = -1$ now $R \sim \mathcal{N}(0, \sigma_S^2)$. The case $m(l) = 1$ is unchanged ($R \sim \mathcal{N}(0, \sigma^2_{\epsilon_I})$). The analysis of section 3 can be used to derive the theoretical decoding performance. However, the two hypotheses are now very close since $\sigma_S^2 = 0.85\sigma^2_{\epsilon_I}$ in average. For reasonable DWR values, the BER is very poor and the systematic interpolation attack is inefficient (for instance, for DWR=38 dB and $L = 1024$, BER=0.43).

6. ROBUSTNESS AND PERFORMANCE COMPARISON

In this section, W-interp is compared in terms of robustness to the classical algorithms DS [3], DS with Wiener prefiltering (DS+W) [3], ST-SCS [6] and LISS [7]. For computational reasons, the BER is sometimes poor (BER= 10^{-2}). It could be improved by decreasing L (more redundancy) or increasing DWR (provided that the imperceptibility constraint is respected). The simulations illustrate two different scenarios. The AWGN attack scenario assumes that σ_B^2 is known at the embedding (Fig. 7). This hypothesis allows for distortion compensation in ST-SCS and LISS and theoretical thresholding in W-interp. The attack channel is unknown in the second (more practical) scenario. The empirical threshold is used in W-interp, whereas no distortion compensation is performed in ST-SCS and LISS. However, the loss of performance in ST-SCS and LISS is low for reasonable values of the spreading factor $P = N_1 N_2 / L$ and WNR [7],[16]. Indeed, the optimal distortion compensation parameter (α in [6] or λ in [7]) is close to 1 when PWNR is large.

While the performance of DS, and to a lesser extent of DS+W, are hampered by the host image interference, ST-SCS and LISS are unaffected by the host image and provide excellent performance facing AWGN (Fig. 7). However, on the whole, ST-SCS and LISS seem vulnerable to attacks, while DS+W is more robust. The use of

quantization results in a high vulnerability of ST-SCS to valumetric attacks such as histogram equalization (Fig. 8). When severe attacks occur, the host-interference rejection in LISS disappears and its performance falls close to that of simple DS (Fig. 9, Fig. 10).

Thanks to its host-rejecting property, W-interp is far more robust to AWGN than DS and DS+W for reasonable noise variance (in Fig. 7, when WNR<0, the noise gets perceptible). In general, LISS and DS+W can outperform W-interp for a large redundancy P, while W-interp is more appropriate for a large payload ($L > 300$). Unlike the two informed watermarking methods, W-interp provides also very good robustness to various attacks such as histogram equalization (Fig. 8), JPEG compression (Fig. 9) and to denoising attacks such as Wiener filtering (Fig. 10). Indeed, the mark embedded in ST-SCS, LISS and W-interp can be modeled as an additive noise. However, in W-interp it is highly correlated with the host image, thus more difficult to remove. W-interp, like all the considered algorithms, is vulnerable to geometric attacks such as rotation, even of a very small angle (Fig. 11).

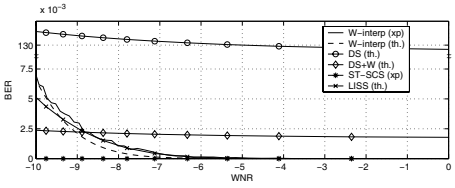

Fig. 7. Robustness to AWGN, $L = 300$, DWR=38 dB

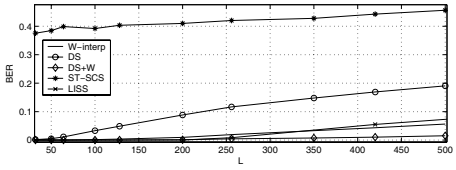

Fig. 8. Robustness to histogram equalization, DWR=38 dB

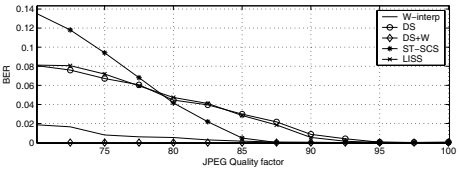

Fig. 9. Robustness to JPEG compression, $L = 64$, DWR=38dB

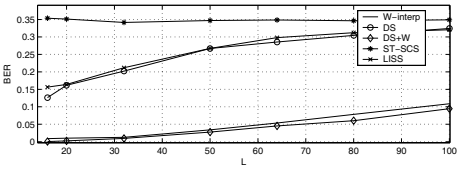

Fig. 10. Robustness to denoising, DWR=38 dB

7. CONCLUSION

A blind, host-interference rejecting watermarking scheme has been proposed. Since interpolation acts like a low-pass filter, W-interp is a particular case of embedding in the high pass coefficients. Interpolation perceptual properties in the spatial domain allow for embedding without need for a perceptual mask. A rigorous study comprising security, imperceptibility and robustness has been presented. The simulations have shown a good robustness of the proposed method to the classical waveform attacks. The theoretical performance and choice of the image-dependent detection or decoding threshold have been derived for the AWGN attack. Robustness to other attacks can also be improved thanks to this kind of study. Specific countermeasures such as interpolation on a geometrically distorted grid will be developed to improve the robustness of W-interp to geometrical attacks. Moreover, this algorithm provides the basis for a new class of watermarking methods based on interpolation. For instance, other interpolation techniques can be used to improve the compromise between perceptual quality and performance. Among them, spline interpolation is especially under study.

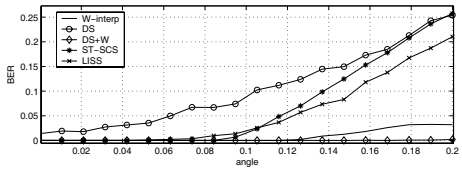

Fig. 11. Robustness to rotation, $L = 64$, DWR=38 dB

8. REFERENCES

[1] I.J. Cox, J. Kilian, F.T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *Image Processing, IEEE Trans. on*, vol. 6, no. 12, pp. 1673–1687, 1997.

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

[3] J.R. Hernández and F. Pérez-González, "Statistical analysis of watermarking schemes for copyright protection of images," *IEEE Proc., Special Issue on Identification and Protection of Multimedia Information*, vol. 87, no. 7, pp. 1142–1166, 1999.

[4] M.L. Miller, I.J. Cox, and J.A. Bloom, "Informed embedding: Exploiting image and detector information during watermark insertion," *IEEE Int. Conf. on Image Processing - ICIP*, vol. 3, pp. 1–4, 2000.

[5] P. Moulin and R. Koetter, "Data-hiding codes," *Proc. of the IEEE*, vol. 93, no. 12, pp. 2083–2127, 2005.

[6] J.J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, "Scalar Costa Scheme for Information Embedding," *IEEE Trans. on Signal Processing*, vol. 51, no. 4, pp. 1003–1019, 2003.

[7] H.S. Malvar and D.A.F. Florêncio, "Improved spread spectrum: a new modulation technique for robust watermarking," *Signal Processing, IEEE Trans. on*, vol. 51, no. 4, pp. 898–905, 2003.

[8] P. Thévenaz, T. Blu, and M. Unser, "Image interpolation and resampling," in *Handbook of Medical Imaging, Processing and Analysis*, I.N. Bankman, Ed., chapter 25, pp. 393–420. Academic Press, San Diego CA, USA, 2000.

[9] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, 1999.

[10] J.J.K. Ó Ruanaith and T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking," *Signal Proc.*, vol. 66, no. 3, pp. 303–317, 1998.

[11] G. Boato, C. Fontanari, and F. Melgani, "Hierarchical deterministic image watermarking via polynomial interpolation," *Proc. of ICIP*, 2005.

[12] R. Ohbuchi, H. Masuda, and M. Aono, "A shape-preserving data embedding algorithm for nurbs curves and surfaces," *Proc. of the Computer Graphics International (CGI)*, pp. 170–177, 1999.

[13] Images, "www.petitcolas.net/fabien/watermarking/image_database/," .

[14] S. Voloshynovskiy, A. Herrigel, N. Baumgartner, and T. Pun, "A stochastic approach to content adaptive digital image watermarking," *International Workshop on Information Hiding*, pp. 212–236, 1999.

[15] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security : Theory and practice," *IEEE Trans. on Signal Processing, Special Issue on Content Protection*, vol. 53, no. 10, pp. 3976–3975, 2005.

[16] F. Pérez-González, F. Balado, and J.R. Hernández Martin, "Performance analysis of existing and new methods for data hiding with known-host information in additive channels," *IEEE Trans. on Signal Processing*, vol. 51, no. 4, pp. 960–980, 2003.