CONTENT ADAPTIVE WATERMARKING BASED ON A STOCHASTIC MULTIRESOLUTION IMAGE MODELING

Sviatoslav Voloshynovskiy, Frédéric Deguillaume, and Thierry Pun
CUI, University of Geneva,
24 rue du Général Dufour, CH 1211 Geneva 4, Switzerland
e-mail: {svolos,Frederic.Deguillaume,Thierry.Pun}@cui.unige.ch

ABSTRACT

In this paper, a wavelet domain robust watermarking technique for still images is presented. The proposed watermarking algorithm is based on 3 key aspects. First, message encoding is accomplished based on iterative error correction codes to reach channel capacity with reasonable decoder complexity. Secondly, watermark embedding is performed in the wavelet domain using a stochastically driven perceptual criterion to provide watermark invisibility. Thirdly, a new principle of watermark spatial allocation, based on the watermark magnitude spectrum, is proposed to recover from general affine geometrical distortions.

1 INTRODUCTION

These last years, the rapidly growing digital multimedia market has revealed an urgent need for efficient copyright protection mechanisms. Two of the most important requirements for watermarking algorithms are *visibility* and *robustness*. To satisfy these two conflicting requirements, attention should primarily be paid to the proper choice of a message encoding scheme, to the use of a perceptual masking criterion, and to a scheme for compensating geometrical attacks.

In earlier solutions the information to be embedded was encoded using either M-ary modulation [4], or mostly algebraic error correction codes (ECC) [3]. The M-ary modulation is potentially able to reach the channel capacity. However, it requires $M \rightarrow \infty$ which can be difficult to implement in a practical watermark demodulator. Therefore, to warrant a practically tractable solution to this problem we propose to use $low - density\ parity - check$ (LDPC) codes [2] or *turbo codes* [1] that are the subject of recent intensive research. The maximum rate at which they can be used is known to be bounded below channel capacity [2]. However, the existence of simple iterative decoding scheme and their outstanding error performance more than compensates this disadvantage. In this paper we will use turbo codes.

Here we propose an extension of our previous work [7] aiming at including a multiresolution paradigm in the stochastic framework described earlier to benefit from a modulation transfer function (MTF) of the human visual system (HVS). This practically means that different watermark strengthes are assigned to different image sub-bands. Such a modification leads to a non-white spectrum of the watermark matched with the MTF, which was not the case for the spatial domain based version of the Noise Visibility Function (NVF) used in [7]. The second reason for using wavelet domain embedding is to exploit the anisotropy of the HVS in different spatial directions in the perceptual mask. The spatial domain version of the NVF [7] uses an isotropic image decomposition. In the wavelet domain the image coefficients in 3 basic spatial directions, i.e. vertical, horizontal and diagonal, are used to better reflect the anisotropy properties of the HVS. As a result, the watermark strength varies for different orientations in the proposed mask.

Robustness to geometrical distortions relied on the use of either a transform invariant domain [5], or of an additional template [6], or of an Autocorrelation Function (ACF) of the watermark itself [4]. The ACF approach is known to have a number of advantages in comparison with the two other methods. In [4] the watermark is replicated in the image in order to create 4 repetitions of the same watermark. This enables to have 9 peaks in the ACF that are used to estimate the undergone geometrical transformations. The descending character of the ACF peaks shaped by a triangular envelope reduces the robustness of this approach to the combination of geometrical attacks and lossy compression. The need for computing two fast Fourier Transforms (FT) of double image size to estimate the ACF creates some problems for real time application in the case of large images. Therefore, to overcome the above difficulties we propose to use a periodical block allocation of a watermark. The known fact that periodical signals have discrete magnitude spectrum makes it possible to obtain a regular grid of reference points that can easily be employed for recovering from general affine transformations. The existence of many peaks in the magnitude spectrum of the periodically repeated wa-

termark increases the probability to detect geometrical transform even after lossy compression attack. This fact indicates the enhanced robustness of the proposed approach. Secondly, it is more difficult to remove the peaks in the magnitude spectrum based on a local interpolation in comparison with a template scheme. Such an attack would create considerable visible distortions in the attacked image.

This paper proposes a new content adaptive wavelet domain robust watermarking technique. The proposed message encoding, the watermark spatial allocation and embedding in wavelet domain according to the NVF is considered in section 2. An efficient method for compensating general affine geometrical distortions is also presented, without need of any additional template, nor explicit ACF computation. Section 3 explains the watermark estimation and overviews the determination of geometrical distortions.

2 MESSAGE ENCODING AND WATERMARK EMBEDDING

2.1 Message encoding

A message (the copyright information) $b = (b_1, ..., b_L)^T$ is first encoded in a codeword $c = (c_1, ..., c_K)^T$ using either M-ary modulation or ECC. The codeword is then mapped from $\{0, 1\}$ to $\{-1, 1\}$ and encrypted by multiplying it with a key-dependent sequence p, followed by a spreading over a square block of size $N_1 \times N_1$ with some density D using a secret key. This block is upsampled by the factor 2 to receive the low-pass watermark and then flipped and copied once in each direction, producing a symmetric block of size $4N_1 \times 4N_1$. Finally, the $4N_1 \times 4N_1$ block is repeated over the whole image size, resulting in a symmetrical and periodical watermark with periods $T_1 = T_2 = 4N_1$. In our modeling we use $L = 64$ bit message that was encoded using its turbo code ($K = 132$). The scheme is very flexible with respect to the encoding, and to compare the performance of the proposed ECC we use also binary modulation ($M = 2$), that is typically used by watermarking community, and BCH code.

2.2 Stochastic multiresolution image modeling and watermark embedding

To embed the resulting watermark in a cover image a linear additive scheme is used in the wavelet domain. Both the cover image and the watermark are first decomposed into a multiresolution sub-band pyramid using the Forward Wavelet Transform (FWT). $N_w = 5$ levels are used for the FWT based on the Daubechies 8-tap filter. Each scale has 3 components corresponding to distinct *orientations* l, for vertical (V), horizontal (H), and diagonal (D) directions. The lowest scale $k = N_w + 1$ consists of only a low-pass component. Figure 1 shows the pyramid of the NVF and of the watermark. The watermarking process is applied and adapted to each k, l sub-band

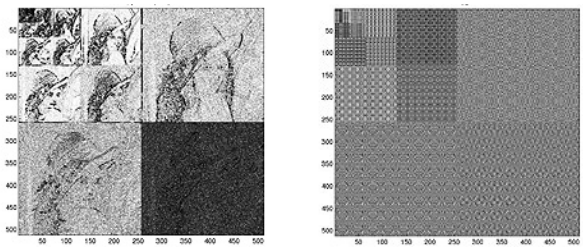

Figure 1: Left: NVF computed for each sub-band of the cover image FWT. Right: watermark FWT pyramid.

component separately. Finally, the stego image is reconstructed by computing the Inverse Wavelet Transform (IWT) of the marked image pyramid.

Perceptual masking is modeled based on a NVF, of pixel (i, j), for each sub-band component k, l:

$$NVF_{k,l}(i,j) = \frac{w_{k,l}(i,j)}{w_{k,l}(i,j) + \sigma^2_{\widetilde{x}_{k,l}}}. \quad (1)$$

The NVF is based on a Stationary Generalized Gaussian (SGG) model described in [7]. $\sigma^2_{\widetilde{x}_{k,l}}$ is the global variance of the wavelet image coefficients from sub-band (k, l), and $w_{k,l}(i, j)$ can be written as $w_{k,l}(i,j) = \gamma_{k,l} \cdot [\eta(\gamma_{k,l})]^{\gamma_{k,l}} \frac{1}{\|r_{k,l}(i,j)\|^{2-\gamma_{k,l}}}$, with $\eta(\gamma) = \sqrt{\frac{\Gamma(\frac{3}{\gamma})}{\Gamma(\frac{1}{\gamma})}}$ where Γ is the Gamma function, and $r_{k,l}(i,j) = \frac{\widetilde{x}_{k,l}(i,j)}{\sigma_{\widetilde{x}_{k,l}}}$ where $\widetilde{x}_{k,l}(i,j)$ are the wavelet cover image coefficients. Figure 1 shows the NVFs of the cover image and watermark pyramids. Finally the weighted watermark is added to the cover image:

$$\begin{aligned} \widetilde{y}_{k,l}(i,j) &= \widetilde{x}_{k,l}(i,j) \\ &+ (1 - NVF_{k,l}(i,j)) \cdot \widetilde{w}_{k,l}(i,j) \cdot S^e_{k,l} \\ &+ NVF_{k,l}(i,j) \cdot \widetilde{w}_{k,l}(i,j) \cdot S^f_{k,l} \end{aligned} \quad (2)$$

where $\widetilde{y}_{k,l}$ are the obtained stego wavelet components and $\widetilde{w}_{k,l}$ are the watermark wavelet components. $S^e{}_{k,l}$ is an embedding strength for the edges and textures, and $S^f{}_{k,l}$ is a strength for the flat regions. Visual masking is ensured first by choosing $S^e{}_{k,l}$ greater than $S^f{}_{k,l}$ for edges and textures hiding, and second by using adapted strengths for each resolution, and even for each orientation. An example of practically used embedding parameters, considering cover image pixels values in the range $[0, 255]$, are:

$$S_e = \begin{bmatrix} 18 & 18 & 2 & 0 \\ 4 & 4 & 6 & 0 \\ 6 & 6 & 8 & 0 \\ 2 & 2 & 4 & 0 \\ 5 & 5 & 7 & 2 \end{bmatrix} \quad S_f = \begin{bmatrix} 0.1 & 0.1 & 0.2 & 0 \\ 0.2 & 0.2 & 0.5 & 0 \\ 0.5 & 0.5 & 1 & 0 \\ 2 & 2 & 4 & 0 \\ 4 & 4 & 6 & 2 \end{bmatrix}$$

where rows denote decreasing resolutions, and columns each orientation.

3 WATERMARK EXTRACTION AND MESSAGE DECODING

The embedded watermark is first estimated from the stego image. Secondly, geometric distortions which may have occurred are retrieved and compensated, by analyzing the FT magnitude of the estimated watermark. The tiled blocks are then averaged in order to get an estimate of the embedded redundant sequence according to the ML-estimate for Gaussian channel. Finally, the message is decrypted and decoded.

3.1 Watermark estimation

To estimate the watermark a maximum *a posteriori* probability (MAP) estimate is used:

$$\widehat{w} = \arg\max_{w \in \Re^N} \{ p_X (y' \mid w) \cdot p_W (w) \} \tag{3}$$

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

$$\widehat{w} = \frac{R_w}{R_w + R_x} (y' - \overline{y}') \tag{4}$$

where it is assumed $\overline{y}' \approx \overline{x}$ to be a local mean, and $\widehat{R}_x = \max\left(0, \widehat{R}_y - R_w\right)$ is the ML estimate of the image covariance matrix.

An important issue is the estimation of the watermark covariance matrix in the above estimate. This can be done based on the available copy of the stego image. However, the severe distortions due to lossy JPEG compression could destroy the information about the texture masking that was used for the watermark embedding, and a histogram modification attack could damage the relevant information about contrast sensitivity masking. Since no reliable information about perceptual mask is available after these attacks, we propose to use a global estimate of the watermark strength based on the available copy of the attacked image. This practically means that we assume spatial stationarity of the watermark $\widehat{R}_w = \widehat{\sigma}_w^2 I$. To estimate a global watermark variance we use the following formula:

$$\widehat{\sigma}_w^2 = \frac{1}{N^2} \sum_{m=1}^{N} \sum_{n=1}^{N} \widehat{\sigma}_y^2 (m, n) \tag{5}$$

where $\widehat{\sigma}_y^2 (m, n)$ is a local variance of the stego image in the coordinates (m, n), for an image of size $N \times M$. The estimate (5) is a global mean value of the watermark variance. Obviously, other robust versions of (5) such as a robust median estimate of the variance could be applied here.

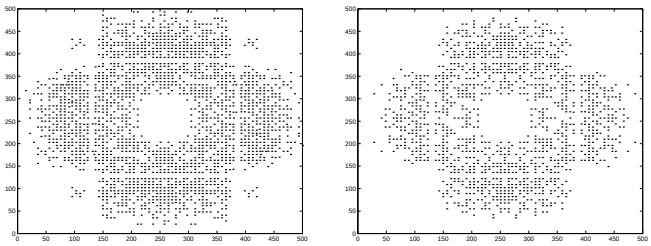

Figure 2: Extracted peaks from magnitude spectrum. Left: peaks computed from the embedded watermark. Right: peaks computed from the Wiener-based estimate of the watermark.

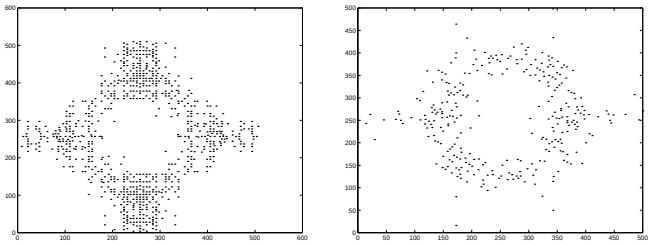

Figure 3: Extracted peaks from magnitude spectrum after JPEG compression with QF = 50%. Left: without geometric distortion. Right: after a rotation of 37^o and autocropping.

3.2 Determining affine geometrical distortions

We compute $|\Im(\widehat{w})|^2$ from the estimated watermark, where $\Im$ is the discrete FT. Due to the periodicity of the embedded information, the estimated watermark spectrum possesses a discrete structure. Assuming that the watermark is white noise within the block, the spectrum of the watermark will additionally be uniform. Therefore, $|\Im(\widehat{w})|^2$ shows aligned and regularly spaced peaks. For a T_1, T_2-periodical watermark $\widehat{w}$, peaks will have periods $\frac{1}{T_1} M_1$ and $\frac{1}{T_2} M_2$ for a 2-D FT domain of size $M_1 \times M_2$. If an affine distortion was applied to the stego image, the peaks layout will be rescaled, rotated and/or sheared, but alignments will be preserved. Therefore, it is easy to estimate any affine geometrical distortion from these peaks by fitting alignments and estimating periods. Figure 2 shows peaks extracted from the magnitude spectrum of the watermark $|\Im(\widehat{w})|^2$. In the left figure, the real embedded watermark w is considered by using the knowledge of the cover image in a non-oblivious approach, while in the right figure the Wiener predicted watermark $\widehat{w}$ is taken. Therefore, these peaks can be extracted from the stego data with high quality from the estimated watermark without knowledge of the cover image. Figure 3 shows peaks extracted after lossy compression, without and with geometric distortions. In experiments peaks could be properly extracted from JPEG compressed images with a quality factor (QF) up to 50%. At the time of publication, no known watermarking method is able to resist to affine transforms combined with such a compression.

3.3 Message decoding

Assuming that attack, prediction and extraction errors could be modeled as additive Gaussian, the detector is designed using the ML formulation for the detection of a known signal (projection sets are known due to the key) in Gaussian noise, that results in a correlator detector

$$r = \langle \widehat{w}, p \rangle \,. \tag{6}$$

In more general cases, the detector should be designed for stationary non-Gaussian noise or for the non-stationary Gaussian case; this is the subject of our ongoing research. Finally, given an observation vector r, the optimum decoder that minimizes the conditional probability of error assuming that all codewords b are equiprobable is given by the ML decoder:

$$\widehat{b} = \arg\max_{\widetilde{b}} p\left(r \mid \widetilde{b}, x\right). \tag{7}$$

Based on the central limit theorem (CLT) most researchers assume that the observed vector r can be accurately approximated as the output of an additive Gaussian channel noise [4, 3] for a large sample space. We use BCJR decoder [1] for the turbo code.

4 RESULTS OF COMPUTER SIMULATION

Experiments have been performed with respect to JPEG compression for QF from 1 to 100. Results have been averaged over 5 test images from the Stirmark benchmark and 50 different keys. The message was encoded using 3 different methods, which are binary antipodal signaling, BCH encoding and turbo encoding at rate 1/2. Figure 4 shows the probability of error of the decoded message relatively to QF, plotted for each encoding method. In the left figure the watermark was estimated from the stego image using the MAP estimate. In the right figure a non-oblivious approach was used, exploiting the knowledge of the original image giving us an upper bound for the reliable watermark decoding. Turbo code gives the best results at low-quality JPEG compression, allowing the decoding of the message with no error up to $QF = 9$, when decoding with the original garanties up to $QF = 8$.

5 Conclusion

We proposed a new approach for content adaptive image watermarking, taking into account edge and texture masking as well as the multiresolution sensitivity of the HVS. A stochastic model is used for both watermark embedding and extraction, combined with the wavelet decomposition. In contrast to earlier non adaptive techniques, we can increase the strength of the watermark while keeping it under the threshold of perceptibility. A new class of the ECC is proposed to be used for message encoding that results in a robust scheme simultaneously satisfying visibility and robustness constraints.

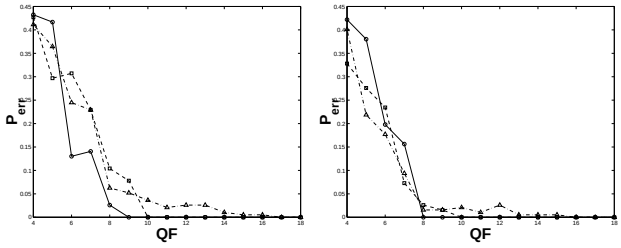

Figure 4: Probability of error P_{err} for a message of 64 bits, for JPEG compression with various QF and various encoding methods. Only $QF = 4 \cdots 18$ are shown, since for higher values P_{err} was equal to 0. Left: oblivious approach. Right: non-oblivious approach. Triangles/dashdot: binary antipodal signaling. Squares/dashed: BCH encoding, and Circles/solid: turbo code encoding, both with rate 1/2.

We also introduced an efficient method for the estimation of general affine distortions, based on the properties of the FT, needing no additional template, and even no explicit ACF. Due to the wavelet decomposition approach used, our method is straightforwardly applicable to JPEG2000.

References

[1] C. Berrou and A. Glavieux. Near optimum error correcting coding and decoding: turbo-codes. *IEEE Trans. Comm.*, pages 1261–1271, October 1996.

[2] R. Gallager. Low-density parity-check codes. In *IRE Transactions on Information Theory*, January 1962.

[3] J. R. Hernández, F. Pérez-González, J. M. Rodríguez, and G. Nieto. The impact of channel coding on the performance of spatial watermarking for copyright protection. *in Proc. ICASSP'98*, 5:2973–2976, May 1998.

[4] M. Kutter. *Digital image watermarking: hiding information in images.* PhD thesis, EPFL, Lausanne, Switzerland, August 1999.

[5] J. Oruanaidh and T. Pun. Rotation, scale and translation invariant spread spectrum digital image watermarking. *Signal Processing*, 66(3):303–317, 1998.

[6] S. Pereira, J. J. K. Ó Ruanaidh, F. Deguillaume, G. Csurka, and T. Pun. Template based recovery of Fourier-based watermarks using Log-polar and Log-log maps. In *Int. Conference on Multimedia Computing and Systems, Special Session on Multimedia Data Security and Watermarking*, Juin 1999.

[7] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T. Pun. A stochastic approach to content adaptive digital image watermarking. In *Lecture Notes in Computer Science: Third International Workshop on Information Hiding*, volume 1768, pages 211–236, Dresden, Germany, September/October 1999. Springer.