

FIDELITY AND ROBUSTNESS ANALYSIS OF IMAGE ADAPTIVE DWT-BASED WATERMARKING SCHEMES

Franco Del Colle and Juan Carlos Gómez
Laboratory for System Dynamics and Signal Processing
FCEIA, Universidad Nacional de Rosario, Argentina

Keywords: Image Digital Watermarking, Discrete Wavelet Transform, Perceptual Distortion Metrics.

Abstract: An Image Adaptive Watermarking method based on the Discrete Wavelet Transform is presented in this paper. The robustness and fidelity of the proposed method are evaluated and the method is compared to state-of-the-art watermarking techniques available in the literature. For the evaluation of watermark transparency, an image *fidelity factor* based on a perceptual distortion metric is introduced. This new metric allows a perceptually aware objective quantification of image fidelity.

1 INTRODUCTION

Digital Watermarking refers to techniques that are used to protect digital data by imperceptibly embedding information (the watermark) into the original data in such a way that always remains present. A set of requirements should be met by any watermarking technique. The main requirements are *perceptual transparency*, *payload of the watermark* and *robustness*. Perceptual transparency refers to the property of the watermark of been imperceptible to the human eye by simple inspection. Payload of the watermark refers to the amount of information stored in the watermark. Finally, robustness refers to the capacity of the watermark to remain detectable after alterations due to processing techniques or intentional attacks. Good overviews on the state of the art of classical watermarking techniques can be found in the recent textbooks (Barni and Bartolini, 2004) and (Cox et al., 2002), and in (Podilchuk and Delp, 2001), (Petitcolas, 2000) and the references therein.

Several techniques have been proposed in the literature for the watermarking of still images. The watermark embedding is achieved by first extracting a set of features from the image to be watermarked, and then modifying them according to an embedding rule. Different approaches have been proposed and they can be classified taking into account: (i) the domain in which the watermark is being embedded, leading to a classification in spatial domain and transform domain techniques; (ii) the watermark adaptation to the particular image leading to Image Adaptive

Watermarking (IAW) methods ((Barni et al., 2001), (Podilchuk and Delp, 2001)), and Image Independent Watermarking (IIW) methods (Cox et al., 1997). This paper will focus on Image Adaptive Discrete Wavelet Transform (IADWT) domain watermarking techniques since they have proved to yield better results regarding transparency and robustness.

In this paper, a watermarking scheme in the DWT domain is proposed as a modification of the one in (Podilchuk and Zeng, 1998). This is done in section 2. A new criterion for watermark transparency evaluation based on perceptual distortion metrics is proposed in section 3. A watermark robustness evaluation criterion is introduced in section 4. A comparison between the proposed method and the one in (Podilchuk and Zeng, 1998) during insertion and detection is performed in section 5. Finally, some concluding remarks are given in section 6.

2 IADWT WATERMARKING

Image adaptive watermarking methods make use of visual models in order to determine the maximum length and power of the watermark according to the image capacity to "hide information" without being perceptible. This capacity is calculated by means of the so called Just Noticeable Differences (JND) thresholds, which measure the smallest difference between images which is perceptually detectable by the human eye. In the DWT domain, these thresholds al-

lows to determine the location of the transform coefficients and the amount that they can vary without being noticeable in the spatial domain.

In the watermark embedding scheme in (Podilchuk and Zeng, 1998), the watermark is modulated by the JND, and the coefficients are marked whenever they are greater than the JND threshold, *i.e.*

$$\widehat{X}^w(u, v) = \begin{cases} \widehat{X}(u, v) + J(u, v)w(\ell) & \widehat{X}(u, v) > J(u, v) \\ \widehat{X}(u, v) & \text{otherwise} \end{cases} \quad (1)$$

where $\widehat{X}(u, v)$ and $\widehat{X}^w(u, v)$ are the DWT coefficients of the original image and the watermarked image respectively, $J(u, v)$ is the JND matrix at the u, v frequency in the DWT domain, and $w(\ell)$ is a zero mean, unit variance, normally distributed random sequence. In this way, the watermark weighted by the JND thresholds has lower power than the maximum power that can be inserted without causing noticeable distortions in the image.

The JND thresholds are computed based on the perceptual model of the Human Visual System (HVS) introduced in (Watson et al., 1997). This model takes into account frequency sensitivity, local luminance and contrast masking effects to determine an image-dependent quantization matrix, which provides the maximum possible quantization error in the DWT coefficients which is not perceptible by the HVS.

In the watermark detection scheme the JND are calculated using the original image, then, the DWT coefficients of the original image are subtracted from the ones of the image suspected to be watermarked, and this difference is divided by the JND in order to obtain the received watermark. The correlation between the extracted watermark and the original one is then performed and the maximum value is determined, *i.e.*

$$w_e(\ell) = \frac{\widehat{X}^w(u, v) - \widehat{X}(u, v)}{J(u, v)} \quad \text{if } \widehat{X}(u, v) > J(u, v) \quad (2)$$

$$r_{w, w_e} = \frac{w_e(\ell) * w(-\ell)}{E_{w_e} \cdot E_w} \quad (3)$$

where E_{w_e} and E_w are the energies of the extracted watermark sequence, $w_e(\ell)$, and the original watermark sequence, $w(\ell)$, respectively.

The following modification to the IADWT insertion scheme in (1) can be introduced

$$\widehat{X}^w(u, v) = \begin{cases} \widehat{X}(u, v) + J(u, v)w(\ell) & \widehat{X}(u, v) > J(u, v) > T \\ \widehat{X}(u, v) & \text{otherwise} \end{cases} \quad (4)$$

This modified insertion scheme will be hereafter denoted as IADWT_T. The *rationale* for the constrain $J(u, v) > T$ is that when the JND thresholds

are too small, the magnitude of the marking term in (4) becomes negligible. The introduction of the lower bound T has then the advantage of reducing the watermark length, improving in this way the fidelity and also the robustness, as will be illustrated in section 5.

The detection scheme in (2) has to be modified to take into account the modification in the insertion scheme, as follows

$$w_e(\ell) = \frac{\widehat{X}^w(u, v) - \widehat{X}(u, v)}{J(u, v)} \quad \text{if } \widehat{X}(u, v) > J(u, v) > T \quad (5)$$

3 FIDELITY EVALUATION

In the evaluation of image watermarking methods it may be of interest to judge the fidelity of the watermarked image, that is the similarity between the images before and after the watermark insertion. To avoid the dependence on human judgement in the fidelity evaluation, it would be desirable to objectively quantify the fidelity of watermarked images based on a metric that takes into account the characteristics of the HVS.

Image fidelity metrics appeared in the context of imaging applications to quantify the distortion in images produced by image processing algorithms such as compression, halftoning, printing, etc. Different metrics have been proposed in the literature to measure image distortion (Winkler, 2005), (Zhang et al., 2004)). Among them, the ones based on the characteristics of the HVS have proved to deliver the best results, since they take into account the different sensitivity of the human eye for color discrimination, contrast masking and texture masking.

A metric widely used to measure image fidelity is the S-CIELAB metric (Zhang, 1996) (based on CIE94 (CIE: International Commission on Illumination, 1995)) that specifies how to transform physical image measurements into perceptual differences (ΔE_{94}) and incorporates the different spatial sensitivities of the three opponent color channels. In (Zhang and Wandell, 1998) the authors test how well the S-CIELAB metric predicts image fidelity for a set of color images by comparison with the widely used root mean square error (RMSE) computed in un-calibrated RGB values.

Since the S-CIELAB metric takes into account the perceptual characteristics of the HVS, such as color discrimination, different spatial sensitivity, etc., this metric represents a natural choice for the quantification, in an objective way, of the fidelity of the watermarked image. To the best of the authors' knowledge,



Figure 1: Left: Original Image. Center: Noisy Image. Right: Distortion Map.

this perceptual metric has not been considered before in the context of watermark fidelity evaluation.

To illustrate the use of the S-CIELAB metric, a region of the left image in Figure 1, delimited by the white square in the center image, is corrupted with zero mean unit variance additive Gaussian white noise. The right image shows the image distortion map corresponding to the noise corrupted image, where the S-CIELAB ΔE_{94} values are shown with a grayscale color map. The pixels where the S-CIELAB ΔE_{94} values are above a specified threshold are then marked in green. For reference purposes the edges of the original image are displayed in white. Note the reader that there are no perceptible differences between the original and corrupted images.

The idea in this paper is to use distortion maps to compare watermarked image fidelity for the two insertion methods described in section 2. Due to the spatial distribution of the S-CIELAB ΔE_{94} errors in the distortion maps (the green marks in the right image of Figure 1) it is difficult to make a comparison of the different methods. To provide a unique parameter quantifying this fidelity, a pooling of the S-CIELAB ΔE_{94} error is proposed as follows:

$$\mathcal{F} \triangleq \left(1 - \frac{\sum_{i=1}^M \sum_{j=1}^N S\Delta E_{94}(i,j) \text{Mask}(i,j)}{\sum_{i=1}^M \sum_{j=1}^N \sqrt{X_L(i,j)^2 + X_a(i,j)^2 + X_b(i,j)^2}} \right) \times 100 \quad (6)$$

where M and N are the rows and columns of the image, $S\Delta E_{94}$ is a matrix with the values of the S-CIELAB ΔE_{94} errors for each pixel, *i.e.* the image distortion map, Mask is a mask with ones in the positions where the S-CIELAB ΔE_{94} errors are above the threshold and zeros otherwise, and X_L , X_a and X_b are the image components in the *Lab* color space.

The performance of the proposed metric is compared with that of a standard non perceptual metric based on the RMS error. This metric, namely Root Mean Square Fit (RMS_{FIT}), is obtained by making a pooling of the RMS errors, resulting in:

$$RMS_{FIT} \triangleq \left(1 - \frac{\sum_{i=1}^M \sum_{j=1}^N \sqrt{\Delta X_R(i,j)^2 + \Delta X_G(i,j)^2 + \Delta X_B(i,j)^2}}{\sum_{i=1}^M \sum_{j=1}^N \sqrt{X_R(i,j)^2 + X_G(i,j)^2 + X_B(i,j)^2}} \right) \times 100 \quad (7)$$

where the subindexes R , G and B denote the corresponding image components in the RGB color space.

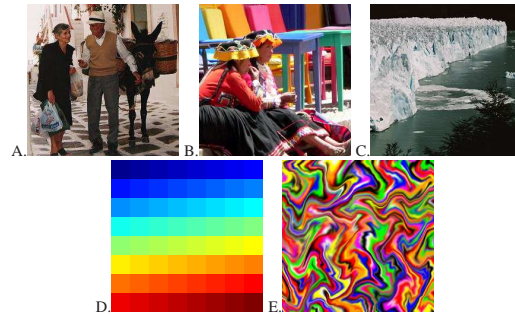


Figure 2: A. Im 1, B. Im 2, C. Im 3, D. Im 4 and E. Im 5.

4 ROBUSTNESS EVALUATION

Another important issue when evaluating image watermarking methods is the robustness, *i.e.*, the capacity of the watermark to survive standard image processing alterations, such as lossy compression, scaling, cropping, printing and scanning, etc..

In this paper, robustness of the watermark against JPEG compression is evaluated by computing a degradation coefficient, D , which quantifies the degradation in the watermark detectability caused by this image processing tasks. To perform the robustness test, the watermarked image is subjected to the above mentioned attack, and then the watermark is extracted following the procedure described in section 2. The normalized cross-correlation, $r_{w,w_e}(k)$, between the original, $w(\ell)$, and the extracted, $w_e(\ell)$, watermarks is then computed. The *detectability degradation coefficient* is then defined as,

$$\mathcal{D} \triangleq (1 - r_{w,w_e}(0)) \times 100. \quad (8)$$

5 RESULTS

In order to compare the performance of the proposed watermarking scheme IADWT_T and the IADWT in (Podilchuk and Zeng, 1998), a set of (256 × 256) natural color images was used (only five are shown due to space limitations, Im 1 to Im 5 in Figure 2). To make the results independent of the particular set of natural images considered, the same tests were also performed on synthetic images with large uniform areas (like Im 4 in Figure 2.D) and images with predominant high frequency regions (like Im 5 in Figure 2.E).

5.1 Fidelity Evaluation Results

In this section two separate tests to evaluate fidelity will be performed. The purpose of *Fidelity Test 1* is



Figure 3: Original (left) and Watermarked (right) images.

to illustrate the fact that the fidelity factor \mathcal{F} defined in (6) provides a much better assessment of image quality than the standard RMS_{FIT} . *Fidelity Test 2* is designed to compare the fidelity of the two DWT based insertion schemes described in Section 2.

Fidelity Test 1: In order to illustrate the fact that the RMS_{FIT} does not provide an objective assessment of image quality, a watermarked image with a strong watermark was generated with the IHW embedding technique proposed in (Cox et al., 1997). The original and the marked images are shown in the left and right sides of Figure 3, respectively. In this case the α parameter was chosen equal to 0.25, resulting in a fidelity factor $\mathcal{F} = 34.04\%$ and a $RMS_{FIT} = 91.26\%$. Based only on the RMS_{FIT} one would expect no noticeable distortions on the watermarked image which is not the case for this example (particularly in the sky portion at the top of the image). The fidelity factor \mathcal{F} in turn gives a better assessment of image quality.

Fidelity Test 2: The values of the watermark length L , the normalized watermark energy E computed as the normalized mean square error between the original and the watermarked images, the fidelity factor \mathcal{F} , and the RMS_{FIT} were computed for the five images in Figure 2, marked using the IADWT and IADWT_T insertion schemes described in Section 2. The results are shown in Table 1.

As can be observed from the fifth column in Table 1 there is no noticeable difference between the fidelity, as measured by the RMS_{FIT} , using both insertion schemes. The difference is more noticeable using the proposed fidelity factor, as can be observed from the values in the fourth column.

The values of the fidelity factor, \mathcal{F} , in Table 1 show that the IADWT_T method consistently outperforms the IADWT method regarding fidelity. Even for the case of images with large uniform color regions, like Im 4 in Figure 2.D, where the image adaptive methods are supposed to work poorly (Podilchuk and Zeng, 1998), the IADWT_T method produces non perceptible watermarks. On the other hand, the IADWT

Table 1: Results on Fidelity Evaluation for Im 1 to Im 5.

	L	E ($\times 10^{-3}$)	\mathcal{F} (%)	RMS_{FIT} (%)
Im 1				
IADWT	8347	1.40	92.27	97.45
IADWT _T	874	0.38	98.37	99.20
Im 2				
IADWT	9314	1.26	94.13	97.52
IADWT _T	1036	0.37	98.50	99.22
Im 3				
IADWT	8196	1.76	92.59	97.11
IADWT _T	1117	0.65	98.03	98.90
Im 4				
IADWT	3002	0.12	99.63	99.17
IADWT _T	1138	0.07	99.82	99.67
Im 5				
IADWT	11336	1.06	95.52	97.56
IADWT _T	1458	0.33	98.63	99.05

method introduces visible distortions, as can be observed from the first row in Figure 4 (see for instance the spots in the green regions in the left image).

The left columns in Figure 4 show the watermarked images corresponding to Im 1 and Im 4 using the above mentioned watermarking schemes. The right columns show the corresponding distortion maps obtained after applying the S-CIELAB ΔE_{94} metric. As expected, the distortion is larger in the regions with high frequency components, which results in a less perceptible watermark due to the masking phenomenon of the HVS.

5.2 Robustness Evaluation Results

In this subsection the robustness of the watermarked images against JPEG compression is evaluated, for both image adaptive DWT-based watermarking schemes. The detectability degradation coefficient \mathcal{D} , as defined in (8), is computed when JPEG-compression with quality factors in the range [95%-75%] is applied. The results for Im 1, Im 4 and Im 5 are shown in Figure 5 from top to bottom respectively. As can be observed the IADWT_T watermarking scheme consistently outperforms the IADWT one regarding robustness against JPEG compression.

6 CONCLUDING REMARKS

An image *fidelity factor* based on the S-CIELAB perceptual distortion metric has been introduced in this paper for the purposes of evaluating the distortion introduced by different IADWT watermark insertion

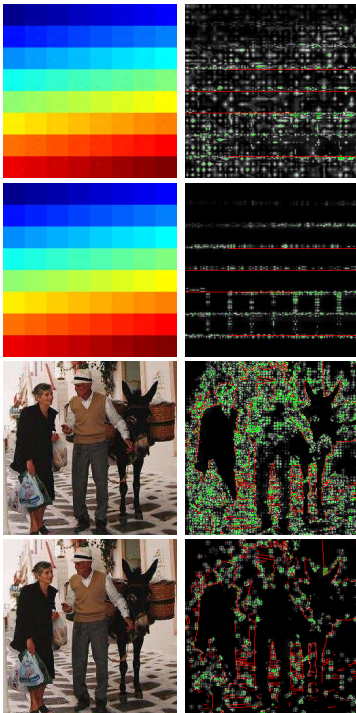


Figure 4: Left Column: IADWT marked Im 4 (first row), IADWT_T marked Im 4 (second row), IADWT marked Im 1 (third row), IADWT_T marked Im 1 (fourth row). Right Column: Corresponding distortion maps.

schemes. The use of this metric allows a perceptually aware objective quantification of image fidelity. Simulation results show the suitability of the proposed metric in the framework of still image digital watermarking. Further, a new IADWT watermarking scheme has been introduced, and its robustness against compression, and fidelity have been investigated. The results show that the proposed technique outperforms other methods available in the literature.

REFERENCES

Barni, M. and Bartolini, F. (2004). *Watermarking Systems Engineering - Enabling Digital Assets and Other Applications*. Marcel Dekker, Inc., New York.

Barni, M., Bartolini, F., and Piva, A. (2001). Improved wavelet-based watermarking through pixel-wise masking. *IEEE Trans. on Image Process.*, 10(5):783–791.

CIE: International Commission on Illumination (1995). Industrial colour difference evaluation. Tech. Rep. CIE 116-95, Austria.

Cox, I., Kilian, J., Leighton, F., and Shamoont, T. (1997). Secure spread spectrum watermarking for multimedia. *IEEE Trans. on Image Process.*, 6(12):1673–1687.

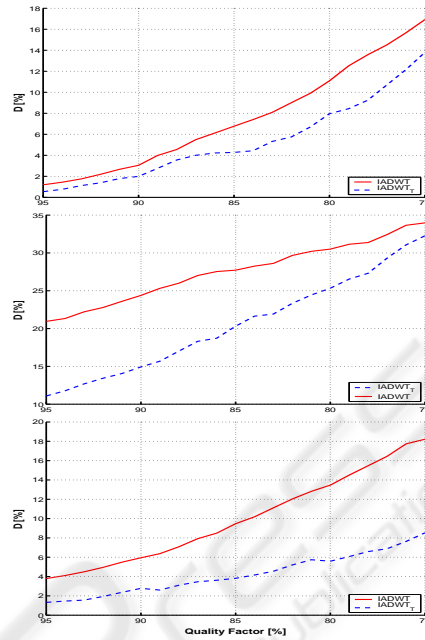


Figure 5: From Top to Bottom: Detectability degradation coefficient vs. JPEG Quality Factor for Im 1, Im 4 and Im 5.

Cox, I., Miller, M., and J.Bloom (2002). *Digital Watermarking*. Morgan Kaufmann, San Francisco.

Petitcolas, F. (2000). Watermarking schemes evaluation. *IEEE Signal Process. Magazine*, 17(5):58–64.

Podilchuk, C. and Delp, E. (2001). Digital watermarking: Algorithms and applications. *IEEE Signal Process. Magazine*, 18(4):33–46.

Podilchuk, C. and Zeng, W. (1998). Image-adaptive watermarking using visual models. *IEEE J. on Selected Areas in Communications*, 16(4):525–539.

Watson, A., Yang, G., Solomon, J., and Villasenor, J. (1997). Visibility of wavelet quantization noise. *IEEE Trans. on Image Process.*, 6:1164–1175.

Winkler, S. (2005). *Digital Video Quality Vision Models and Metrics*. John Wiley & Sons Ltd, Chichester, UK.

Zhang, X. and Wandell, B. (1998). Color image fidelity metrics evaluated using image distortion maps. *Signal Process.*, 70:201–214.

Zhang, Z. (1996). A spatial extension to CIELAB for digital color image reproduction. *Society for Information Display Symposium Technical Digest*, 27:731–734.

Zhang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Trans. on Image Process.*, 13(4):600–612.