Subjective Validation of Perceptual Fidelity Metrics for Image Watermarking

Franco Del Colle * Juan Carlos Gómez

Laboratory for System Dynamics and Signal Processing
FCEIA, Universidad Nacional de Rosario
CIFASIS, CONICET
Riobamba 245 bis, 2000 Rosario, Argentina
{delcolle, jcgomez}@fceia.unr.edu.ar

Abstract. The main contribution of this paper is the validation and comparison of several state-of-the-art watermark perceptual metrics used to quantify watermark transparency in still images. The validation of the metrics is carried out through subjective assessment by adapting a standardized technique for image quality assessment. Simulation results show that a metric based on S-CIELAB perceptual distortion maps exhibits the best correlation between subjective tests and the objective values. The Image Adaptive Watermarking methods in the Discrete Wavelet Transform Domain considered in this paper are compared using the proposed perceptual metric. Robustness against JPEG compression of the watermarking methods is also analyzed.

1 Introduction

The inexpensive and massive distribution of digital data has cause content owners to protect their information against piracy. Digital Watermarking has become the most efficient and widely used technique to copyright data in an imperceptible way, i.e. the embedded information (watermark) should be always present and detectable but users should not be supposed to perceive its existence. The main requirements that any watermarking technique should met are:

- *Perceptual transparency*: Property of the watermark of been imperceptible in the sense that humans can not distinguish the watermarked images from the original ones by simple inspection.
- *Robustness*: Capacity of the watermark to remain detectable after alterations due to processing techniques or intentional attacks.
- *Payload of the watermark*: Amount of information stored in the watermark, which in general depends on the application.

Good overviews on the state of the art of classical watermarking techniques can be found in the recent textbooks [1] and [2], and in [3], [4], [5] and the references therein.

* Author to whom all correspondence should be addressed.

XXXIV Conferencia Latinoamericana de Informática (CLEI 2008), pp. 1070-1079, 2008.

Among the different approaches that have been proposed in the literature for the watermarking of still images, the ones in the transform domain which are adapted to the particular image have proved to deliver better results regarding transparency and robustness. In these methods the length, location and amplitude of the watermark is adapted to the image characteristics, [6], [5], [7], [8]. This paper will focus on Image Adaptive Discrete Wavelet Transform (IADWT) domain watermarking techniques. In particular the methods in [7] and the modification in [9] are considered in this paper.

This paper focus on the validation of several perceptual and non-perceptual metrics for watermark image fidelity evaluation through subjective tests. In particular, the standard non perceptual Root Mean Square Error and the perceptual metrics introduced in [9], in [10], and in [11] are considered. Simulation results show that the perceptual metric in [9] outperforms the other metrics regarding its correlation to the subjective tests. In addition, the perceptual metric introduced by the authors in [9] is used to compare the fidelity of two different IADWT watermarking insertion schemes. The robustness of both watermark schemes against JPEG compression is also evaluated.

The rest of the paper is organized as follows. In section 2, both IADWT watermarking techniques are briefly described. In section 3, the perceptual metrics used for the evaluation of the fidelity performance are presented. For the sake of completeness the metric in the very recent reference [9] (only available online) is described in more detail. The general conditions for the subjective tests used to validate the different fidelity metrics are also described in this section. In section 4, the performance of the different metrics for the evaluation of fidelity of both IADWT schemes is illustrated. Finally, some concluding remarks are given in section 5.

2 Image Adaptive DWT 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 allow to determine the location of the transform coefficients and the amount that they can variate without being noticeable in the spatial domain.

In the watermark embedding scheme in [7], the watermark is modulated by the JND, and the coefficients are marked whenever they are greater than the JND threshold, *i.e.*

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

where $\widehat{X}(u,v)$ and $\widehat{X}^w(u,v)$ are the DWT coefficients of the original image and the watermarked image respectively, and $J(u,v)$ is the JND matrix at the u,v frequency in the DWT domain.

In this scheme, the watermark sequence $w(\ell)$ is generated from a zero mean, unit variance, normally distributed random sequence. In this way, the watermark sequence

weighted by the JND thresholds has lower power than the maximum power that can be inserted without causing noticeable distortions in the image. Figure 1 schematically depicts the image adaptive watermarking embedding scheme, where $X(i,j)$ denotes the original image and $X^w(i,j)$, the watermarked image.

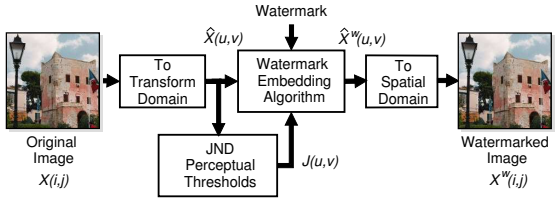

Fig. 1: Image Adaptive Watermarking Embedding Scheme.

The JND thresholds are computed based on a perceptual model of the Human Visual System (HVS). A widely used perceptual model is the one introduced by Watson in [12]. 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. This model has been used by the image compression standard JPEG2000, where the JND thresholds determine the optimal quantization step sizes or bit allocations for different parts of the image to be compressed.

The IADWT method in [7] has been compared with the Image Independent Watermarking (IIW) method in [13], showing some advantages regarding robustness and fidelity.

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

$$\widehat{X}^w(u,v) = \begin{cases} \widehat{X}(u,v) + J(u,v)w(\ell) & \widehat{X}(u,v) > J(u,v) > T \\ \widehat{X}(u,v) & \text{othewise} \end{cases} \tag{2}$$

This modified insertion scheme will be hereafter denoted as IADWT$_T$. The *rationale* for the constraint $J(u,v) > T$ is that when the JND thresholds are too small, the magnitude of the marking term in (2) becomes negligible. The introduction of the lower bound T has then the advantage of reducing the watermark length, improving in this way the fidelity. Through simulation trials a value of T equals 12 proved to be the most suitable for all tested images.

3 Watermarking Fidelity Assessment

In the evaluation of image watermarking methods it is of interest to judge the fidelity of the watermarked image. Basically, the fidelity is a measure of the similarity between the images before and after the watermark insertion. For some applications, fidelity is

the primary perceptual measure of concern, and in these cases the watermarked image must be indistinguishable from the original.

The natural way to assess the fidelity of watermarked images is to run a subjective test where observers are asked to rank the distortion of the images in a given scale. This type of evaluation involves large number of individuals in order for the results to be statistically significant and demands considerable time.

As an alternative to this, an objective assessment based on a metric that quantifies the watermarked image fidelity can be performed since it is less time consuming and does not require the involvement of human beings. However this objective assessment is usually validated with a subjective test.

Several metrics have been proposed in the literature to quantify image quality, see for instance [14] and the reference therein. The most successful ones are those that take into account the perceptual characteristics of the HVS. These techniques could eventually be used to quantify watermark fidelity. In section 3.2 the perceptual metrics used in this paper are described.

3.1 Subjective Assessment

As pointed out before the straightforward way to assess the fidelity of watermarked images is to run a subjective test. There are standardized techniques to perform subjective tests for general image quality assessment. For instance the Recommendation ITU-R BT.500-11 [15] issued by ITU [16] specifies a methodology for the subjective assessment of still images' quality. On the other hand no standards are available for subjective assessment of watermarked images' quality. Since watermarked images can be considered as the result of some processing operations (the watermark embedding algorithms) applied to the original image, these general subjective quality assessment techniques could in principle be applied to watermarked images. In this paper, the Double Stimulus Impairment Scale (DSIS) protocol, described in [15], is used. This protocol has also been used by Marini and coauthors in [17] in the same context.

The experiment was carried out in a room designed according to the recommendation ITU-R BT.500-11 [15]. Fourteen observers were enrolled to do the test and fifteen different natural images were watermarked using the two IADWT algorithms described in section 2. This resulted in 20 minute sessions where observers were asked to rate 30 images at an observation distance of six times the display size of the images. Both the original and the watermarked images were displayed side by side on the viewing monitor and the observers were asked to judge the quality of the marked image compared to the quality of the original on a scale of five categories, namely 5=Imperceptible, 4=Perceptible but not annoying, 3=Slightly annoying, 2=Annoying, and 1=Very annoying.

The results of these experiments are included in section 4.1.

3.2 Objective Assessment using Perceptual Metrics

As pointed out above, to avoid the dependence on human judgement, the objective assessment of watermarked image fidelity using a metric that takes into account the characteristics of the HVS is desirable. Several perceptual metrics have been proposed to quantify image quality. The metrics: Komparator introduced in [10], SSIM introduced

in [11], and the S-CIELAB based metric introduced in [9] will be briefly described in this section. They take into account the different sensitivities of the human eye for color discrimination, contrast masking and texture masking.

- **Komparator**: This metric consists of two main steps. The first one computes visual representations of the original and the watermarked images, using a model based on results from psychophysics experiments on color perception and masking effects. The second step performs a pooling of the errors in order to obtain a single value representing the distortion between the two images. This pooling is based on the density of errors and their structure.

- **Structural SIMilarity (SSIM)**: This metric works under the assumption that the human eye is highly adapted for extracting structural information from an image. Thus, quality evaluation is based on the degradation of this structural information assuming that error visibility should not be equated with loss of quality as some distortions may be clearly visible but not so annoying. The SSIM metric does not attempt to predict image quality by accumulating the errors associated with psychophysically understood simple patterns, but proposes to directly evaluate the structural changes between two complex-structured signals.

- **S-CIELAB based metric**: A widely used metric to measure fidelity is the CIELAB metric [18] that specifies how to transform physical image measurements into perceptual differences (ΔE). The metric was derived from perceptual measurements of color discrimination of large uniform targets. An extension of CIELAB, named S-CIELAB [19], includes the spatial-color sensitivity of the human eye. The S-CIELAB metric incorporates the different spatial sensitivities of the three opponent color channels by adding a spatial pre-processing step before the standard CIELAB ΔE calculation. As a result a S-CIELAB ΔE_{94} distortion map, indicating where the visible distortions are in the image and how large this distortions are, is obtained. Due to the spatial distribution of the S-CIELAB ΔE_{94} errors in the distortion maps it is difficult to make a comparison with other metrics. To provide a unique parameter quantifying the fidelity, a pooling of the S-CIELAB ΔE_{94} errors is proposed by defining the following *fidelity factor*:

$$\mathcal{F} \triangleq \left(1 - \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(S\Delta E_{94}(i,j)Mask(i,j))}{\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{X_L(i,j)^2+X_a(i,j)^2+X_b(i,j)^2}}\right)\times 100 \tag{3}$$

  where $S\Delta E_{94}$ 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, X_L, X_a and X_b are the image components in the *Lab* color space. Values of $\mathcal{F}$ close to 100 % indicates that non perceptible distortion is present in the watermarked image.

The performance of the above described perceptual metrics will be compared in section 4 with that of the standard (non perceptual) Root Mean Square (RMS) error. The non perceptual metric RMS Fit (RMS_{FIT}) is obtained by making a pooling of the

RMS errors, resulting in:

$$RMS_{FIT} \triangleq \left(1 - \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{\Delta X_R(i,j)^2+\Delta X_G(i,j)^2+\Delta X_B(i,j)^2}}{\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{X_R(i,j)^2+X_G(i,j)^2+X_B(i,j)^2}}\right)\times 100 \tag{4}$$

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

4 Simulation Results

The metrics described in subsection 3.2 are used in this section to evaluate the fidelity of the IADWT watermarking described in section 2. A set of fifteen (256×256) natural color images was used. Three of these images, called Image 1 to Image 3, are shown in Figure 2. The complete image dataset have not been included here due to space limitations but it can be downloaded from the website of the group (not included here to preserve anonymousness for the double blind review system).

Fig. 2: From left to right: Image 1, Image 2 and Image 3.

Results from two separate tests are presented in this section. The purpose of Test 1 in subsection 4.1 is to compare the four fidelity metrics, namely, the standard RMS_{FIT}, and the perceptual metrics, SSIM, Komparator and the one defined in eq. (3). On the other hand, Test 2 in subsection 4.2 is designed to compare the fidelity of the two IADWT insertion schemes described in section 2 using the S-CIELAB based metric, which is the one that best matches the subjective tests.

4.1 Test 1: Fidelity Metrics Comparison

In order to illustrate which metric provides the best objective assessment of image quality for both watermarking methods, the four metrics are computed and compared to the mean opinion score [1] (MOS) for the fifteen images. The corresponding 97.5 % confidence intervals (CI) were also calculated to specify intervals of values with the highest likelihood of containing the true value of the general MOS. These intervals, centered in the MOS, are shown in blue solid line in Figure 3; the non perceptual RMS_{FIT} is denoted with green triangles, the SSIM values with orange squares, the Komparator

[1] The Mean Opinion Score for each image is the average of the scores assigned by the observers.

values with brown circles, while the Fidelity Factor $\mathcal{F}$ with red crosses. The values in Figure 3 are normalized in the range $[1,5]$.

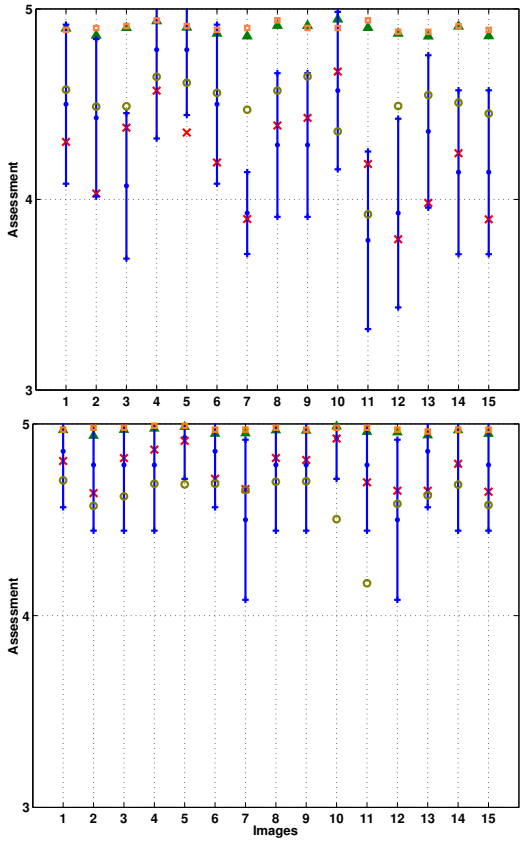

Fig. 3: Comparison of Objective and Subjective Assessment for Watermarking Method IADWT (top) and for Watermarking Method IADWT$_T$ (bottom). *CI*: Blue solid line, RMS_{FIT}: green triangles, SSIM: orange squares, Komparator: brown circles, $\mathcal{F}$: red crosses.

From Figure 3 it is clear that the RMS_{FIT} does not give a correct assessment of fidelity as the values fail to fall in the confidence intervals for twelve out of thirty watermarked images. The number of points that fall outside the confidence intervals and the average distance (d) of each metric to the MOS were calculated for both Watermarking algorithm and the corresponding values are shown in Table 1.

From Figure 3 and Table 1, it is clear that the fidelity factor $\mathcal{F}$ is the metric that best fits the subjective results, although the Komparator metric gives also acceptable results.

Table 1: Performance of the metrics

	IADWT		IADWT$_T$	
	Points outside *CI*	*d*	Points outside *CI*	*d*
RMS_{FIT}	10	0.59	2	0.18
SSIM	9	0.29	2	0.12
Komparator	3	0.26	3	0.20
$\mathcal{F}$	1	0.23	0	0.08

4.2 Test 2: Watermarking Schemes Comparison

In this section, the fidelity factor, $\mathcal{F}$, is used to compare the performance of the IADWT and IADWT$_T$ insertion schemes. In Figure 4, the values of $\mathcal{F}$ for the IADWT and IADWT$_T$ insertion schemes are represented by red circles and blue crosses, respectively.

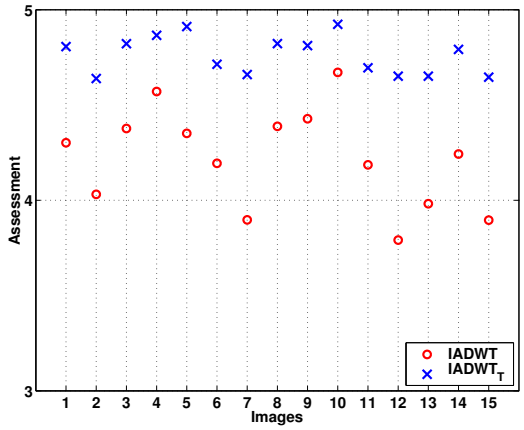

Fig. 4: Objective Assessment based on $\mathcal{F}$ for Watermarking Method IADWT (red circles) and for Watermarking Method IADWT$_T$ (blue crosses).

As it can be observed, the IADWT$_T$ method consistently outperforms the IADWT one regarding fidelity. Even for the case of images with large uniform color regions, as Image 3 in Figure 2, where the image adaptive methods are supposed to work poorly [7], the IADWT$_T$ method produces non perceptible watermarks ($\mathcal{F} = 4.822$).

In order to examine if the better fidelity performance of the IADWT$_T$ method results in a loss of robustness, as the watermark length is reduced, the robustness against JPEG compression at different rates is evaluated. Both IADWT methods were compared by computing a degradation coefficient which quantifies the degradation in the watermark

detectability caused by this image processing task. Therefore, to perform the robustness test, the watermarked image is compressed at a given rate, and then the watermark is extracted. 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*, D, is then defined as

$$\mathcal{D} \triangleq (1 - r_{w,w_e}(0)) \times 100. \tag{5}$$

As it can be observed in Figure 5 the IADWT$_T$ watermarking scheme has a lower value of D, for all the images and the three compression rates considered, than the IADWT, proving to be more robust against JPEG compression.

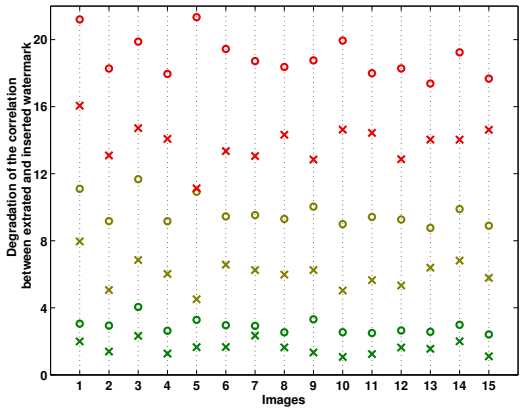

Fig. 5: Robustness against JPEG Compression at compression rates 90% (green), 80% (brown) and 70% (red) based on the degradation of the correlation between the extracted and the inserted watermark. Watermarking Method IADWT (circle) and for Watermarking Method IADWT$_T$ (cross).

5 Concluding Remarks

Several image quality perceptual metrics have been tested in this paper for the purpose of evaluating the transparency of image watermarking insertion schemes. In particular IADWT watermark insertion algorithms were tested. The evaluation has been carried out by performing subjective tests using the protocol described in the Recommendation ITU-R BT.500-11 [15] and comparing the MOS to the result of each metric. Simulation results show that the image *fidelity factor* based on the S-CIELAB ΔE_{94} perceptual distortion maps has a better correlation with the subjective tests for the purposes of quantifying still image watermarking fidelity. In addition, a comparison of the fidelity and the robustness of the two IADWT watermarking schemes has been done showing that the IADWT$_T$ outperforms the method in [7] regarding image fidelity but not at a cost of a decrease in the robustness.

References

1. Barni, M., Bartolini, F.: Watermarking Systems Engineering - Enabling Digital Assets and Other Applications. Marcel Dekker, Inc., New York (2004)
2. Cox, I., Miller, M., J.Bloom: Digital Watermarking. Morgan Kaufmann, San Francisco (2002)
3. Langelaar, G., Setyawan, I., R.Lagendijk: Watermarking digital image and video data. IEEE Signal Processing Magazine **17**(5) (2000) 20–46
4. Petitcolas, F.: Watermarking schemes evaluation. IEEE Signal Processing Magazine **17**(5) (2000) 58–64
5. Podilchuk, C., Delp, E.: Digital watermarking: Algorithms and applications. IEEE Signal Processing Magazine **18**(4) (2001) 33–46
6. Barni, M., Bartolini, F., Piva, A.: Improved wavelet-based watermarking through pixel-wise masking. IEEE Transactions on Image Processing **10**(5) (2001) 783–791
7. Podilchuk, C., Zeng, W.: Image-adaptive watermarking using visual models. IEEE Journal on Selected Areas in Communications **16**(4) (1998) 525–539
8. Swanson, M., Zhu, B., Tewfik, A.: Transparent robust image watermarking. In: Proceedings International Conference on Image Processing. Volume 3. (1996) 211–214
9. Del Colle, F., Gómez, J.C.: DWT based digital watermarking fidelity and robustness evaluation. Journal of Computer Science & Technology **8**(1) (2008) 15–20
10. Le Callet, P., Barba, D.: A robust quality metric for color image quality assessment. In: Proceedings of the IEEE International Conference on Image Processing. Volume 1. (2003) 437–440
11. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4) (2004) 600–612
12. Watson, A., Yang, G., Solomon, J., Villasenor, J.: Visibility of wavelet quantization noise. **6**(8) (1997) 1164–1175
13. Cox, I., Kilian, J., Leighton, F., Shamoon, T.: Secure spread spectrum watermarking for multimedia. IEEE Transactions on Image Processing **6**(12) (1997) 1673–1687
14. Winkler, S.: Digital Video Quality Vision Models and Metrics. John Wiley & Sons Ltd, Chichester, UK (2005)
15. ITU: Recommendation ITU-R BT.500-11: Methodology for the subjective assessment of the quality of television pictures. Technical report, International Telecommunication Union (2002)
16. International Telecommunication Union: (http://www.itu.int)
17. Marini, E., Autrusseau, F., Le Callet, P., Campisi, P.: Evaluation of standard watermarking techniques. In: Security, Steganography, and Watermarking of Multimedia Contents IX, Proc. of SPIE-IS& Electronic Imaging. Volume 6505., San Jose, CA, USA (2007) 1–10
18. CIE: International Commission on Illumination: Recommendations on uniform color spaces, color difference equations, psychometrics color terms. Technical Report CIE 15 (E.-1.3.1), Supplement No.2, Vienna, Austria (1971)
19. Zhang, Z.: A spatial extension to CIELAB for digital color image reproduction. Society for Information Display Symposium Technical Digest **27** (1996) 731–734