SUBJECTIVE VERIFICATION OF PERCEPTUAL METRICS FOR IMAGE WATERMARKING FIDELITY

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

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

Abstract: In this paper, the performance of several state-of-the-art watermark perceptual transparency metrics is evaluated through subjective assessment. Simulation results show that a metric based on S-CIELAB distortion maps proved to be better correlated to the subjective tests than other objective metrics available in the literature. The paper focus on Image Adaptive Watermarking methods in the Discrete Wavelet Transform Domain since they yield better results regarding robustness and transparency than other watermarking schemes.

1 INTRODUCTION

Digital Watermarking has become the most efficient and widely used technique addressing the issue of digital data protection. The idea is to imperceptibly embed information (the watermark) into the original data in such a way that always remains present and detectable. A set of requirements should be met by any watermarking technique (Barni and Bartolini, 2004). The main requirements are *perceptual transparency, payload of the watermark and robustness*. Perceptual transparency refers to the 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. Payload of the watermark refers to the amount of information stored in the watermark, which in general depends on the application. 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 (Langelaar et al., 2000), (Petitcolas, 2000) and the references therein.

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 (Barni et al., 2001) and (Podilchuk and Zeng, 1998). This paper will focus on Image Adaptive Discrete Wavelet Transform (IADWT) domain watermarking techniques.

In this paper, several perceptual metrics for watermark image fidelity evaluation are validated through subjective tests. In particular, the perceptual metrics introduced in (Le Callet and Barba, 2003), in (Del Colle and Gómez, 2008) and in (Wang et al., 2004) are considered. Simulation results show that the perceptual metric in (Del Colle and Gómez, 2008) outperform the other metrics regarding correlation to the subjective tests.

2 IADWT WATERMARKING

In this paper, the watermark embedding scheme in the DWT domain in (Podilchuk and Zeng, 1998) is considered. Here, the watermark is modulated by the Just Noticeable Differences (JND) thresholds, and the coefficients are marked whenever they are greater than the JND threshold, *i.e.*

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

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

Del Colle F. and Carlos Gómez J. (2008).
SUBJECTIVE VERIFICATION OF PERCEPTUAL METRICS FOR IMAGE WATERMARKING FIDELITY.
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pages 397-400
DOI: 10.5220/0001940903970400
Copyright © SciTePress

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.

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 in (Watson et al., 1997), which 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.

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

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

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 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. The natural way to assess fidelity 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 (Winkler, 2005) 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.

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 (ITU, 2002) specifies a methodology for the subjective assessment of still image quality. On the other hand no standards are available for subjective assessment of watermarked image 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 be applied to watermarked images. In this paper, the Double Stimulus Impairment Scale (DSIS) protocol, described in (ITU, 2002), is used. This protocol has also been used by Marini and coauthors in (Marini et al., 2007) in the same context.

The experiment was carried out in a room designed according to the recommendation ITU-R BT.500-11 (ITU, 2002). 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 min sessions where observers were asked to rate 30 images at an observation distance of six times the display size of the images. The original and the watermarked images were displayed side by side on the monitor and the observers were asked to rate the quality of the marked image compared to that 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.

3.2 Objective Assessment

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 S-CIELAB based metric introduced by the present authors in (Del Colle and Gómez, 2008) will be briefly described and compared to the Komparator metric introduced in (Le Callet and Barba, 2003) and the SSIM metric introduced in (Wang et al., 2004). All of them take into account the different sensitivities of the human eye for color discrimination, contrast masking and texture masking.

The S-CIELAB (Zhang, 1996) metric is an extension of the CIELAB metric (CIE, 1971) which 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 a pooling of the standard Root Mean Square (RMS) error, namely, the RMS Fit (RMS_{FIT}) defined as:

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

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

4 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. The complete image dataset can be downloaded from: http://www.fceia.unr.edu.ar/lsd/mrg/watermark/.

Results from two separate tests are presented in this section. The purpose of Test 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 is designed to compare the fidelity of the two IADWT insertion schemes described in section 2 using the

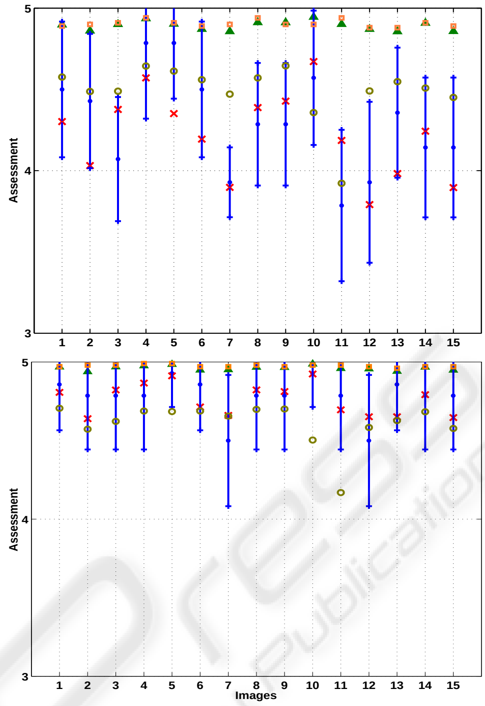

Figure 1: Comparison of Objective and Subjective Assessment for methods IADWT (top) and IADWT$_T$ (bottom). *CI*: Blue solid line, RMS_{FIT}: green triangles, SSIM: orange squares, Komparator: brown circles, $\mathcal{F}$: red crosses.

S-CIELAB based metric, which is the one that best matches the subjective tests.

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 Fig. 1; the non perceptual RMS_{FIT} is denoted with green triangles, the SSIM values with orange squares, the Komparator values with brown circles, while the Fidelity Factor $\mathcal{F}$ with red crosses. The values in Fig. 1 are normalized in the range $[1,5]$.

From Fig. 1 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 calcu-

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

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

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

lated for both Watermarking algorithms and the corresponding values are shown in Table 1. From Fig. 1 and Table 1, it is clear that the metric $\mathcal{F}$ is the one that best fits the subjective results, although the Komparator metric gives also acceptable results.

Test 2 - Watermarking Schemes Comparison. The fidelity factor, $\mathcal{F}$, is used to compare the performance of the IADWT and $IADWT_T$ insertion schemes. In Fig. 2, the values of $\mathcal{F}$ for the IADWT and $IADWT_T$ insertion schemes are represented by red circles and blue crosses, respectively.

As it can be observed, the $IADWT_T$ method outperforms the IADWT one regarding fidelity. This holds even for images with large uniform color regions, where the image adaptive methods are supposed to work poorly (Podilchuk and Zeng, 1998) (results are not shown here due to space limitation).

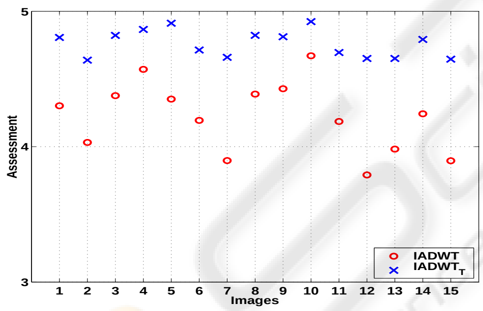

Figure 2: Objective Assessment based on $\mathcal{F}$ for methods IADWT (red circles) and $IADWT_T$ (blue crosses).

5 CONCLUDING REMARKS

Several image 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 (ITU, 2002) 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 of the two IADWT watermarking schemes has been done showing that the $IADWT_T$ outperforms the method in (Podilchuk and Zeng, 1998) regarding image fidelity.

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 pixelwise masking. *IEEE Transactions on Image Processing*, 10(5):783–791.

CIE (1971). Recommendations on uniform color spaces, color difference equations, psychometrics color terms. Technical Report CIE 15 (E.-1.3.1), Vienna, Austria.

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

Del Colle, F. and Gómez, J. C. (2008). DWT-based digital watermarking fidelity & robustness evaluation. *Journal of Computer Science & Technology*, 8(1):15–20.

ITU (2002). Recommendation ITU-R BT.500-11: Methodology for the subjective assessment of the quality of television pictures. Technical report, International Telecommunication Union.

Langelaar, G., Setyawan, I., and R.Lagendijk (2000). Watermarking digital image and video data. *IEEE Signal Processing Magazine*, 17(5):20–46.

Le Callet, P. and Barba, D. (2003). A robust quality metric for color image quality assessment. In *Proceedings of the IEEE International Conference on Image Processing*, volume 1, pages 437–440.

Marini, E., Autrusseau, F., Le Callet, P., and Campisi, P. (2007). Evaluation of standard watermarking techniques. In *Proc. of SPIE-IS& Electronic Imaging*, volume 6505, pages 1–10, San Jose, CA, USA.

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

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

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

Watson, A., Yang, G., Solomon, J., and Villasenor, J. (1997). Visibility of wavelet quantization noise. *IEEE Transactions on Image Processing,*, 6(8):1164–1175.

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

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