HAL
open science

A Multi-purpose Objective Quality Metric for Image Watermarking

Vinod Pankajakshan, Florent Autrusseau

To cite this version:

Vinod Pankajakshan, Florent Autrusseau. A Multi-purpose Objective Quality Metric for Image Watermarking. International Conference on Image Processing, Sep 2010, Hong Kong SAR China. pp.2589-2592. hal-00495957

HAL Id: hal-00495957

https://hal.science/hal-00495957v1

Submitted on 12 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cover Page

1) Title of the paper:

A MULTI-PURPOSE OBJECTIVE QUALITY METRIC FOR IMAGE WATERMARKING

2) authors' affiliation and address:

IRCCyN-IVC, (UMR CNRS 6597), Polytech' Nantes
Rue Christian Pauc, La Chantrerie, 44306 NANTES, France.
Tel : 02.40.68.30.52
Fax : 02.40.68.32.32

3) e_mail address:

Florent.Autrusseau@univ-nantes.fr

4) Conference & Publisher information:

IEEE ICIP 2010
http://www.icip2010.org/
http://www.ieee.org/

5) bibtex entry:

```
@conference{ICIP2010_CPA2,
  author = {V. Pankajakshan and F. Autrusseau},
  title = {A Multi-purpose Objective Quality Metric for Image
Watermarking},
  booktitle = {IEEE International Conference on Image
Processing, ICIP'2010},
  year = {2010}
}
```

A MULTI-PURPOSE OBJECTIVE QUALITY METRIC FOR IMAGE WATERMARKING

Vinod Pankajakshan, Florent Autrusseau

Laboratoire IRCCyN, Université de Nantes
Rue Christian Pauc, 44306Nantes, France

ABSTRACT

Knowing that the watermarking community use simple statistical quality metrics in order to evaluate the watermarked image quality, the authors have recently proposed a simplified objective quality metric (OQM), called "CPA", for watermarking applications. The metric used the contrast sensitivity function, along with an adapted error pooling, and proved to perform better than state-of-the-art OQMs. In this work, we intend to improve the performance of the CPA metric. The new metric includes the most important steps of Human Visual System (HVS) based quality metric, namely spatial frequency consideration and masking effects. Besides, this work goes further than classical image quality assessment, and several objective quality metrics will be tested in a watermarking algorithm comparison scenario. We will show that the proposed metric is both able to accurately predict the observers score in a quality assessment task, and is also able to compare watermarking algorithms altogether on a perceptual quality viewpoint.

1. INTRODUCTION

Digital watermarking techniques operates in various transformed domains, in various frequency ranges and includes different considerations for ensuring invisibility. Although it is quite easy to compare the robustness of several embedding techniques by simply using a robustness benchmark and counting the percentage of correct watermark detection, it is much more difficult to compare the perceptual quality of images watermarked with different embedding techniques [1]. Despite the importance of the watermark invisibility, it was shown that several data hiding techniques induce strong perceptual distortions. Objective Quality Metrics (OQM) have recently been very extensively used for various image processing applications. Such metrics were mostly designed for coding artefacts annoyance assessment. The watermarking community recently started showing some interest in using OQM, either for the watermarked media quality assessment [2, 3] , or for assessing the quality of attacked images [4]. However, for watermarking applications, the Objective Quality Metrics have mostly been used for determining the perceived quality of watermarked images (using a particular embedding technique). We have shown that in such context, there is a large discrepancy of metrics performances [5], and most importantly, that OQM must be chosen very carefully when comparing several images marked with distinct embedding algorithms (supposedly modifying distinct parts of the frequency spectrum).

A different framework was proposed in [1], where several quality metrics were tested in order to rank the visibility of watermarking algorithms. The "Komparator" metric [6] proved to provide the best ranking according to subjective experiments. It actually appeared that although the watermark invisibility is an important requirement, many embedding techniques distort the watermarked image so severely that the Mean Opinion Score (MOS) provided by human observers during subjective experiments could be as low as 2 on an 1 to 5 annoyance scale (a score of 2 corresponding to "Annoying" distortions) with default embedding parameter (embedding strength).

In brief, researches on quality assessment for digital watermarking were able to either determine the best OQM for assessing the perceptual quality for a specific embedding algorithm, or to specify the best metric when comparing quality performances of several watermarking algorithms. However, the optimal metric differs in these two scenarios.

In this work, we intend to propose a multi-purpose OQM providing on one hand an accurate estimation of the observers' quality score and on the other hand, being able to rank watermarking techniques based on a perceptual quality criterion. This work is based on a preliminary study [5] proposing a simplified quality metric, CPA, for watermarking application. Although the CPA metric was exploiting very basic HVS features, it appeared to perform significantly better than many state-of-the-art OQMs. The proposed work is aimed at improving the performance of the CPA metric by incorporating a simplified contrast masking model.

This paper is structured as follows: In Section 2 we present the improvements brought to the CPA metric. In Section 3 the experimental setup is briefly reminded (the setup was already presented in previous works), and the added databases are described. Section 4 shows the improvements of the new metric regarding two distinct quality assessment tasks (regular image quality assessment and watermarking algorithm comparison regarding perceptual quality). Finally, the main contributions of the work are highlighted in Section 5.

2. ENHANCED QUALITY METRIC

The previously proposed CPA metric omitted the contrast masking property of the HVS for simplicity. The contrast masking is an important feature when modeling HVS behavior, it basically discriminates the frequency discrepancies. In brief, a high frequency distortion would inevitably have a much lower impact on the visibility when added up on a high frequency image area than in the lower frequencies. The contrast masking is thus particularly important in watermarking framework, where noise-like watermark could be embedded uniformly on the image without regard to the image masking capabilities. In HVS-based OQMs, the contrast masking is exploited by a threshold elevation step after a perceptual sub-band decomposition [7], which is computationally expensive. So, in this work the perceptual sub-band decomposition is replaced with a simple block-based frequency decomposition.

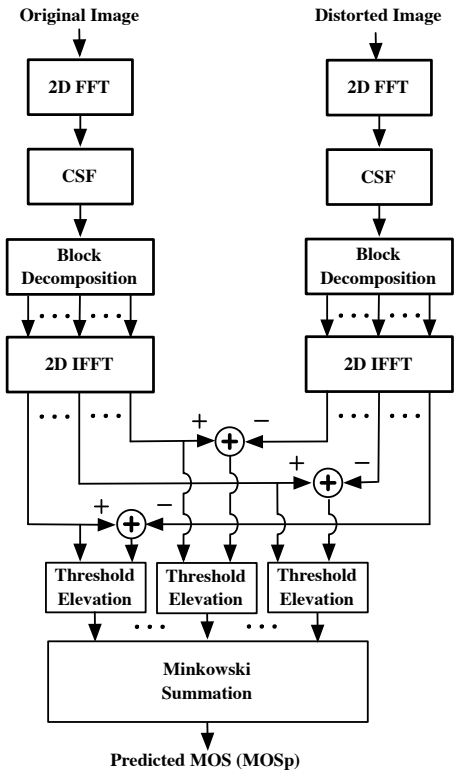

Fig. 1. Block diagram of the proposed OQM.

Figure 1 shows different steps composing the proposed metric, referred as CPA2 hereafter. Both the original and the distorted images are first subjected to 2D-FFT and each of the resulting spectrum is weighted by 2D-CSF. The CSF filtered spectra are then divided into non-overlapping blocks of size $N \times N$, with the base-band centered at the DC coefficient. Inverse FFT is applied to each $N \times N$ block, except the base-band, to get sub-band images. Let $X_k(i,j)$ and $Y_k(i,j)$be respectively the kth sub-band image at location (i,j) corresponding to the original and the distorted image. The threshold elevation for the kth sub-band image is computed as:

$$T_k(i,j) = \left(1 + \left(a_1 \left(a_2 \mid X_k(i,j) \mid\right)^s\right)^b\right)^{\frac{1}{b}} \tag{1}$$

where $a_1 = 0.0153, a_2 = 392.5$ and s, b are frequency dependant parameters [7]. Finally, the metric value M is obtained by generalized Minkowski summation given as:

$$M = \left(\sum_{i,j}\left(\left(\sum_k \left(\frac{X_k(i,j) - Y_k(i,j)}{T_k(i,j)}\right)^{\alpha_f}\right)^{\frac{1}{\beta_f}}\right)^{\alpha_s}\right)^{\frac{1}{\beta_s}} \tag{2}$$

where α_f, β_f are Minkowski parameters for frequency summation and α_s, β_s are the parameters for spatial summation.

3. EXPERIMENTAL SETUP

The experimental dataset is composed of 7 subjective databases. Among this dataset, four databases were tested in [5]. In this section, we will briefly summarize all tested databases, as well as the performance tools used for the metric evaluation. Four databases are considering watermarking applications, and the remaining three only includes coding distortions for comparison purpose with state of the art metrics, which have mostly been assessed on such kind of artefacts. Interested readers might refer to [5] for further details on the different setups. The first database is composed of 210 images watermarked in three distinct frequency ranges. The watermarking technique basically modulates a noise-like watermark onto a frequency carrier, and additively embeds the watermark in different regions of the Fourier spectrum (Database 1). The second database is composed of 120 distorted images watermarked using the "Broken Arrows" watermarking technique [8], operating in the wavelet domain (Database 2). Similarly, the third database exploits both the Wavelet transform, and the Dual-Tree Complex wavelet transform for multiplicative watermark embedding and generates 120 distorted images (DB3). In the fourth database, three image coding techniques were used (JPEG, JPEG200, and LAR coding) and generated 120 distorted images (DB4). Two more coding databases were used, in order to ensure a proper quality assessment for a large amount of distinct databases. The fifth database is composed of 168 distorted images (84 JPEG coded, and 84 JPEG2000 coded). The Sixth database is composed of 190 distorted images coded with JPEG, JPEG2000 and LAR coding (DB6). Note that although the last 3 databases consist of images with similar kind of distortions, we consider them separately because they have different image contents and different subjective experimental setup.

Finally, as previously explained, in this work, we intend to find the optimal OQM in a perceptual ranking task. A seventh

database was thus added [1]. This database includes 100 distorted images watermarked with 10 watermarking techniques on 5 input images, each image being watermarked with 2 embedding strengths, the default strength and 1.5 times the default strength of the embedding algorithm. The watermarking algorithms and the experimental setup for the subjective experiment are detailed in [1]. An in-depth objective quality assessment test was performed (see Section 4.1) and it appeared that due to the very wide range of distortions in this database (10 watermarking techniques operating in various frequency domains, various frequency ranges, and with different embedding formulas) no metric could perform well on such database. The correlation between the MOS and the predicted MOS (MOSp) for this database was varying between 0.3 with the PSNR and 0.7 for the CPA2 metric. We thus hereby use this database only for algorithm comparison.

4. RESULTS

In this section, we provide an evaluation of the proposed CPA2 metric with regard to two distinct assessment tasks. The first task is common quality assessment, where the OQM is basically used to predict the observers subjective scores (MOS). In the second task, we intend to find the OQM providing the more accurate watermarking algorithm ranking (on a perceptual quality criterion). The goal of this task being to compare different watermarking techniques altogether based on a perceptual quality criterion. These two distinct tasks are described in the following. We have evaluated the performance of the CPA2 metric for different sets of Minkowski parameters and the results presented here correspond to the values $\alpha_s = 5, \beta_s = 10, \alpha_f = 5,$ and $\beta_f = 10$ [5]. The block-size for the frequency decomposition in CPA2 was 15×15 .

4.1. Image Quality assessment task

Sixteen publicly available quality metrics (including 12 from the metrix_mux[1] package, as well as WPSNR, Komparator, C4 and CPA) were compared with the proposed CPA2 metric. Based on previous works, only the three metrics presenting overall best performances were selected here for OQM comparison along with the PSNR, in order to clearly outline the limitations of this statistical metric. The VIF [9] and C4 [7] metrics have shown interesting predictions [5] of the MOS values and will be used here. The CPA metric although very simple in its design also presented very interesting quality assessment performances and will be compared here with its improved version "CPA2". Figure 2 shows respectively the Pearson and Spearman correlation, the RMSE and Outlier Ratio between the MOSp and the MOS for six selected databases, and the six tested OQMs. Several observations can be raised on this Figure. The modifications brought on the

[1] http://foulard.ece.cornell.edu/gaubatz/metrix_mux/

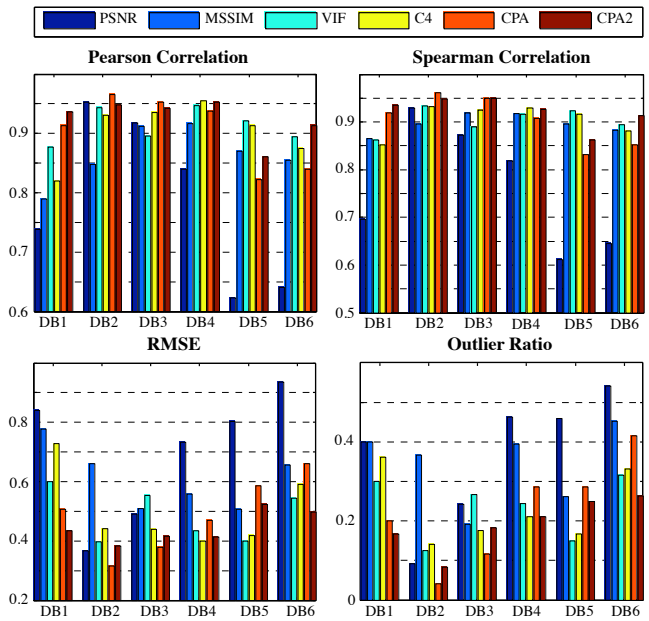

Fig. 2. Performance evaluation of the 6 tested metrics.

CPA2 metric allows a clear enhancement of the metric capabilities on four databases (out of the six tested). No improvements were observed on DB 2 & 3 for the CPA2 metric, but it is important to notice that both metrics present very good performances on these databases, and any further improvement could hardly be achieved. Both the C4 and VIF metrics, which have mostly been tested and designed for coding applications shows overall good performances on the three coding databases (DB 4, 5 and 6). Only CPA and CPA2 presents acceptable performances on the multi-frequency watermarks database (DB1). The PSNR is among the worst metrics for most cases, and especially has larger RMSE and outlier ratio (inconsistency of the PSNR values). However, the PSNR presented acceptable results on the Databases 2 &3, as these embedding techniques fixes a quality condition on the PSNR value, thus regularly spacing the distribution of plots in the MOS versus MOSp (see detailed explanations in [5]).

4.2. Watermarking algorithms invisibility ranking task

Although the watermark invisibility is of great importance, some watermarking techniques may emphasis the robustness features, at the expense of the watermark imperceptibility. It is thus important to have a tool being able to rank several watermarking techniques altogether regarding the perceptual aspect of the watermarked images. Previous works were conducted on this matter [1], and showed that due to the complexity and wide variety of the distortions induced by the watermarking techniques, ranking algorithms altogether is not an easy task. The Komparator metric [6] proved to be the more suitable in this "perceptual ranking" task, whereas the C4

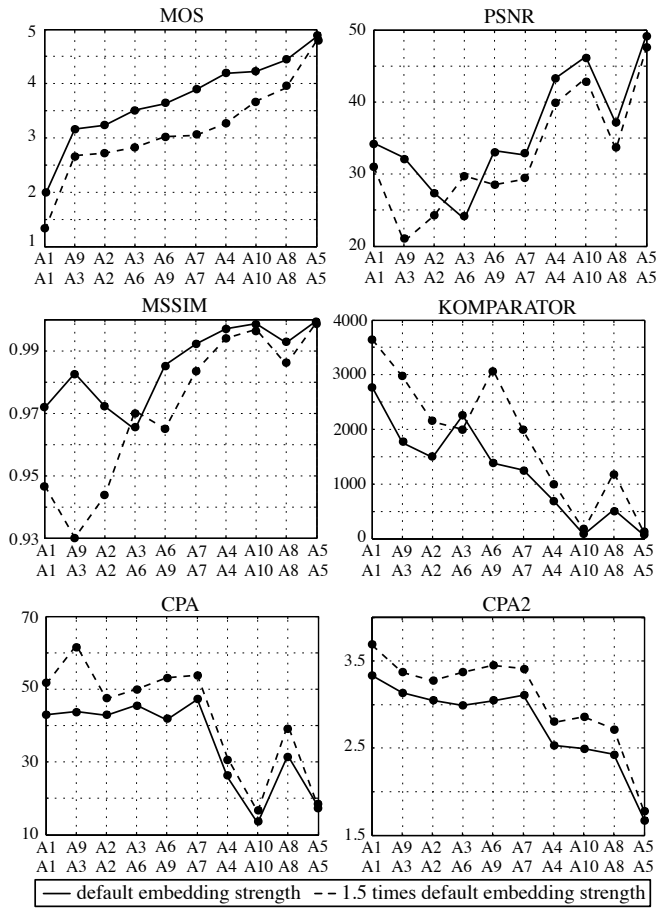

Fig. 3. Watermarking algorithms ranking according to perceptual quality.

metric proved to have better performances for typical quality assessment task. With regard to the recently proposed OQMs we hereby want to reevaluate the metrics performances for this specific task.

The MOS and the OQM values for the 5 watermarked images corresponding to the default embedding strength are averaged and plotted in solid lines in the figure as a function of the algorithm numbers, in the ascending order of the MOS values. Similarly, the values corresponding to 1.5 times default embedding strength are plotted in dashed lines. The first row in the X-axis label represent the Algorithm numbers for the default embedding strength and the second row represent those for 1.5 times the default embedding strength. Note that the PSNR and the MSSIM are similarity measures which gives higher metric values for better quality images whereas the Komparator, CPA and CPA2 are distortion measures which give lower metric values for better quality images.

As evident from the non-monotonic plots, none of the OQMs perfectly rank the watermarking algorithms according to the perceptual quality point of view. For instance, consider the MOS and the MSSIM values for algorithms 3 and 1 (default embedding strength). The MOS values for A1 and A3 are respectively 2 and 3.5 whereas the MSSIM and PSNR values for A3 are lower than that of A1. It can be noticed that among the tested OQMs, the CPA2 provides the best ranking of the algorithms in accordance with the MOS values. Furthermore, for the CPA2 metric, the plots corresponding to the two embedding strengths are not crossing each other, thus following the same trend in the MOS values.

5. CONCLUSIONS

We have presented in this work a new objective quality metric, and tested its performances regarding two distinct quality assessment tasks. The proposed OQM is an improved version of a recently proposed simplified quality metric named CPA. The improvements appeared to significantly enhance the metric's performance in most scenarios, and more specifically for the perceptual ranking task. To the best of our knowledge, the CPA2 metric is the only objective quality metric having good performances for both regular quality assessment and watermarking techniques perceptual comparison.

6. REFERENCES

[1] P. Le Callet, F. Autrusseau, and P. Campisi, *Multimedia Forensics and Security*, chapter IX, pp. 164–193, Idea Group Publishing, 2008.

[2] E. Drelie Gelasca, T. Ebrahimi, M. Corsini, and M. Barni, "Objective Evaluation of the Perceptual Quality of 3D Watermarking," in *IEEE International Conference on Image Processing (ICIP)*, 2005.

[3] S. Winkler, E. Drelie Gelasca, and T. Ebrahimi, "Toward perceptual metrics for video watermark evaluation.," in *Applications of Digital Image Processing*. 2003, vol. 5203 of *Proc. of SPIE*, pp. 371–378, SPIE.

[4] A. D'Angelo, L. Zhaoping, and M. Barni, "A full-reference quality metric for geometrically distorted images," *IEEE Trans. on Image Processing*, 2010 (to appear).

[5] M. Carosi, V. Pankajakshan, and F. Autrusseau, "Toward a simplified perceptual quality metric for watermarking applications," in *Proc. of the SPIE Electronic Imaging*, 2010, vol. 7542.

[6] D. Barba and P. Le Callet, "A robust quality metric for color image quality assessment," in *IEEE International Conference on Image Processing*, 2003, pp. 437–440.

[7] M. Carnec, P. Le Callet, and D. Barba, "Objective quality assessment of color images based on a generic perceptual reduced reference," *Signal Processing: Image Communication*, vol. 23(4), pp. 239–256, 2008.

[8] T. Furon and P. Bas, "Broken arrows," *EURASIP Journal on Information Security*, vol. 2008, pp. 1–13, 2008.

[9] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," in *IEEE Trans. on Image Processing*, May 2005.