

# Towards a Simplified Perceptual Quality Metric for Watermarking Applications

Maurizio Carosi, Vinod Pankajakshan, Florent Autrusseau

June 19, 2009

IRCCyN-IVC, Polytech'Nantes, rue Ch. Pauc, 44306, Nantes, FRANCE

## Abstract

This work is motivated by the limitations of statistical quality metrics to assess the quality of images distorted in distinct frequency range. Common quality metrics, which basically have been designed and tested for various kind of global distortions, such as image coding may not be efficient for watermarking applications, where the distortions might be restricted in a very narrow portion of the frequency spectrum. We hereby want to propose an objective quality metric which performances do not depend on the distortion frequency range, but we nevertheless want to provide a simplified objective quality metric in opposition to the complex HVS based quality metrics recently made available. The proposed algorithm is generic (not designed for a particular distortion), and exploits the contrast sensitivity function (CSF) along with an adapted Minkowski error pooling. The results show a high correlation between the proposed objective metric and the mean opinion score (MOS). A comparison with relevant existing objective quality metrics is provided.

## 1 Introduction

During the last couple of decades, digital watermarking has significantly been improved regarding various aspects. Early works have mostly focused on robustness improvements, more recent research targeted an optimal security. However, although of great importance, the watermark invisibility is commonly ensured by means of MSE or PSNR. Objective quality metrics (OQM) have sometimes been exploited to ensure the watermark invisibility. The most commonly used quality metrics for digital watermarking are WPSNR [7] or SSIM [1]. The limitations of these metrics for watermarking applications were recently pointed out [2].

Evidently the most efficient way to assess the perceived quality of watermarked images is to run a subjective experiment where human observers are

asked to judge the quality of the displayed images. Subjective experiments are very restrictive and require a specific setup. Objective quality assessment have recently been of great interest to the human vision community, OQM are designed to provide the best approximation of the observers' quality score. We hereby propose an objective quality metric taking a very simple but yet important HVS feature into account, and a comparison with several widely accepted objective quality metrics.

Common objective quality metrics are usually made of five distinct steps: A preprocessing step can be used (such as the screen non linearity function), CSF filtering is then performed, masking effects are considered (via a perceptual channel decomposition), error normalisation can then be computed to provide a distortion map, and finally an error pooling on this map provides a predicted quality score (commonly called MOSp), which supposedly correlates well with the Mean Opinion Score (MOS) provided by human observers.

Objective quality metrics can be of three different kinds. No reference quality metrics would compute a MOSp based on a distorted image only, Reduced Reference Quality Metrics use both the distorted images and some features collected from the original image. Finally, Full Reference Quality Metrics use both the original and distorted image to predict the quality score. In watermarking context, Full Reference Quality Metric can be considered.

## 2 Proposed Model

Basically, two very important steps in objective metrics design are the contrast sensitivity consideration and a proper error pooling. The proposed metric is restricted to these two considerations incorporating appropriate Minkowski pooling parameters.

The sensitivity of the Human Visual System varies with many factors, including luminance level or contrast. The Contrast Sensitivity Function (CSF) (depicted in Figure 1) represents our sensitivity to spatial frequency. Using the CSF is essential when designing an objective quality metric, especially for watermarking applications, where the distortions might occur in different frequency range or different image areas. The 2D CSF (inset in Fig. 1) basically filters the Fourier spectrum of an image, thus enhancing the portions of the image having peak sensitivity.

The behavior of common OQM is strongly linked to the frequency range of the distortions. Figure 2 shows very distinctively three clusters corresponding to the predicted MOS for PSNR and SSIM when watermarks are added either in Low Frequency (LF), Middle Frequency (MF) or High Frequency (HF) bands (see details about the "Fourier database" in section 3). Watermarked images having the same PSNR (or SSIM) value might have very different perceived quality depending on the watermark frequency.

The Minkowski error pooling is commonly used as the final step of OQM, it is defined as

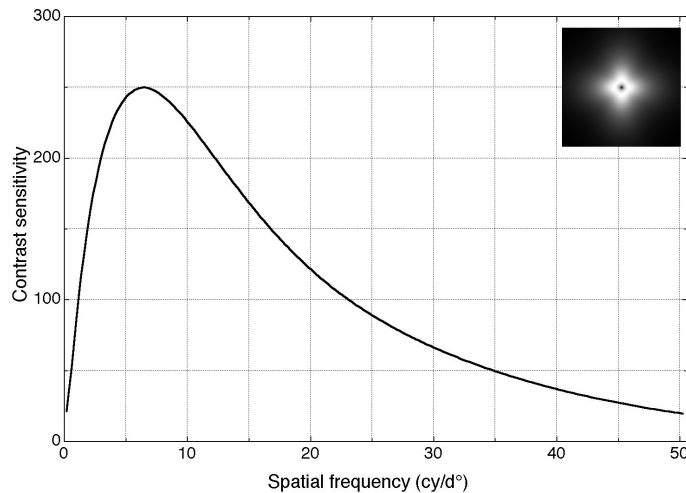


Figure 1: 1D CSF along with the 2D version.

$$Mink(P, R) = \sqrt[R]{\frac{1}{N} \sum_{i=1}^N |v_i|^P} \quad (1)$$

$N$  being the number of pixel in the image, and  $v(i)$  the absolute difference value at the  $i^{th}$  spatial location in the difference map. The generalized Minkowski summation, where  $P \neq R$ , provides additional flexibility for adjusting the response of individual parameters to changes in perceived quality. As a special case, Equation 1 reduces to the mean absolute error (MAE) when  $R = P = 1$ , and the MSE if  $P = 2$ . Finally if  $R = P = 2$ , the Minkowski summation is equivalent to the RMSE. As  $P$  increases, more emphasis will be put at the image regions of high distortions. Some appropriate values of  $R$  and  $P$  should provide a reasonable estimation of how humans rate image quality.

The different steps of the proposed objective quality metric are depicted on Figure 3. Both input images goes through a Fourier transform. Each spectrum is weighted by the CSF, thus emphasizing perceptually important frequencies. An inverse Fourier transform is computed, and an adapted Minkowski summation on the so-obtained weighted error map provides the MOSp.

### 3 Experimental setup

Among the four tested subjective databases<sup>1</sup>, three are based on digital watermarking algorithms operating in various transformed spaces, and a coding database was also used for comparison purpose. All subjective scores (MOS)

<sup>1</sup>Available online: <http://www.ircyn.ec-nantes.fr/~autrusse/Publications.html#DB>

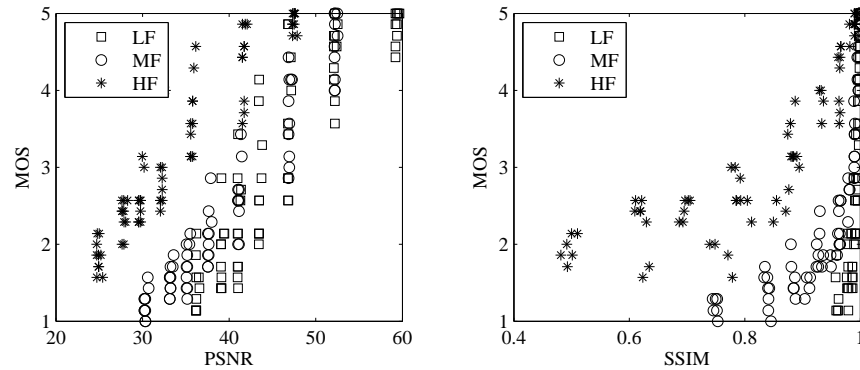


Figure 2: Subjective scores as a function of PSNR and SSIM for Fourier domain watermarking in three distinct frequency ranges.

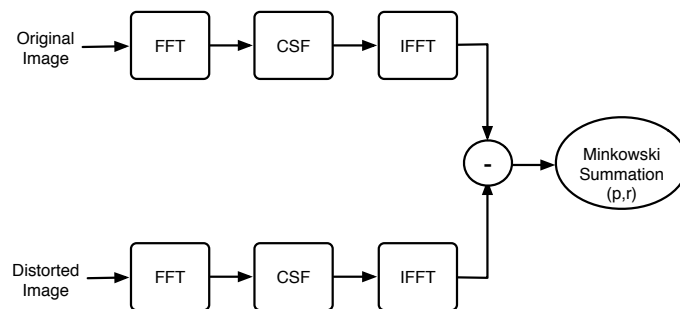


Figure 3: Block diagram representing the steps of the proposed OQM.

were collected in normalized viewing conditions as recommended by the International Telecommunications Union (ITU). Interested readers should refer to [2] for further details on the experimental setup. We hereby give a brief description of every database:

- Database 1<sup>2</sup>: 5 original grayscale images, 210 distorted images were generated from Gaussian noise addition in 3 different Fourier frequency range and two orientations per frequency range, with 7 embedding strengths. Noise patches were modulated onto frequency carriers and embedded independently in the six Fourier perceptual sub-bands ([2]) (7 observers participated in this experiment).
- Database 2<sup>3</sup>: 10 original grayscale images (5 natural images and 5 art images), 120 distorted images were generated with 6 different embedding strength using the “Broken Arrows” embedding technique [4] operating

<sup>2</sup>Details are provided here: <http://www.irccyn.ec-nantes.fr/~autrusse/Databases/FourierSB/>

<sup>3</sup>Thanks to Patrick Bas for providing the marked images

in the wavelet domain (target PSNR: 20, 24, 28, 32, 36 and 40 dB), and either with or without weighting coefficients (17 observers).

- Database 3<sup>4</sup>: 12 input grayscale images, 120 distorted images were generated from 5 embedding strength in two distinct embedding domains (DWT & Dual-Tree Complex Wavelet Transform) the same embedding equation applies in both domains:  $y = x + \alpha \times |x| \times w$  (4 observers).
- Database 4: 8 original color images were used (4 natural images and 4 art images), 120 distorted images were generated from 3 different codecs (JPG, JP2K and LAR [5]), and 5 compression rates (19 observers). This database was used for testing purpose, and to ensure the validity of the metric for non-watermarking distortion type.

The performance of the proposed quality metric was compared with twelve metrics from the “metrix\_mux” package<sup>5</sup> (MSE, PSNR, SSIM, MSSIM, VSNR, VIF, VIFP, UQI, IFC, NQM, WSNR, SNR), as well as C4[3] and WPSNR. Due to space limitation, we hereby present results for three widely used quality metrics in watermarking domain (PSNR, WPSNR, SSIM) and an advanced HVS based OQM (C4).

In this experiment, the Minkowski  $P$  and  $R$  parameters were tested in the range [1, 12] with a step of 0.5, best overall results were obtained with  $P = 5$  and  $R = 10$ . Such unusual range for Minkowski parameters is explained by the higher dynamic range of the CSF filtered images.

## 4 Results

In this work, the metric’s performances were evaluated by the analytical tools recommended by VQEG [6], we hereby present results regarding Linear/Spearman Correlations, RMSE and Outlier Ratio. Figure 4 presents the MOS plotted as a function of the MOS<sub>p</sub> for every tested metric on database 2. Evidently a narrow and linear distribution of points is highly desirable. Contrary to similarity metrics, the proposed metric provides a measure of the differences between original and distorted images, thus yielding to a negative slope.

Figure 5 presents the metrics’ performances in terms of Correlation, RMSE and Outlier Ratio for all databases. In order to match the objective scores within the subjective range a fitting function was used, as recommended by the VQEG [6].

Some important observations can be raised on this plot:

- Although very simple in its design, the proposed metric presents very good performances on all databases with regard to the four analytical measures. It is moreover the only metric providing reasonable MOS prediction on database 1.

<sup>4</sup>Thanks to Peter Meerwald for providing the marked images.

<sup>5</sup>[http://foulard.ece.cornell.edu/gaubatz/metrix\\_mux/](http://foulard.ece.cornell.edu/gaubatz/metrix_mux/)

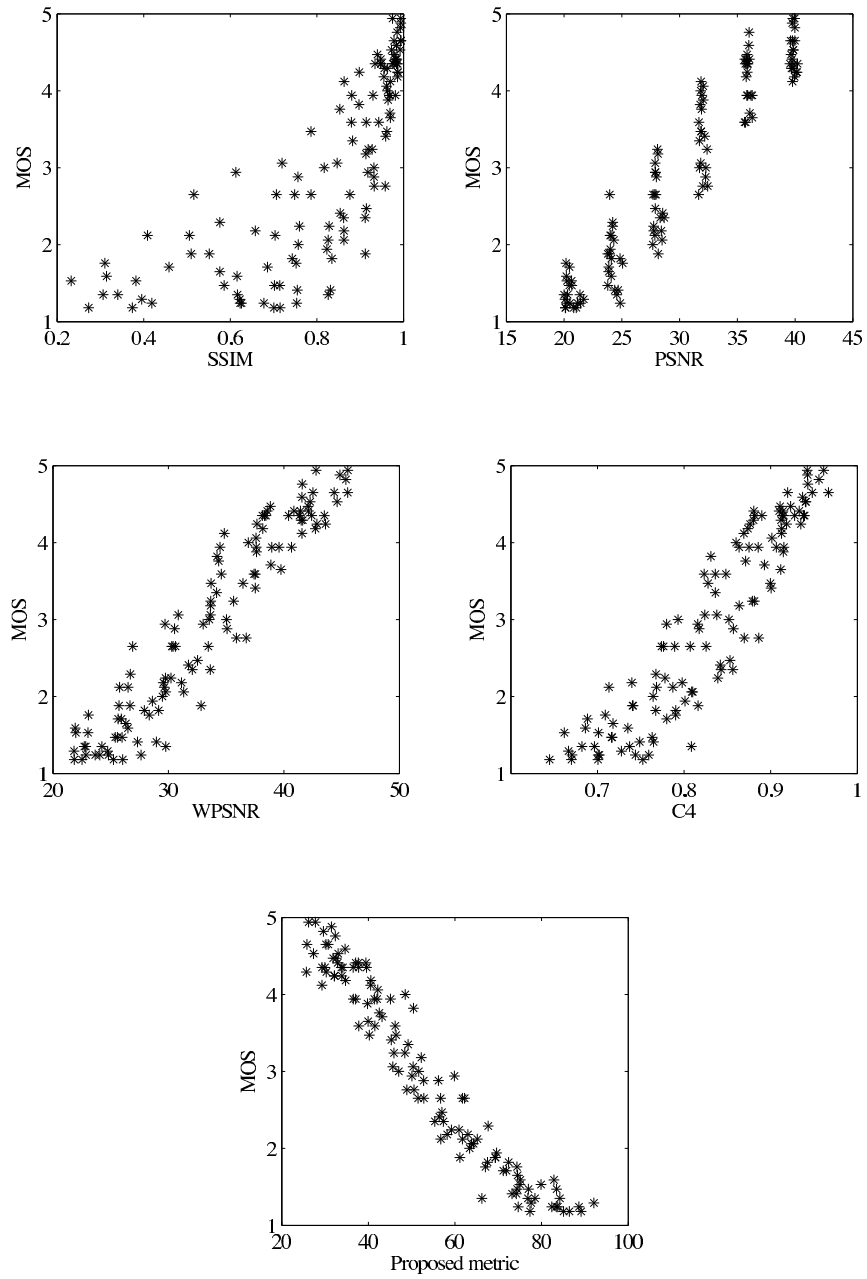


Figure 4: MOS versus MOSp plots for the 5 metrics on Database 2.

- SSIM and PSNR overall presents quite bad correlation with human judgement, and perform badly when the watermark is modulated on different frequency carriers (database 1).
- An important performance discrepancy is noticeable on database 1, this is actually due to the different frequency embedding, only the metrics considering the CSF might perform well on such database (C4 and the proposed metric provide the best correlation).
- The C4 metric, although being a reduced reference metric, provides overall satisfactory results. This metric has mostly been evaluated for coding distortions, and shows best performances on database 4.
- PSNR performs well on database 2, which is expected as the MOS<sub>p</sub> values are evenly distributed (“Broken Arrows” fixes a condition on the PSNR for watermark embedding). However, we can notice on Figure 4 a wide spreading of the MOS values, leading thus to higher outlier ratio and RMSE as compared to the proposed metric (cf. in Fig 5).

## 5 Conclusion

We have proposed in this work a simple yet efficient objective quality metric exploiting basic features of the Human Visual System. The proposed metric was successfully compared to four state-of-the-art metrics on four subjective databases of various distortion content. Our metric proved to be efficient on the four tested subjective databases, although its complexity is significantly lowered compared to complex HVS modeling exploited in C4.

The narrow distribution of MOS versus MOS<sub>p</sub> for the proposed metric suggests that a visibility threshold (JND mask) could be set. The next step of our research will focus on improving the metric by incorporating a simplified masking model.

## References

- [1] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, Eero P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity”, *IEEE Transactions on Image Processing*, 13, (4), 2004
- [2] Patrick Le Callet, Florent Autrusseau, Patrizio Campisi, “Visibility control and Quality assessment of watermarking and data hiding algorithms”, Chap IX in *Multimedia Forensics and Security*, Idea Group, pp. 163-192, 2008. ISBN: 978-1-59904-869-7
- [3] Mathieu Carnec, Patrick Le Callet, Dominique Barba, “Objective quality assessment of color images based on a generic perceptual reduced reference”, *Signal Processing: Image Communication*, 23 (2008) pp. 239–256.

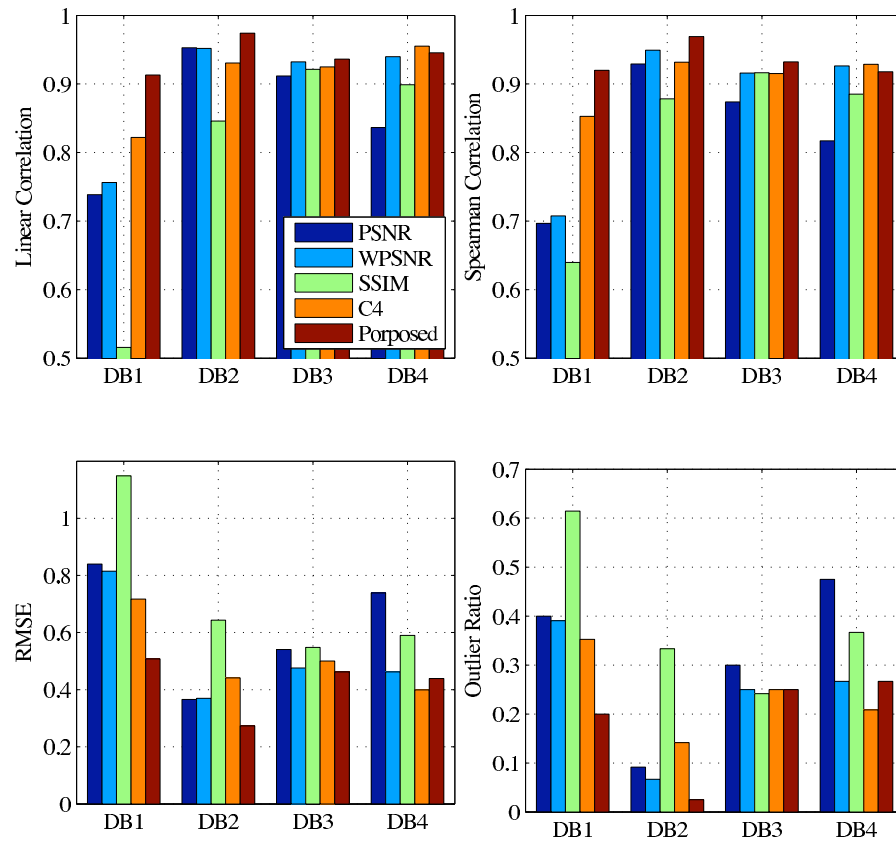


Figure 5: Metrics performances in terms of Linear Correlation, Spearman correlation, RMSE and Outlier Ratio for all databases.

- [4] Teddy Furon and Patrick Bas, “Broken Arrows”, *EURASIP Journal on Information Security* (2008), Article ID597040, 13p, doi:10.1155/2008/597040
- [5] Olivier Déforges, Marie Babel, Laurent Bédard, and Joseph Ronsin, “Color LAR Codec: A Color Image Representation and Compression Scheme Based on Local Resolution Adjustment and Self-Extracting Region Representation,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, (8), pp. 974–987, 2007.
- [6] VQEG. (2000) ”Video Quality Experts Group: Current Results and Future Directions” , VCIP 2000.
- [7] Sviatoslav Voloshynovskiy, Shelby Pereira, Alexander Herrigel, Nazanin Baumgartner, Thierry Pun, “Generalized watermarking attack based on watermark estimation and perceptual remodulation”, *SPIE Security and watermarking of multimedia contents*, 3971, 358-370, 2000 .