# Pseudo No Reference image quality metric using perceptual data hiding

Alexandre Ninassi, Patrick Le Callet, Florent Autrusseau

Ecole Polytechnique de l'Universite de Nantes, IRCCyN lab., Rue Christian Pauc, La Chantrerie, BP 50609, 44306, Nantes Cedex3, France

## ABSTRACT

Regarding the important constraints due to subjective quality assessment, objective image quality assessment has recently been extensively studied. Such metrics are usually of three kinds, they might be Full Reference (FR), Reduced Reference (RR) or No Reference (NR) metrics. We focus here on a new technique, which recently appeared in quality assessment context: data-hiding-based image quality metric. Regarding the amount of data to be transmitted for quality assessment purpose, watermarking based techniques are considered as pseudo no-reference metric: A little overhead due to the embedded watermark is added to the image. Unlike most existing techniques, the proposed embedding method exploits an advanced perceptual model in order to optimize both the data embedding and extraction. A perceptually weighted watermark is embedded into the host image, and an evaluation of this watermark allows to assess the host image's quality. In such context, the watermark robustness is crucial; it must be sufficiently robust to be detected after very strong distortions, but it must also be sufficiently fragile to be degraded along with the host image. In other words, the watermark distortion must be proportional to the image's distortion. Our work is compared to existing standard RR and NR metrics in terms of both the correlation with subjective assessment and of data overhead induced by the mark.

## 1. INTRODUCTION

The main goal of image quality assessment is to find an automatic metric which provides computed quality scores well correlated with the ones given by human observers. Image quality metrics can be divided in three categories:

- full reference metrics (FR) for which both the original image and the distorted one are required,
- reduced reference metrics (RR) for which a description of the original image into some parameters and the distorted image are both required,
- no reference (NR) metrics which only require the distorted image.

In QoS monitoring, only RR and NR metrics are acceptable since transmitting the whole reference image is not realistic at all. Ideally for such applications, NR metrics are preferred since no extra data is added to the bit stream. As an alternative to NR metrics, we propose an objective quality metric based on data hiding. But such technique can neither be considered as NR metric, nor as RR metric : no content description is transmitted, but a little overhead is added to the image. Actually, this work is based on the idea that the embedded data is affected by distortions in the same way as the initial content. Thus, assessing the embedded data quality corresponds to assess the host image's quality.

Data hiding techniques are typically used for several purposes: fingerprinting, multimedia indexing, context base retrieval, etc. Recently, embedding techniques have also been used to estimate video quality at the receiver [1, 2, 3, 4, 5].

An embedding system with copyright protection purpose has to satisfy two main constraints:

- Invisibility : the mark should not affect the perceptual quality.
- Robustness : the mark cannot be altered by malicious (an attempt to alter the mark) or unintentional (compression, transmission) operations.

The requirement are quite similar for quality assessment purpose. Invisibility, for example, is very important. Concerning the robustness requirement, the mark has to be sufficiently robust, to be detected in a very poor quality image, but it also has to be sufficiently fragile to be distorted proportionally to the image distortions. Furthermore, it is important to notice that increasing the robustness generally increases the watermark's visibility. If the mark is too fragile, the extracted mark will be lost for small distortions making it difficult to differentiate between medium or highly degraded videos. We expect the embedded watermark to be semi-fragile and degrade at around the same rate as the host media.

One of the most advanced work in this topic have been proposed by Farias et al. [1, 2] for video. In their work, a two-dimensional binary mark is embedded in the DCT domain. A spread-spectrum technique is employed to hide the mark, by using a set of uncorrelated pseudo-random noise (PN) matrices (one per frame) which are later multiplied by the reference mark (the same for the whole video). Unfortunately, embedding marks into images or video may introduce unwanted distortions or artifacts degrading the perceived quality. The visibility and annoyance of these artifacts depend on several factors like the domain where the mark is being inserted, the embedding algorithm, and the mark's strength. To tackle this issue, Farias have included in the design of the system a psychophysical experiment to evaluate the visibility and annoyance of the artifacts caused by the embedding algorithm. The results show that the choice of either mark image does not significantly affect the visibility and annoyance of the embedding impairments. The annoyance and psychometric functions considerably vary depending on the physical characteristics of the particular video. This is probably due to the masking effect that varies along with the content. To avoid such empirical approach and take benefit from recent models of masking effect, we propose in this paper an embedding method based on a psychovisual model that allows to analytically control the mark visibility. We exploit this technique to assess quality of still color images and we compare quality metric performances with classical metrics found in the literature.

This paper is decomposed as follows. Section 2 is devoted to the watermarking technique, the watermark embedding and detection processes are both presented. The quality assessment metric is presented in section 3, where the choice of the frequency sub-bands as well as the watermark size are argued. Finally, some results are given in section 4, where comparisons with other existing techniques are shown.

## 2. THE WATERMARKING TECHNIQUE

The embedding technique used here is based on a robust perceptual watermarking scheme designed for copyright protection purpose [6]. Here, the authors opted for a strictly localized frequency content watermark embedding. To fulfill the optimal perceptual constraint, the watermark strength is adapted using a visual mask established from an advanced human visual system model. This visual mask provides quantization noise visibility thresholds for each spatial image site. This perceptual model takes into account very advanced features of the Human Visual System (HVS), fully identified from psychophysics.

### 2.1. Perceptual mask

Like in most approaches, we use a sub-band decomposition defined by analytic filters for luminance supposed to describe the different channels of the human vision system and so the visual filtering. Previous study have been conducted in our lab in order to characterize this decomposition (see figure 2), the experiments were based on the measurement of the masking effect between two complex narrow band limited signals. For still images, we need to use four radial frequency channels, one low-pass called I with radial selectivity 0 cy/deg to 1.5 cy/deg and three bandpass called II, III, IV with radial selectivity respectively 1.5 cy/deg to 5.7 cy/deg, 5.7 cy/deg to 14.1 cy/deg, 14.1 cy/deg to 28.2 cy/deg. The three bandpass are decomposed into angular sectors associated with orientation selectivity. The angular selectivity is 45deg for sub-band II and 30deg for sub-bands III and IV. The masking effect model is based on the visibility produced by quantizing the content of a particular sub-band rather than the visibility of any increments or any white gaussian noises. We have previously shown that perception of quantization noise on $L_{i,j}$ at location (m,n) is directly dependent on the ratio between $L_{i,j}$ and the average luminance at this location. This latter is computed from the sub-bands having a lower radial frequency. This ratio is therefore a local contrast $C_{i,j}$ given by :

$$C_{i,j}(m,n) = \frac{L_{ij}(m,n)}{\sum_{k=0}^{i-1} \sum_{l=0}^{Card(l)} L_{k,l}(m,n)} \tag{1}$$

Psychovisual tests performed on the different visual channels have shown that local contrasts must always be uniformly quantized in order to achieve a just noticeable quantization law, the quantization step being dependent of the considered visual sub-band. Inter-channels luminance masking effect is partially taken in account by this model, the model fails for masking effect along directional adjacency. So we have completed this model with further experiments and it has been successfully implemented in a visual coding scheme. In a watermarking context the model is very useful since it provides the maximum luminance variation that can be applied for each (i,j) sub-band and for each (m,n) pixel position without providing visible artifacts. We can define a spatial mask given by

$$\Delta L_{i,j}(m,n) = \Delta C_{i,j} \times \overline{L_{i,j}}(m,n) \tag{2}$$

$\Delta C_{i,j}$ are the quantization thresholds measured from psychophysics experiments for each i,j sub-band and $\overline{L_{i,j}}(m,n)$ is the local mean luminance for the i,j sub-band and for each (m,n) position.

### 2.2. Watermark embedding

According to the chosen watermark embedding algorithm [6], a frequency domain watermark (noise) restricted into a single visual sub-band is then built and its spatial representation is computed. Finally, this spatial watermark with limited frequency content is scaled according to its corresponding visual mask.

As mentioned in the introduction, our aim in this work is to use a perceptually optimized watermarking scheme able to resist to most attacks and especially to geometrical distortions. The watermark amplitude must be weighted according to the visual mask. Although this visual mask is spatially defined, the Fourier transform linearity allows to use the same weighting coefficient independently in the spatial or Fourier domain. A Fourier coefficients watermark is then built and modulated onto a frequency carrier. Finally a perceptual weighting coefficient $K_{i,j}$ is computed from the watermark's spatial domain representation and the sub-bands dependent visual mask. $K_{i,j}$ is given in equation 3

$$K_{i,j} = argmin_{m,n}\left(\left|\frac{\Delta L_{i,j}(m,n)}{W_S(m,n)}\right|\right) \tag{3}$$

where $\Delta L_{i,j}(m,n)$ represents the previously defined visual mask and $W_S(m,n)$ depicts the watermark's spatial representation before weighting process by factor $K_{i,j}$ for each $(m,n)$ spatial position. Figure 1 summarizes both the perceptual mask creation steps (upper branch) and the watermark weighting process (lower branch).

### 2.3. Watermark detection

Regarding to the embedding process, the extraction technique is straightforward, the cross-correlation function computed between the watermark and the extracted marked and possibly attacked Fourier coefficients. A correlation peak will prove the watermark presence in these coefficients.

The main advantages of this method are:

- The control of the mark visibility
- The watermark is content independent (whereas the weighting coefficient is image dependent).

For image quality assessment, only the watermarked image is transmitted. The detection process performs a cross-correlation between the stored watermark and the Fourier coefficients surrounding the known frequency carrier extracted from the marked image. This cross-correlation values are then compared to a detection threshold in order to guarantee the watermark presence in the modified coefficients. The only needed data for the retrieval procedure are the original watermark, and its frequency carrier.

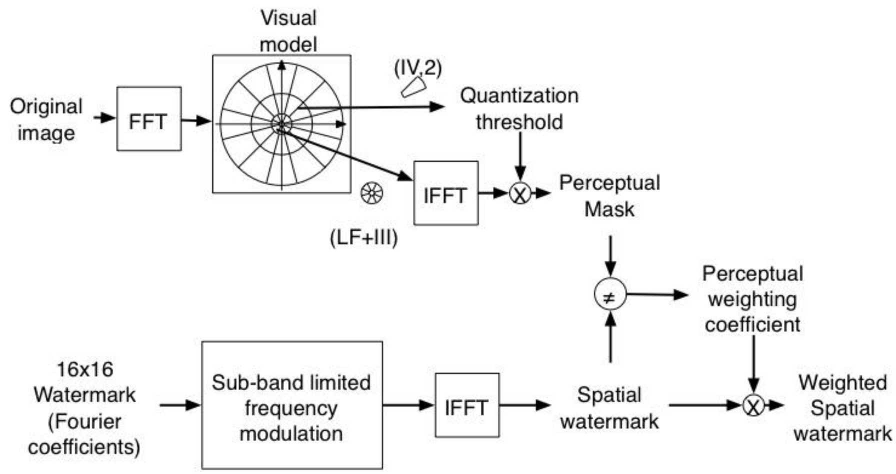



**Figure 1.** Watermark embedding technique

## 3. THE QUALITY METRIC

The proposed watermarking technique allows to embed the watermark in different frequency and orientation range. For quality assessment purpose, we have chosen to embed several marks, as distortions may affect different parts of the Fourier spectrum.

In our metric we embed marks both in the middle and high frequencies of the image. For each mark, the previous watermarking scheme is used. We select several sub-bands of the perceptual channel decomposition (PCD) so that the presented metric has several measuring points on the frequency content of the source image.

The mark modifies the original image in an invisible way for an observer, however the image content is definitely modified. The visual mask being content dependant, it is not possible to compute the visual masks once and for all, and then to embed all the marks. The watermark embedding in a given sub-band of the chosen perceptual channel decomposition will modify the visual masks of all the higher sub-bands. In order to guarantee the invisibility of the multiple embedding technique, we must calculate a new visual mask after each single watermark embedding, respecting masking precedence relationship.

For this study, 8 watermarks have been embedded :

- 6 watermarks (10x10 coefficients) in high frequencies (one mark per sub-band)
- 2 watermarks (8x8 coefficients) in the middle frequencies

The figure 2 schematically depicts the 8 chosen watermarks superimposed to the HVS decomposition (PCD). Obviously, the Fourier spectrum symmetry is respected and the bottom part of the spectrum is filled with the symmetric watermarks.

After applying an image processing (compression scheme or filtering), we measure for each mark the cross-correlation (cf. figure 3) between the original mark and the corresponding coefficients of the watermarked image. Therefore, we get 8 cross-correlation maximum values (one per mark). The quality score is obtained in two steps: we first compute the mean *Mhf* of the cross-correlation obtained with the six high frequency marks, and the

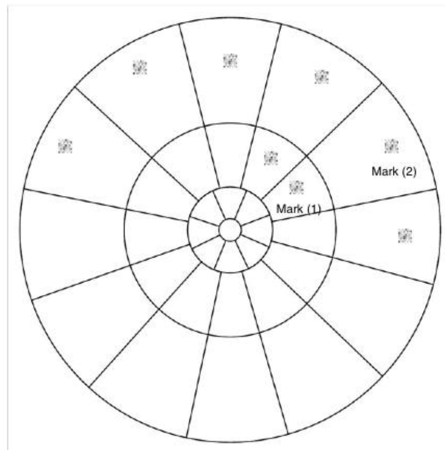
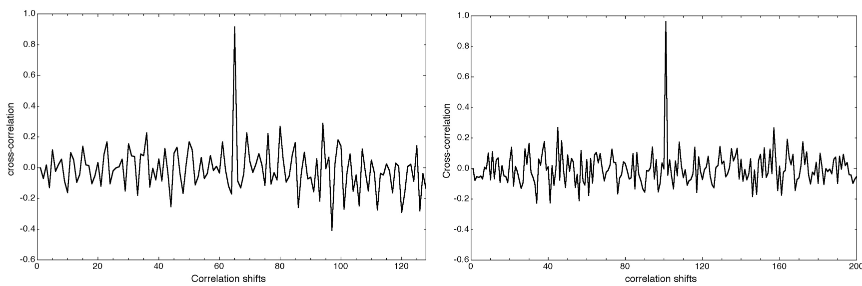



**Figure 2.** Spectrum of a multi-embedding. Frequency carries and watermarks




(a) Mark (1) from figure 2 (b) Mark (2) from figure 2

**Figure 3.** Cross-correlation example of a MF (a) and a HF (b) mark

mean $Mmf$ of the cross-correlation obtained with the two middle frequency marks. The quality score is finally given by computing the mean between $Mhf$ and $Mmf$.

The obtained quality score $Q$ is in the range [0, 1]. A psychometric function should be used in order to map objective quality scores $Q$ in the range of subjective quality scores MOS (Mean Observer Score). This methodology is approved and recommended by VQEG (Video Quality Experts Group)*. The psychometric function used in our case is the function with 3 parameters given by the equation:

$$MOSp = \frac{b1}{1 + e^{-b2*(Q-b3)}}, \tag{4}$$

where MOSp is the predicted MOS, Q is the quality score given by the metric, and b1, b2 and b3 are the parameters of the psychometric function.

MOS have been obtained conducting subjective quality assessment experiments in our lab in normalized conditions. From 10 original images, we get 170 images from 3 different processing : JPEG, JPEG2000 and Blurring. These algorithms have the advantage to generate very different type of distortions. Subjective evaluations were made at viewing distance of 6 times the screen height using a DSIS (Double Stimulus Impairment Scale) method

*http://www.its.bldrdoc.gov/vqeg/projects/rrnr-tv/index.php with 5 categories and 15 observers. Distortions for each processing and each image have been optimised in order to uniformly cover the subjective scale. When the MOSp are calculated, it is possible to compare the metric scores with subjective scores. In our case, we use RMSE between MOS and MOSp, linear correlation coefficient (CC) between MOS and MOSp. We use also outlier ratio that denotes the relative number of conditions for which the difference between MOS and MOSp is greater than twice the interval of confidence on MOS value.

Before finding the final combination of the cross-correlation values, various combinations were tested :

- $Q1$ : the mean of the 8 values *Mhf* and *Mmf*.
- $Q2(\alpha)$ : a linear combination of the mean *Mhf* (6 HF values) and *Mmf* (2 MF values) †

| Distortions | | $Q1$ | $Q2(\alpha)$ |
|---|---|---|---|
| All database | RMSE | 0,840 | 0,784 |
| | CC | 0,739 | 0,777 |
| All color database | RMSE | 0,860 | 0,791 |
| | CC | 0,722 | 0,771 |
| JPEG2000 | RMSE | 0,955 | 0,955 |
| | CC | 0,743 | 0,774 |
| JPEG | RMSE | 0,687 | 0,688 |
| | CC | 0,824 | 0,823 |
| Blur | RMSE | 0,828 | 0,828 |
| | CC | 0,926 | 0,926 |

**Table 1.** Comparison between the different combinations

Overall, in the above table the best value is $Q2(\alpha)$, which minimizes the RMSE (Root Mean Square Error) and maximizes the Correlation Coefficient (CC) (cf figure 4). This value is obtained for $\alpha = 0,5$, so $Q2$ happens to be the mean of *Mhf* (6 HF values) and *Mmf* (2 MF values).

By comparing $Q1$ and $Q2$, we can notice that $Q2$ is much better on all the database, as well in terms of minimization of the RMSE, as in terms of maximization of the CC. Besides, we can do the same observation on the subsets of images corresponding to the color images and the JPEG2000 attacks. For the JPEG and blur attacks subsets, the results of $Q1$ and $Q2$ are roughly the same. This is why we chose the combination $Q2$ for the quality score computation.

## 4. RESULTS

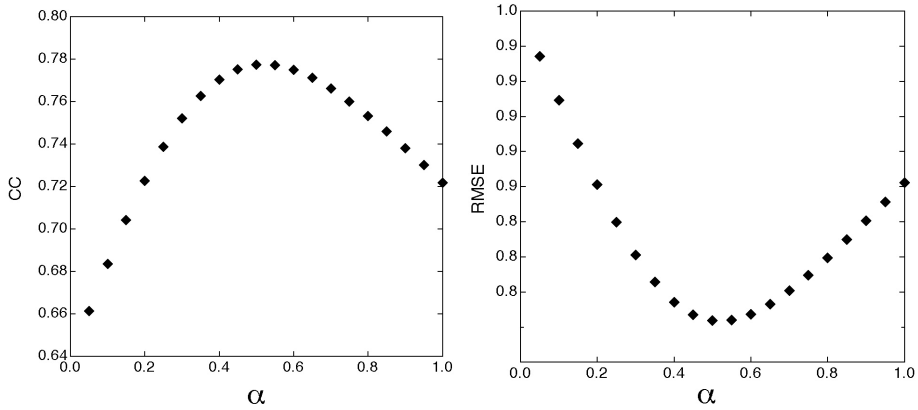
Our metric is compared with 4 others metrics of the literature. Two of them are NR metric, one dedicated to JPEG distortions [7] and the other to JPEG2000 distortions[8]. The other two are generic RR metrics [9, 10].

### 4.1. Full database

For the full database, only RR metrics can be compared with our metric. This latter clearly outperforms Wang RR and is slightly better than Carnec RR (see table 2). However, it is important to notice that the Carnec RR metric is designed for color images quality assessment, whereas 20 monochromatic images are included in the used database. Thus, it appears that the presented metric is appropriate for both gray scale and color images, and it does not need any *a priori* knowledge on the distortions. When we restrict the database to only color images, the Carnec metric is clearly the best (see table 3).

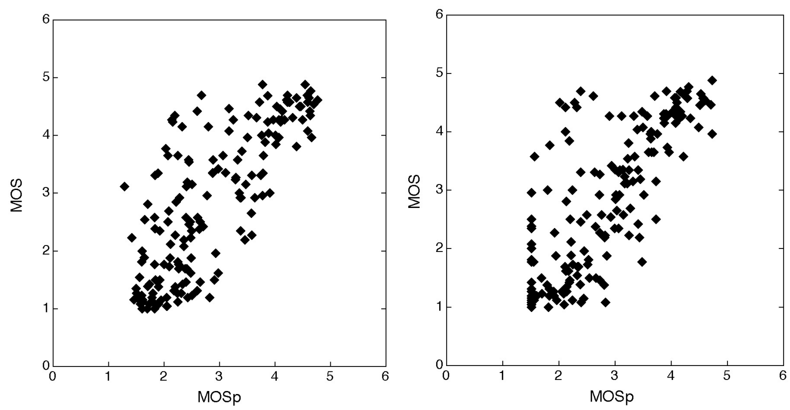
We notice on figure 5, that the set of points of our metric and that Carnec RR metric are the closest to the line $MOS = MOSp$.

† The linear combination is : $Q(\alpha) = \alpha * Mhf + (1-\alpha) * Mmf$, with $\alpha \in [0,1]$.

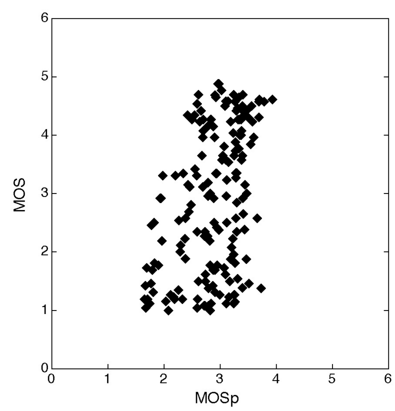



(a) (b)

**Figure 4.** CC and RMSE variation according to $\alpha$ for $Q2(\alpha)$




(a) proposed metric (b) Carnec RR metric




(c) Wang RR metric

**Figure 5.** MOS according to MOSp for all database.

Figures 6 and 7 represents the MOS and MOSp obtained with our metric for JPEG2000 and JPEG distortions respectively. On these plots, x-axis refers to the degraded images and y-axis refers to the MOSp value. For each image, 5 degraded versions are represented ordered by level of increasing degradation, the degraded images are gathered according to their original version. For example, on the figure 6, the first 5 ticks on the x-axis correspond to 5 increasing JPEG2000 compression rates for the *plane* image.

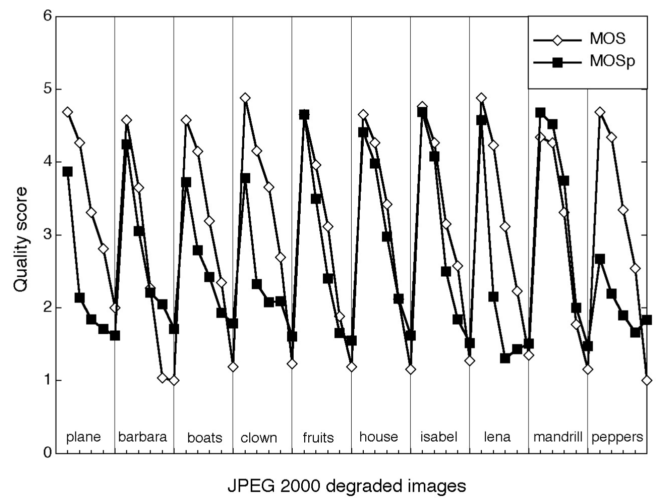



**Figure 6.** MOS and MOSp according to JPEG2000 distortions

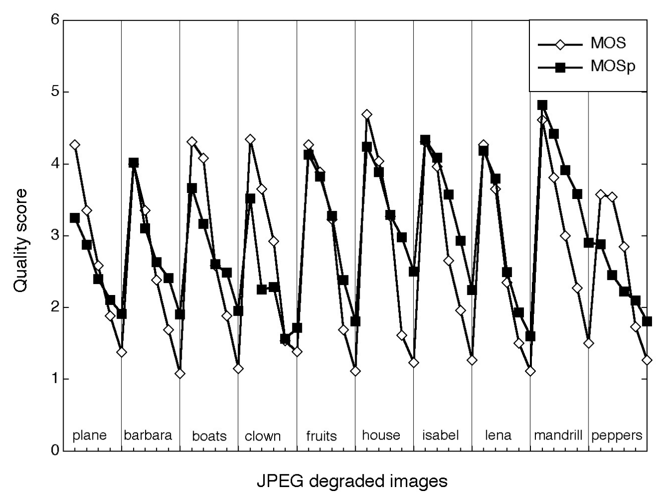
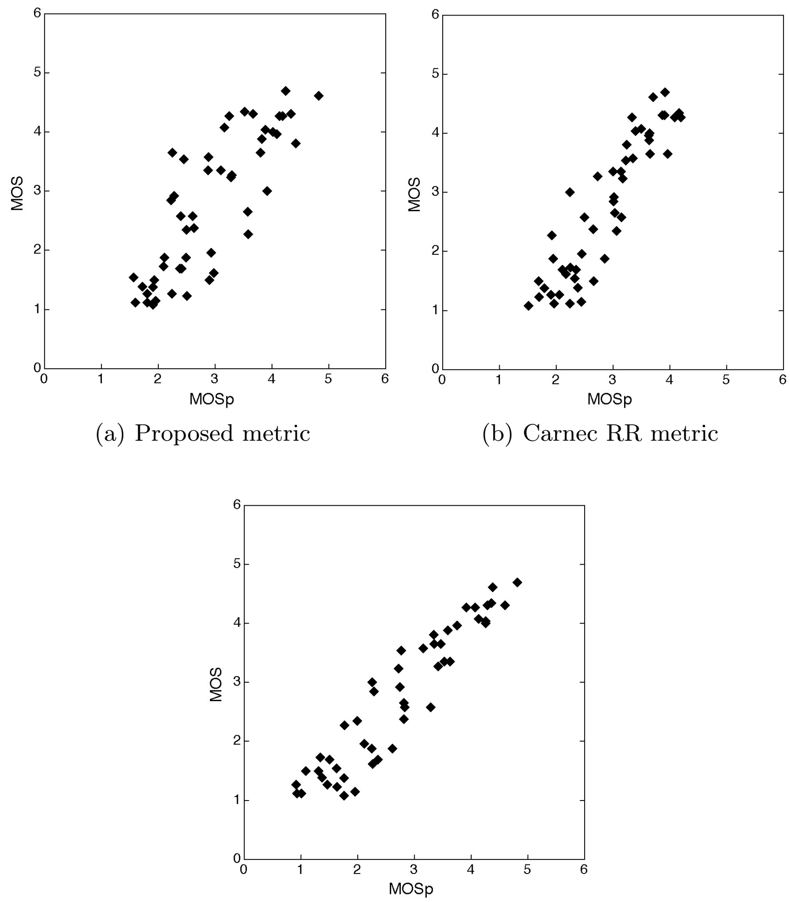



**Figure 7.** MOS and MOSp according to JPEG distortions

Performance are affected by image content. Images like *fruit*, *house* and *isabel* give excellent results where the MOSp and the MOS almost overlap. On the other hand, for some images like *plane* and *peppers* the matching is not that accurate. It appears that the images containing many textures and contours (thus most of HF) gets a much better evaluation than the images containing many uniform areas and few textures and contours (like *peppers*). This is not so surprising regarding to the watermark embedding frequencies. Nevertheless, the HVS model used for visual mask computation is not robust enough for low frequency embedding.

### 4.2. JPEG distortions

On the JPEG distortions subset, the comparison can be performed with Wang NR metric specifically designed for JPEG compressed images assessment.




(a) Proposed metric (b) Carnec RR metric

(c) Wang NR metric

**Figure 8.** MOS according to MOSp for JPEG distortions.

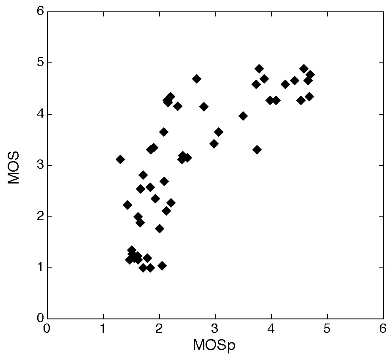
According to table 4 the Wang NR metric presents the best results for all three indicators. The Carnec RR metric also presents interesting results. Our metric obtains acceptable results, but it remains weaker than the two others. It is interesting to notice that the best MOS match is given by a NR metric, note though that this NR metric is based on an *a priori* knowledge on the distortions.

These performances are clearly depicted on figure 8, where the Wang NR metric provides a very good correspondence between the MOS and the MOSp.

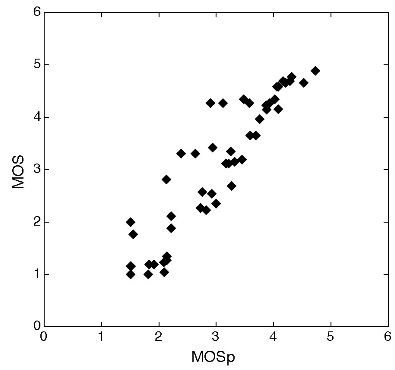
### 4.3. JPEG2000 distortions

Concerning the JPEG2000 distortions subset, we can add the Sheikh NR metric for a performance comparison. This latter is adapted to JPEG2000 compressed images assessment.

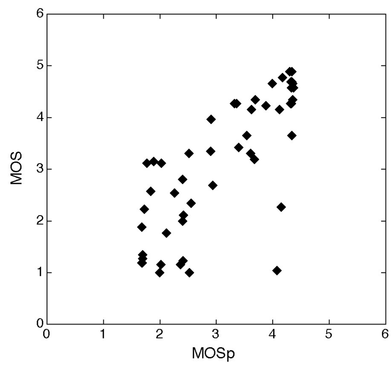
The Carnec RR metric appears to be the only efficient one, it reaches acceptable values for all three indicators (table 5). Figure 9 clearly shows the best results from the Carnec metric, as well as the bad outlier ratio obtained by both the Sheikh metric and ours. Performances of both Sheikh NR metric and ours are comparable (the Sheikh metric gets slightly better results).




(a) Proposed metric




(b) Carnec RR metric




(c) Sheikh NR metric

**Figure 9.** MOS according to MOSp for JPEG2000 distortions.

## 5. CONCLUSION

We proposed in this paper a new image quality assessment metric exploiting the data hiding principle. This metric can be regarded as *pseudo* NR, and it is not based on any *a priori* knowledge on the distortions. The watermarking technique exploits an advanced HVS model in order to ensure both the mark's invisibility and its robustness. In this application, the watermark has to be sufficiently robust to be retrieved after strong image distortions, but it also must be distorted proportionally to the host image. The quality metric performances have been compared to other standard metrics of the literature, and a correlation factor with the predicted visual quality (subjective assessment) is given. Overall, our metric provides the best results on the whole subjective database (several distortion types on color and monochromatic images). Nevertheless, while restricting the database to color images or JPEG2000 distortions the Carnec metric is the most suitable. The Wang metric is the most appropriate for JPEG distortions. Even though our metric is the most robust, the obtained results point out the difficulty to design an efficient generic NR or RR metric.

## References

[1] M. Farias, M. Carli, J. Foley, and S. Mitra, "Detectability and annoyance of artifacts in watermarked digital videos," in *EUSIPCO'2002, Proc. XI European Signal Processing Conference (EUSIPCO), Toulouse, France*, 2002.

[2] M. Farias, M. Carli, and S. Mitra, "Video quality objective metric using data hiding," in *MMSP'2002, Proc. IEEE Workshop on Multimedia Signal Processing (MMSP)* **3**, pp. 464–467, 2002.

[3] M. Farias, S. Mitra, M. Carli, and A. Neri, "A comparison between an objective quality measure and the mean annoyance values of watermarked videos," in *ICIP'2002, in Proc. IEEE Intl. Conf. on Image Processing, Rochester, NY, USA* **3**, pp. 469–472, 2002.

[4] O. Sugimoto, R. Kawada, M. Wada, and S. Matsumoto, "Objective measurement scheme for perceived picture quality degradation caused by mpeg encoding without any reference pictures," in *SPIE Electronic Imaging, in Proc. SPIE Human Vision and Electronic Imaging, SanJose, CA, USA* **4310**, pp. 923–939, 1998.

[5] M. Holliman and M. Young, "Watermarking for automatic quality monitoring," in *SPIE Electronic Imaging, Proc. SPIE Security and Watermarking of Multimedia Contents, San Jose, CA, USA* **4675**, 2002.

[6] F. Autrusseau and P. LeCallet, "Quantization noise visibility thresholds applied to fourier domain watermarking technique," *submitted in IEEE, International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 14-19, Toulouse, France*, 2006.

[7] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *ICIP'2002, IEEE International Conference on ImageProcessing*, 2002.

[8] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: Jpeg2000," *IEEE Transactions on ImageProcessing* , 2002.

[9] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statisticmodel," in *EUSIPCO'2002, Proc. XI European Signal Processing Conference (EUSIPCO), Toulouse, France*, 2002.

[10] M. Carnec, P. LeCallet, and D. Barba, "A new image quality assessment method with reduced reference," in *SPIE VCIP, VCIP 2003*, 2003.

| Metrics | RMSE | CC | Outlier Ratio |
|---|---|---|---|
| Our metric | 0,784 | 0,777 | 45,29% |
| Carnec RR | 0,807 | 0,762 | 44,71% |
| Wang RR | 1,122 | 0,434 | 70,00% |

**Table 2.** Comparison between different quality metrics on the full database.

| metrics | RMSE | CC |
|---|---|---|
| Wang RR | 1.115 | 0.443 |
| Carnec | 0.628 | 0.890 |
| proposed | 0.791 | 0.771 |

**Table 3.** Comparison between different quality metrics on the color database.

| Metrics | RMSE | CC | Outlier Ratio |
|---|---|---|---|
| Our metric | 0,688 | 0,823 | 46,00% |
| Carnec RR | 0,585 | 0,911 | 34,00% |
| Wang NR | 0,396 | 0,940 | 22,00% |

**Table 4.** MOS according to MOSp for JPEG distortions.

| Metrics | RMSE | CC | Outlier Ratio |
|---|---|---|---|
| Our metric | 0,955 | 0,774 | 54,00% |
| Carnec RR | 0,560 | 0,921 | 36,00% |
| Sheikh NR | 0,822 | 0,771 | 44,00% |

**Table 5.** Comparison between different quality metrics on JPEG2000 distortions.