Which Semi-Local Visual Masking Model For Wavelet Based Image Quality Metric?

Alexandre Ninassi, Olivier Le Meur, Patrick Le Callet, Dominique Barba

WHICH SEMI-LOCAL VISUAL MASKING MODEL FOR WAVELET BASED IMAGE QUALITY METRIC?

A. NINASSI[1,2], *O. LE MEUR*[1]

[1]Thomson Corporate Research
1 Avenue Belle Fontaine
35511 Cesson-Sevigne, France

P. LE CALLET[2], *D. BARBA*[2]

[2]IRCCyN UMR 6597 CNRS
Ecole Polytechnique de l'Universite de Nantes
rue Christian Pauc, La Chantrerie
44306 Nantes, France

ABSTRACT

Properties and models of the Human Visual System (HVS) are the fundaments for most of efficient objective image or video quality metrics. Among HVS properties, visual masking is a sensitive issue. Many models exist in literature. Simplest models can only predict visibility threshold for very simple cue while for natural images one should consider more complex approaches such as semi-local masking. Our previous work has shown the positive impact of incorporating semi-local masking in image quality metric according to one subjective study. It is important to consolidate this work with different subjective experiments. In this paper, different visual masking models, including contrast masking and semi-local masking, are evaluated according to three subjective studies. These subjective experiments were conducted with different protocols, different types of display devices, different contents and different populations.

Index Terms— Quality Assessment, Human Visual System, Contrast Masking, Semi-local Masking, Entropy Masking

1. INTRODUCTION

The purpose of an objective image quality evaluation is to automatically assess the quality of images or videos in agreement with human quality judgments. Over the past few decades, image quality assessment has been extensively studied and many different objective criteria have been built. Quality metrics based on models of the HVS are an important part of the different approaches in image quality assessment. HVS models may be categorized into mono-channel or multi-channel models, and this work focuses on the latter. In order to simulate the multi-channel behavior of the HVS and to well qualify the visual masking effects, this kind of quality metrics rests on a perceptual subband decomposition. In a previous work [1], we have studied the impact of semi-local masking in a wavelet based quality assessment (WQA) metric. This work has shown, among others things, that contrast masking is positively completed by semi-local masking. In image or video quality assessment, the results are sensitive to the subjective data used. It is important to complete this work with other subjective data. In this new study, experimental data are collected from several subjective experimentations. These experiments were conducted using two protocols of test (DSIS and ACR), on two populations (French and Japanese), with two types of display devices (LCD and CRT), and with two image databases.

In this paper, an image quality metric based on a multi-channel model of the HVS using wavelet domain is described. The HVS model of the low-level perception used in this metric includes subband decomposition, spatial frequency sensitivity, contrast masking and semi-local masking. The subband decomposition of this multi-channel approach is based on a spatial frequency dependent wavelet transform. The spatial frequency sensitivity of the HVS is simulated by a wavelet contrast sensitivity function (CSF) derived from Daly's CSF [2]. Masking effects include both contrast masking and semi-local masking. Semi-local masking allows to consider the modification of the visibility threshold due to the semi-local complexity of an image. This phenomenon is also called entropy masking [3], activity masking, texture masking, or local texture masking [4]. Due to the influence of the neighborhood characteristics, this masking effect will be called semi-Local Masking (sLM) in the rest of the paper. The focus of this work is to evaluate the impact of semi-local masking on image quality assessment, and to consolidate the results of a previous study [1] with three subjective studies.

In order to investigate its efficiency, the WQA metric is compared with subjective ratings and the state-of-the-art measure of structural similarity (SSIM) [5]. The WQA metric is tested with and without semi-local masking, giving insight into the relevance of the semi-local masking.

This paper is organized as follows. Section 2 is devoted to the description of the WQA metric. Several versions of the WQA metric using different masking functions are compared in section 3. Finally, general conclusions are provided.

2. QUALITY METRIC DESCRIPTION:WQA

In this section the WQA metric is described. Its structure is illustrated in Figure 1. As mentioned before, the HVS model of the low level perception used in this metric includes subband decomposition, spatial frequency sensitivity, contrast masking and semi-local masking.

The versions of WQA used in this work are achromatic versions. The first step consists of adaptation. Adaptation describes the changes that occur due to different illumination levels in the visual sensibility of lightness.

2.1. Subband decomposition

A subband decomposition defined by wavelet filters is used and supposed to describe the different channels in the human vision system. The correspondence between the visual system and the wavelet domain is known to be only approximate [6][7]. However, it is still possible to build a quality metric based on wavelet filters which leads to good performance as explained in [1]. This subband decomposition is based on a spatial frequency dependent wavelet transform approximating the Perceptual Subband Decomposition (PSD) characterized

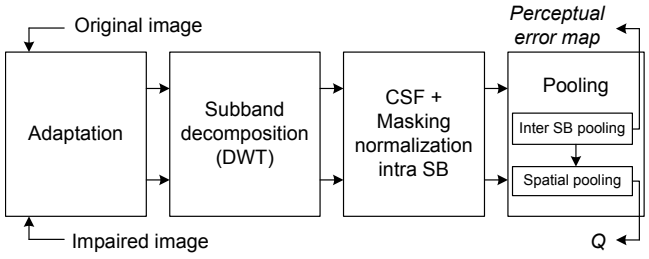


Fig. 1. Structure of the wavelet based quality assessment method (WQA)

in previous works [8], and defined by analytic filters. The Discrete Wavelet Transform (DWT) used is the CDF 9/7 (Cohen-Daubechies-Feauveau). The number of decomposition levels L is chosen so that the low frequency (LF) DWT subband matches to the LF subband of the PSD.

2.2. Contrast sensitivity function

The CSF describes the variations in visual sensitivity as a function of spatial frequency and orientation. As complete frequency representation of the images is not available, the CSF is applied over the DWT subband. The wavelet coefficients $c_{l,o}(m,n)$ are normalized by the CSF using one value by DWT subband:

$$\tilde{c}_{l,o}(m,n) = c_{l,o}(m,n) \cdot N_{l,o}^{CSF} \,, \tag{1}$$

For each subband a CSF value $N_{l,o}^{CSF}$ is calculated from the 2D CSF defined by Daly [2]. This value is the average of the 2D CSF values over the covered frequency range for each subband.

2.3. Masking functions

Masking is a rather well known effect that refers to the changes of visibility increase (pedestal effect) or decrease (masking effect) of a signal due to the presence of background (masking signal). The visual masking effects concern here both contrast masking and semi-local masking. The former is used to take into account the modification of the visibility threshold due to the contrast value, whereas the latter allows to consider the modification of the visibility threshold due to the neighborhood characteristics. Four masking functions were tested. The first two are adaptations of Daly's masking model [2] using or not the neighborhood characteristics. The last two are adaptations of Nadenau's masking [9] model using or not the neighborhood characteristics.

2.3.1. Contrast masking by using Daly's model (Daly)

As proposed by Daly[2], the visibility threshold elevation $T_{l,o}(m,n)$ at site (m,n) in the subband (l,o), where l is the level and o is the orientation, is given by:

$$T_{l,o}(m,n) = \left(1 + \left(k_1 \cdot \left(k_2 \cdot |\tilde{c}_{l,o}(m,n)|\right)^s\right)^b\right)^{\frac{1}{b}} \,, \tag{2}$$

where $\tilde{c}_{l,o}(m,n)$ is the CSF-normalized wavelet coefficient at site (m,n), k_1 and k_2 determine the pivot point of the curve, and the parameter b determines how closely the curve follow the asymptote in the transition region. In the initial work of Daly, a value for the learning slope is chosen depending on the subband (cortex subband). Ideally, this value should depend on the uncertainty of the signal masking. One way to deal with the semi-local masking is to locally adapt the slope s in function of the neighborhood activities.

2.3.2. Semi-Local masking by modifying Daly's model (Daly sLM)

In [2] Daly has noted that the parameter s corresponds to the slope of the high masking contrast asymptote, which ranges between 0.65 and 1.0. For a high uncertainty (low learning level), the slope is 1.0 and as the learning increases, the slope (and uncertainty) reduces to 0.65. The visibility threshold elevation $T_{l,o}(m,n)$ at site (m,n) in the subband (l,o), where l is the level and o is the orientation, is given by:

$$T_{l,o}(m,n) = \left(1 + \left(k_1 \cdot \left(k_2 \cdot |\tilde{c}_{l,o}(m,n)|\right)^{s(m,n)}\right)^b\right)^{\frac{1}{b}} \,, \tag{3}$$

where the parameters are the same as in the Equation (2), except for parameter $s(m,n)$ which depends on the neighborhood according to:

$$s(m,n) = S + \Delta s(m,n) \in [0.65; 1] \,, \tag{4}$$

where $\Delta s(m,n)$ is the semi-local complexity parameter. The semi-local activity values of a n-by-n neighborhood are computed on the achromatic component for both the reference and the impaired image. The semi-local activity value $E(m,n)$ is evaluated through the entropy on a n-by-n neighborhood. Then, the entropy values $E(m,n)$ are mapped to the values $\Delta s(m,n)$ through a sigmoid function.

2.3.3. Nadenau : Intra-Channel Model (Nadenau)

In his work [9], Nadenau proposed a simple intra-channel (IaC) contrast model applied on the wavelet coefficients. The non-linearity of the threshold elevation function is approximated by two piece-wise linear functions:

$$T_{l,o}(m,n) = \max(1, \tilde{c}_{l,o}(m,n)^{\varepsilon}) \,, \tag{5}$$

where ε is the slope-parameter.

2.3.4. Nadenau : Intra-Channel Model with semi-Local Masking (Nadenau sLM)

In his work [9], Nadenau also proposed an intra-channel contrast model applied on the wavelet coefficients and using the semi-local activity. This model is inspired from the so called extended masking [10] in the framework of J2K. Basically, it considered the point-wise contrast masking as captured by the IaC-model, but applies additionally an inhibitory term that takes the neighborhood activity into account:

$$T_{l,o}(m,n) = \max(1, \tilde{c}_{l,o}(m,n)^{\varepsilon}) \cdot (1 + \omega_{\Gamma}) \,. \tag{6}$$

where ω_{Γ} is the correction term for the influence of an active or homogeneous neighborhood. ω_{Γ} is the normalized sum of the neighboring coefficients that were taken to the power of ϑ:

$$\omega_{\Gamma} = \frac{1}{(k_L)^{\vartheta} N_{\Gamma}} \sum_{\Gamma} |\tilde{c}_{l,o}|^{\vartheta} \,. \tag{7}$$

The parameter k_L determines the dynamic range of ω_{Γ}, while N_{Γ} specifies the number of coefficient in the neighborhood Γ. Contrary to Nadenau's work, the neighborhood Γ is not chosen causal in this study, but as in the section 2.3.2, a n-by-n neighborhood around site (m,n) is used.

Subjective Experiments	Format	Distortions	#Contents / #Distorted images	Protocol	Viewing Conditions	Display Devices	Population (#)
IVC	512×512	DCT Coding, DWT Coding, Blur	10 / 120	DSIS	ITU-R BT 500.10 6H	CRT	French (20)
OriginalToyama	768×512	DCT Coding, DWT Coding	14 / 168	ACR	ITU-R BT 500.10 4H	CRT	Japanese (16)
NewToyama	768×512	DCT Coding, DWT Coding	14 / 168	ACR	ITU-R BT 500.10 4H	LCD	French (27)

Table 1. Description of the three subjective studies

		IVC (DSIS)			*NewToyama (ACR)*			*OriginalToyama (ACR)*		
		CC	SROCC	RMSE	CC	SROCC	RMSE	CC	SROCC	RMSE
MOS	WQA Daly	0.892	0.896	0.562	0.851	0.855	0.571	0.837	0.844	0.71
	WQA Daly sLM	0.923	0.921	0.48	0.937	0.941	0.38	0.919	0.923	0.514
	WQA Nadenau	0.877	0.876	0.597	0.819	0.818	0.623	0.805	0.806	0.768
	WQA Nadenau sLM	0.918	0.914	0.492	0.876	0.873	0.523	0.861	0.857	0.66
	PSNR	*0.768*	*0.77*	*0.795*	0.699	0.685	0.777	0.685	0.678	0.943
	SSIM	*0.832*	*0.844*	*0.691*	0.823	0.826	0.618	0.814	0.82	0.754
DMOS	WQA Daly	—	—	—	0.874	0.874	0.535	0.85	0.85	0.68
	WQA Daly sLM	—	—	—	0.943	0.942	0.367	0.932	0.93	0.468
	WQA Nadenau	—	—	—	0.84	0.84	0.596	0.81	0.81	0.756
	WQA Nadenau sLM	—	—	—	0.888	0.888	0.508	0.863	0.862	0.652
	PSNR	—	—	—	0.73	0.717	0.752	0.691	0.683	0.931
	SSIM	—	—	—	0.833	0.838	0.61	0.805	0.81	0.766

Table 2. Results on all datasets (MOS and DMOS)

2.4. Error pooling

Prior to this stage, for each subband (l, o), the masking normalization is applied on the error between the CSF normalized wavelet coefficients of the reference image and the impaired image. The goal of this stage is to provide both a distortion map expressed in term of visibility, stemming from the wavelet subbands, and a quality score. The inter subband pooling is divided in three steps (orientation pooling, level pooling and spatial pooling). As the pooling stage is not the focus of this work, the solution chosen is rather simple. It consists in using different Minkowski summations for each pooling steps.

The sequence of the orientation pooling and the level pooling provides a unique perceptual error map, then the spatial pooling is computed resulting in the quality score Q.

3. RESULTS

3.1. Quantitative analysis : MOS/MOSp

The performances of the WQA metric using the four masking functions presented in the previous section are evaluated according to mean observer score (MOS) and quality difference score (DMOS). MOS have been obtained by conducting three subjective quality assessment experiments in normalized conditions (ITU-R BT 500.10). The three subjective experiments are called *IVC*, *OriginalToyama* and *NewToyama*, and are described in Table 1. All observers had normal or corrected to normal vision. All were inexperienced observers (in video processing) and naive to the experiments. Two image databases with various contents, called IVC database and Toyama database, were used in these experiments. The Toyama database comes from the university of Toyama in Japan [11]. The images were displayed on two type of display devices (CRT and LCD). In order to deal with the influence of subjective assessment methodology, these experiments were conducted with two standardized protocols of test, the Absolute Category Rating (ACR) and the Double Stimulus Impairment Scale (DSIS). The main difference between these protocols is that the reference is hidden in ACR and clear in DSIS. Moreover, the impact of cultural factors is explored thanks to the two populations (Japanese and French) tested in these experiments.

Prior to evaluate the objective image quality measures, a psychometric function $f(Q)$ is used to transform the objective quality score Q in predicted MOS (MOSp) or in predicted DMOS (DMOSp), as recommended by the Video Quality Expert Group [12](VQEG). The objective quality metrics are evaluated using three performance metrics recommended by VQEG. The three performance metrics are the linear correlation coefficient (CC), the Spearman rank order correlation coefficient (SROCC), and the root-mean-square-error (RMSE).

Results, presented in Table 2, are reported for the different methods and for the three experiments. For information and to allow readers to make their own opinions on the image dataset, PSNR and SSIM [5] are also evaluated for the three experiments.

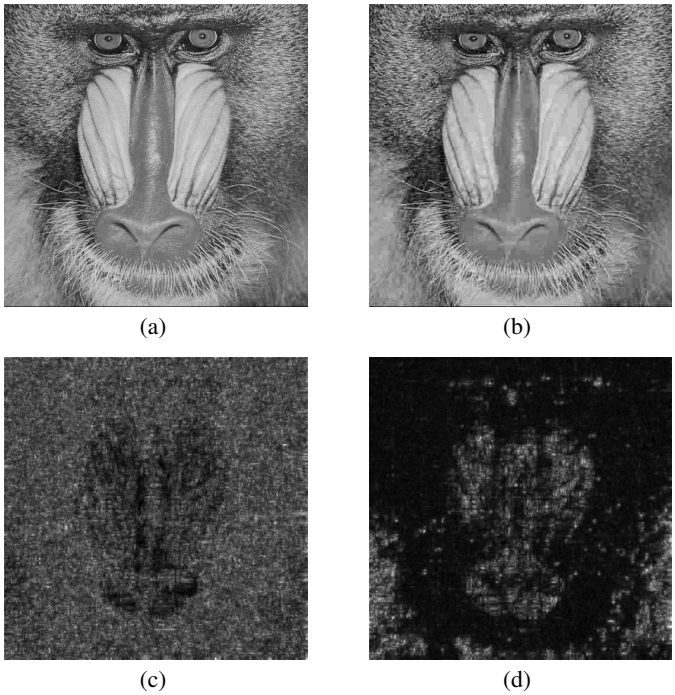
The four multi-channel models outperform PNSR in terms of CC, SROCC and RMSE. SSIM is outperformed by all multi-channel models in terms of CC, SROCC and RMSE for almost all datasets. The exception is WQA Nadenau without sLM concerning MOS on the *NewToyama* and *OriginalToyama* datasets, where there are no significant difference. ΔCC between the multi-channel models and the SSIM goes from -0.009 to $+0.127$. It is not surprising since SSIM do not simulated the multi-channel structure of the HVS.

The use of the semi-local masking in the two configurations (WQA Daly vs WQA Daly sLM, and WQA Nadenau vs WQA Nadenau sLM) consistently increases the performance of the model

in terms of CC, SROCC and RMSE. This observation is done with MOS and with DMOS on the three datasets. On the *IVC* dataset ΔCC between with and without sLM are respectively $+0.031$ and $+0.041$ with WQA Daly and WQA Nadenau. On the *NewToyama* dataset ΔCC between with and without sLM concerning MOS are respectively $+0.086$ and $+0.057$ with WQA Daly and WQA Nadenau, and ΔCC between with and without sLM concerning DMOS are respectively $+0.069$ and $+0.048$ with WQA Daly and WQA Nadenau. On the *OriginalToyama* dataset ΔCC between with and without sLM concerning MOS are respectively $+0.082$ and $+0.056$ with WQA Daly and WQA Nadenau, and ΔCC between with and without sLM concerning DMOS are respectively $+0.082$ and $+0.053$ with WQA Daly and WQA Nadenau. The same trend is observed in terms of SROCC and RMSE. These observations show the positive impact of the semi-local masking, and prove that the masking effect must not be limited to contrast masking.

3.2. Qualitative analysis (semi-Local Masking)

Figure 2(a,b) represents the original image Mandrill and a JPEG compressed version of Mandrill respectively. The difference between the perceptual error map of the WQA Daly model (cf. Figure 2(c)), and the perceptual error map of WQA Daly sLM (cf. Figure 2(d)) is significant. The masking effect in the most active areas like the beard areas, is underestimated with the WQA Daly model, but it is closer to the reality with WQA Daly sLM model.



(a) (b) (c) (d)

Fig. 2. (a) is Mandrill, (b) is Mandrill with JPEG compression, (c) and (d) are WQA perceptual error maps with Daly masking and with Daly sLM masking respectively

4. CONCLUSION

The positive impact of the semi-local masking on some images is important and complementary to contrast masking. Integration of this type of masking in quality metrics improves both the prediction performance of the metrics, and the relevance of their perceptual error maps. The same observations have been done on three subjective studies. Is is interesting to note that the results are independent of the subjective assessment methodology, and cultural factors.

This results lead to the conclusion that semi-local masking must be incorporate in image quality metrics. Another conclusion is that doing a subband decomposition using DWT to simulate the multi-channel structure of the HVS leads to good prediction performance. A spatial transform such as DWT can be considered as a good alternative to reduce computation effort.

Future work includes further investigation to find more revealing measures of the surround influences on masking effect. Moreover, other masking models exist in literature, as [4], and have to be tested.

5. REFERENCES

[1] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "On the performance of human visual system based image quality assessment metric using wavelet domain," *Proc. SPIE Human Vision and Electronic Imaging XIII.*, 2008.

[2] S. Daly, "The visible differences predictor : an algorithm for the assessment of image fidelity," *Proc. SPIE*, vol. 1666, pp. 2–15, 1992.

[3] A. B. Watson, R. Borthwick, and M. Taylor, "Image quality and entropy masking," in *Human Vision, Visual Processing, and Digital Display VIII*, San Jose, CA, USA, 1997, vol. 3016.

[4] M. D. Gaubatz, D. M. Chandler, and S. S. Hemami, "Spatial quantization via local texture masking," 2005, vol. 5666, pp. 95–106.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, pp. 600–612, 2004.

[6] W. Zeng, S. Daly, and S. Lei, "An overview of the visual optimization tools in JPEG2000," *Signal Processing: Image Communication*, vol. 17, no. 1, pp. 85–104, 2002.

[7] A. P. Bradley, "A wavelet visible difference predictor," *IEEE Transactions on Image Processing*, vol. 8, no. 5, pp. 717–730, 1999.

[8] H. Senane, A. Saadane, and D. Barba, "The computation of visual bandwidths and their impact in image decomposition and coding," *International Conference and signal Processing Applications and Technology*, pp. 766–770, 1993.

[9] M. Nadenau, *Integration of Human Color Vision Models into High Quality Image Compression*, Ph.D. thesis, École Polytechnique Fédérale de Lausanne, 2000.

[10] S. Daly, W. Zeng, and S. Lei, "Visual masking in wavelet compression for JPEG2000," *Proc. SPIE Image and Video Communications and Processing*, vol. 3974, 2000.

[11] Z. M. Parvez Sazzad, Y. Kawayoke, and Y. Horita, "Spatial features based no reference image quality assessment for jpeg2000," *International Conference on Image Processing*, vol. 3, pp. 517–520, 2007.

[12] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," March 2000, http://www.vqeg.org/.