IMAGE QUALITY ASSESSMENT WITH MEAN SQUARED ERROR IN A LOG BASED PERCEPTUAL RESPONSE DOMAIN

Wufeng Xue, Xuanqin Mou

Institute of Image Processing and Pattern Recognition
Xi'an Jiaotong University, Xi'an, 710049, China
xqmou@mail.xjtu.edu.cn

ABSTRACT

Up to now, there existing a lot of models that predict subjective quality of the contents of natural images which have undergone some unknown distortion procedures. These models, no matter fall in to the bottom-up mechanism or belong to the top-down functional modelling, fail to provide an easy-applied and reliable solution. The complex computation procedure prevents them from being widely used in related image processing areas such as image enhancement, image reconstruction and video coding.

In the present work, we start from a two stage nonlinear perception model, which transforms the input image into a decorrelated one and then further reduces the redundancy between neighboring pixels by another nonlinear normalization procedure which transforms the previous output into a perceptual response domain. The final quality prediction is computed as the Euclid distance of the reference image and the distorted one in this response domain, this will make the new model be readily applied in other areas.

Index Terms— image quality, response domain, correlation, mutual information

1. INTRODUCTION

Quality assessment of image content aims to provide each image an objective score that indicating the subjective satisfactory of an observer with that image. When this subjective feeling is conducted by comparing the test image with a reference pristine image, the quality assessment is called full reference image quality assessment (FRIQA). Other conditions when reference image is not totally assessable is beyond the scope of this work. In the scenario of FRIQA, the traditional Euclid distance based metrics, i.e. MSE or PNSR, although quite convenient to utilize, show quite inconsistent with the subjective experience [1, 2]. This is why perceptual relevant FRIQA model is expected.

Among the existing literatures, there are mainly two mechanisms [3]: bottom-up ones, which process the input signal with sophisticated sub-models that try to behave like the human visual system (HVS) does in each procedure of the visual pathway; and the top-down ones, which only models the assumed functionality of the whole visual procedure. Due to the limited knowledge about HVS, the former ones often get inferior performance with much demanding computation. It's not until recently that FRIQA models following this way show state-of-art models but with cost of high complexity [4, 5]. On the contrast, the later ones only model the assumed properties of the HVS, and bring us a promising solution [6] for FRIQA and lead to more improved models along this line [7, 8, 9]. These models are all of high consistence with subjective experience, but they still failed to be easily applied. Other models that based on the information theory in the wavelet domain also achieve good results [13, 14].

The application of a FRIQA model is not limited to just a quality indicator. There are fields like image restoration, image super-resolution, and perceptual video coding, resources allocation in the network, where the FRIQA model is expected to be some optimization function that we can easily find the algorithm. This is actually the reason why after development for decades, MSE still is the first choice in many algorithm design than FRIQA models. Unfortunately, none of the existing models has the potential to be applied in such case due to their intrinsic mathematic properties.

In our previous work [10, 11, 12], the image quality is computed based on the receptive fields related laplacian of Gaussian (LOG) signals. NSER [10] and rNSE [11] measure the quality through the ratio of the edge point that preserved in the distorted image in a single optimal scale and in multi-scale, respectively. The underlying assumption is that quality assessment is conducted by the earliest stage of HVS. Due to the highly nonlinear procedure of the edge detection, the edge based models cannot be conveniently applied to other perceptual related problems. The LOG-COR model in [12] predicts the image quality by the correlation between two counterpart LOG signals. Although the correlation computation has been employed in other models, comparing to the Euclid distance, it is still less flexible in practice.

We will show in this paper that, after a normalization procedure, which transform the LOG signal into the

(a) Image in the spatial domain (b) Image after linear transform (c) Image in the response domain

Figure 1. Example of a natural image in the spatial domain (a), the LOG transformed domain (b) and the nonlinear normalization domain (c).

perceptual response domain, more dependency of neighboring location is reduced and the correlation of the perceptual domain signals become more predictive of image quality. Even more surprising result is that in the perceptual response domain, the Euclid distance is sufficient to predict the image subjective quality well. This will pave the way of IQA models' application in other areas that mentioned above.

The rest of the paper is organized as follows: Section 2 introduces the normalized procedure and propose the IQA models based on this normalized signal. Section 3 shows the IQA performance of the proposed models in three publicly available IQA databases and conducts comparison with other existing state-of-art models. Then Section 4 concludes the paper.

2. PERCEPTUAL RESPONSES AND IQA MODELS

To build an IQA model that are reasonable and reliable, a perceptual model of HVS is prerequisite. The perceptual model we utilized in this work is originally proposed by Teo and Heeger [1] by fitting the empirical measurement of the response properties of neurons in the primary visual cortex (V1 area) and the psychophysics of spatial pattern detection. This model is comprised of two basic steps [1, 15]:

$$I \xrightarrow{H} W \xrightarrow{R} r \quad (1)$$

In the first step, the input image $\boldsymbol{I}$ is transformed into a new local frequency domain with the filter bank $\boldsymbol{H}$. Examples for the filter include discrete cosine transform (DCT), principal component analysis (PCA), Fourier transform (FT).

Unfortunately, these linear filters only decorrelate the input images. There still exists much redundancy and interaction between the transform domain coefficients, which will prevent the linear representation from being appropriate to employ the Euclid distance as the final distortion measurement. Therefore, a further nonlinear transform $\boldsymbol{R}$ is required the transform the coefficients $\boldsymbol{W}$ into the new nonlinear perceptual domain $\boldsymbol{r}$. During the nonlinear transform $\boldsymbol{R}$, the coefficient in the linear domain is normalized by the root of a weighted energy summation. After this nonlinear normalization, the dependency of neighboring coefficients is much reduced.

2.1 The linear transformation

After the long evolution, human visual system is believed to have been shaped according to the statistical properties of the surrounding environments [16]. The functionality of HVS is to efficiently represent these external information and serves to more complex high level visual tasks. In the earliest stage of visual pathway, the input visual signal is first received by the rods and cones that densely distributed in the retina and then forward to the ganglion cells and neurons in lateral geniculate nucleus (LGN), and further to the visual cortex area [17]. The receptive fields, which means the responsible area in the retina of the neuron cells in the visual pathway, of the ganglion cells and the LGN have been shown to resemble the Laplacian of Gaussian (LOG) filter [18]. Besides, the LoG signal is considered to carry crucial structural information of the image on the position of its zero-crossing according to Marr's theory. The LOG filter is as follows:

$$h(x, y, \sigma_1) = \frac{1}{2\pi\sigma_1^2} \frac{x^2 + y^2 - 2\sigma_2^2}{\sigma_1^4} \exp\left(-\frac{x^2 + y^2}{2\sigma_1^2} \right) \quad (2)$$

Then the linear representation with LOG filter is computed as by convolution between the input image and the filter $\boldsymbol{h}$.

$$W = H(I) = I \otimes h \quad (3)$$

In the original perceptual model, the linear transform is applied by a bank of frequency and orientation selected filters, such as the steerable pyramid decomposition. We assume that the perception of HVS to the distortion structure in the image is not sensitive to the orientation of the structure. Therefore, we prefer the physiology related LOG filter and build an isotropy perceptual model for image quality assessment. Besides, we applied the LOG filter in a single scale for the reason that the test images in the database are all of the similar size, in which case a single optimal scale will be sufficient.

2.2 Normalization model

To further reduce the redundancy between the neighboring coefficients $\boldsymbol{W}$, a nonlinear divisive normalization $\boldsymbol{R}$ is

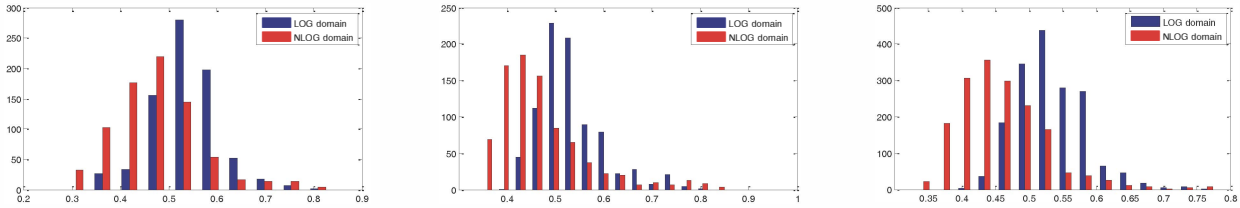

Figure 2. The mutual information between neighboring coefficients in the LOG domain and the NLOG domain for LIVE database (left), CSIQ database (middle), and TID2008 database (right).

applied to each coefficient in $\boldsymbol{W}$. The normalization factor is computed as the root of the weighed summation of the energy of neighboring coefficients. The nonlinear transform $\boldsymbol{R}$ matches well the making effect and the limit dynamic range of HVS. After adapted to the LOG filter that used in this work, the transform $\boldsymbol{R}$ can be computed as:

$$r = R(W) = \frac{W}{\sqrt{W^2 \otimes g + c_1}}, \tag{4}$$

where g is a Gaussian kernel that weight the neighboring coefficients and c_1 is a constant to avoid dividing by zero. The constant also accounts for the neuron noise in HVS.

$$g(x, y, \sigma_2) = \frac{1}{2\pi\sigma_2^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_2^2}\right). \tag{5}$$

In Figure 1 we show examples of the LOG response $\boldsymbol{W}$ and the normalized LOG (NLOG) response $\boldsymbol{r}$. It is clear that although the LOG response can extract the image structure well, it fails to equalize the dynamic range of the structure contrast across all the spatial locations (the steps and leaves are with lower contrast than the people and other objects on the stage). On the contrary, in the NLOG response, the contrast of the structures in difference area have been equalized. The mutual information existing between neighboring locations is also reduced, which is more clearly demonstrated in Figure 2. In all of the three databases, we computed the mutual information between neighboring pixels (average of the horizontal and vertical directions) in the LOG domain and NLOG domain. The normalization procedure further reduces the mutual information between neighboring pixels.

2.3 The FRIQA models

Based on the above perceptual models, we proposed two FRQIA models NLOG-MSE and NLOG-COR. Denote the reference image and the distorted image as C and D, respectively, the two models are computed as follows:

$$\textbf{NLOG-MSE}(C, D) = \frac{1}{n}\sum_{i=1}^{n}\left(r_C(i) - r_D(i)\right)^2 \tag{6}$$

$$\textbf{NLOG-COR}(C, D) = \frac{1}{n}\sum_{i=1}^{n}\frac{2r_C(i)r_D(i) + c_2^2}{r_C^2(i) + r_D^2(i) + c_2^2} \tag{7}$$

where i is the pixel index, n is the total number of pixels in the image and c_2 is a constant to avoid dividing by zero.

3. EXPERIMENT AND ANALYSIS

3.1 Databases and evaluation protocols

We can evaluate how closely the predicted image quality scores by a FRIQA model correlate with human judgments, which are usually recorded as (Difference) Mean Opinion Scores (DMOS/MOS). Several subjective image quality evaluation databases have been established in the IQA community. Here we use the three largest and mostly widely used ones: the LIVE database [19], the CSIQ database [4] and the TID2008 database [20].

The LIVE database consists of 779 distorted images, generated from 29 original images by processing them with 5 types of distortions at various levels. These distortions reflect a broad range of image impairments, such as edge smoothing, block artifacts, image-dependent distortions, and additive random noise. The CSIQ database consists of 30 original images and their distorted counterparts with six types of distortions at five different distortion levels each. The TID2008 database is composed of 25 reference images and their distorted counterparts with 17 types of distortions at four levels each.

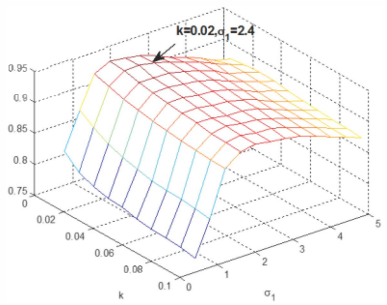

Figure 3. Performance variation of NLOG-MSE in a subset of LIVE database.

To evaluate the performance of a FRIQA model, three scores that measure the consistency between the results of a BIQA model and the subjective DMOS/MOS scores are generally used: the Spearman rank order correlation coefficient (SRC), the Pearson correlation coefficient (PCC),

Table 1. Performance comparison of the proposed IQA models and other competitors.

	LIVE(779)			CSIQ(866)			TID2008(1700)			Weighted average	
	SRC	PCC	RMSE	SRC	PCC	RMSE	SRC	PCC	RMSE	SRC	PCC
PSNR	0.876	0.872	13.36	0.806	0.751	0.173	0.553	0.523	1.144	0.694	0.664
IFC[13]	0.926	0.927	10.26	0.767	0.837	0.144	0.568	0.203	1.314	0.703	0.537
NSER[10]	0.942	0.939	9.362	**0.934**	**0.947**	**0.084**	0.740	0.796	0.813	0.837	0.868
SSIM[6]	0.948	0.945	8.95	0.876	0.861	0.133	0.775	0.773	0.851	0.841	0.836
VIF[14]	**0.964**	**0.960**	**7.61**	0.919	0.928	0.098	0.749	0.808	0.790	0.844	0.875
MAD[4]	**0.967**	**0.968**	**6.907**	**0.947**	**0.950**	**0.082**	0.834	0.829	0.751	0.894	0.893
IW-SSIM [9]	0.957	0.952	8.35	0.921	0.914	0.106	**0.856**	**0.858**	**0.689**	**0.896**	**0.895**
FSIM[7]	**0.963**	**0.960**	**7.67**	0.924	0.912	0.108	**0.880**	**0.874**	**0.653**	**0.911**	**0.904**
LoG-MSE	0.927	0.913	11.136	0.876	0.646	0.200	0.757	0.737	0.907	0.827	0.754
LoG-COR	0.934	0.928	10.214	0.922	0.916	0.105	0.835	0.826	0.756	0.881	0.873
NLOG-MSE	0.947	0.945	8.952	**0.945**	**0.947**	**0.085**	**0.851**	**0.836**	**0.736**	**0.898**	**0.890**
NLOG-COR	0.941	0.937	9.536	0.931	0.933	0.094	0.838	0.827	0.755	0.886	0.880

and the root mean squared error (RMSE). Both SRC and PCC lie in the range [-1, 1]. A good BIQA model should give a correlation with the subjective DMOS/MOS scores as close to 1 (or -1) as possible. When calculating these indices, a nonlinear logistic regression [21] is built between the predicted scores and the subjective scores.

3.2 Parameter settings

During the experiment, all the computations are applied to the luminance channel. There are totally three parameters in the nonlinear perception model: σ_1, σ_2, c_1. In our experiment, we set $\sigma_2=2\sigma_1$, and $c_1=(255\times k)^2$. Note that 255 is the range of the luminance value of the input image. Then σ_1 and c_1 is fine-tuned with a subset (20%) of the distorted images in LIVE database. Figure 3 shows the SRC performance of NLOG-MSE in this subset. The optimal values occur at k=0.02 and σ_1=2.4. We can draw from Figure 3 that the performance is not sensitive to k. For the parameter c_2 in NLOG-COR, we keep k and σ_1 fixed, and fine-tune it with the same subset. The optimal value for c_2 is 0.72.

3.3 Performance analysis and comparison

We evaluate the performance of the proposed models in the three public database. As comparison, we list the results of other state-of-art models as well as the LOG based models [12] without the nonlinear normalization. The results are demonstrated in Table 1. In the last columns show the average results across the three database weighted by the image number of each database. In each column, the top 3 values are highlighted in boldface. With respect to the average SRC performance, NLOG-MSE outperforms other popular models, including the information based VIF and IFC [13, 14], the structure similarity based SSIM [6], IWSSIM [9], the well-designed bottom-up model MAD [4]. NLOG-MSE only fails to beat FSIM [7], which is based on the computation demanding phase congruency feature.

Compared to the two models LOG-MSE and LOG-COR, which are based on the immediately output of the LOG filter, the proposed ones have obvious advantage. Another interesting results should be noticed that without the normalization procedure, LOG-MSE performs poorly while the LOG-COR performs much better. As we have investigated, there is still much dependency in the LOG signals and it's not appropriate to apply the MSE computation to this kind of signals. Indeed, a simple reformulation of the correlation computation will reveal that fact that correlation computation is a point-wise energy normalized MSE computation, where the energy is the summation of the squared LOG coefficients of the reference image and the distorted image. This computation can reduce the effect of image content to some extent and thus improve the performance and is widely applied in the IQA community. However, after the nonlinear procedure is conducted, the benefits that the correlation computation bring is not so much, and actually it results in worse performance. This reason lies in the fact that the NLOG signal is already much redundancy-reduced compare to the LOG signal and the MSE is not applicable to the signal, the point-wise normalization is no longer necessary.

4. CONCLUSION

In this work, we propose a perceptual model for the task of FRIQA, which is comprised of two basic steps: the linear LOG filtering and the nonlinear normalization. This new model largely reduced the mutual dependency between neighboring coefficients and is efficient for quality assessment. Based on this representation, we show that both the MSE computation and correlation give excellent performance. Besides, the simple MSE computation between the NLOG makes IQA model easily-computed and has a great potential to be widely used in other real-time applications and perceptual based optimization problems.

5. ACKNOWLEDGEMENTS

This work is partly supported by Natural Science Foundation of China (No.90920003 and No. 61172163).

6. REFERENCES

[1] Patrick C. Teo, and David J. Heeger. "Perceptual image distortion." *IS&T/SPIE 1994 International Symposium on Electronic Imaging*, 1994.

[2] Z. Wang and A. C. Bovik, "Mean Squared Error: Love It or Leave It?" *IEEE Signal Processing Magazine*, no. January, pp. 98–117, 2009.

[3] Z. Wang and A. C. Bovik, Modern image quality assessment, vol. 2, no. 1. *Morgan & Claypool Publishers,* 2006, pp. 1–156.

[4] Eric C. Larson and Damon M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy", *J. Electron. Imaging* 19, 011006, Jan 07, (2010).

[5] D. M. Chandler and S. S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–98, Sep. 2007.

[6] Z. Wang, A C. Bovik and H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, pp. 600-612, 2004.

[7] Lin Zhang, Lei Zhang, X. Mou and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," *IEEE Trans. Image Process*. 20(8):2378-2386 (2011).

[8] W. Xue, L. Zhang, X. Mou and A. Bovik, "Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index," *IEEE Transactions on Image Processing*, vol.23, no.2, pp.684-695, Feb. 2014.

[9] Z. Wang, and Q. Li. "Information content weighting for perceptual image quality assessment." *IEEE Transactions on Image Processing*, 20.5, 1185-1198 (2011).

[10] M. Zhang, X. Mou, and L. Zhang. "Non-Shift Edge Based Ratio (NSER): An Image Quality Assessment Metric Based on Early Vision Features." *Signal Processing Letters, IEEE* 18.5, 315-318 (2011).

[11] W. Xue, and X. Mou. "An image quality assessment metric based on Non-shift Edge." *IEEE International Conference on Image Processing (ICIP)*, 2011.

[12] W. Xue, X. Mou, C. Chen and Lei Zhang, "LoG acts as a good feature in the task of image quality assessment," *SPIE Electronic Imaging*, 2014.

[13] H.R. Sheikh, A.C. Bovik and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol.14, no.12pp. 2117- 2128 (2005).

[14] H.R. Sheikh. and A.C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing,* vol.15, no.2,pp. 430- 444 (2006).

[15] Andrew B. Watson, and Joshua A. Solomon. "Model of visual contrast gain control and pattern masking." *JOSA A* 14, no. 9 (1997): 2379-2391.

[16] J. J. Atick and A. N. Redlich, "What does the retina know about natural scenes?" *Neural Computation*, vol. 4, no. 2, pp. 196–210 (1992).

[17] G. S. Brindley. "Physiology of the retina and the visual pathway." *London: Edward Arnold*, (1960).

[18] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman, "Receptive field dynamics in the central visual pathways", *Trends Neurosci*.,vol. 18,pp. 451-458 (1995).

[19] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. (2005) "Live Image Quality Assessment Database Release 2." [Online]. http://live.ece.utexas.edu/1research/quality.

[20] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, "TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics", *Advances of Modern Radio electronics*, Vol. 10, pp. 30-45, (2009).

[21] VQEG, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment – Phase II," August 2003, available at http://www.vqeg.org/.