ON BETWEEN-COEFFICIENT CONTRAST MASKING OF DCT BASIS FUNCTIONS

Nikolay Ponomarenko (*), Flavia Silvestri(**), Karen Egiazarian (***), Marco Carli (**), Jaakko Astola (***) and Vladimir Lukin (*)

(*) National Aerospace University, Kharkov, Ukraine
(**) University of Rome "Roma TRE", Rome, Italy
(***) Tampere University of Technology, Tampere, Finland

ABSTRACT

In this paper we propose a simple and effective model of visual between-coefficient contrast masking of DCT basis functions based on a human visual system (HVS). The model operates with the values of DCT coefficients of 8x8 pixel block of an image. For each DCT coefficient of the block the model allows to calculate its maximal distortion that is not visible due to the between-coefficient masking. A modification of the PSNR is also described in this paper. The proposed metric, PSNR-HVS-M, takes into account the proposed model and the contrast sensitivity function (CSF). For efficiency analysis of the proposed model, a set of 18 test images with different effects of noise masking has been used. During experiments, 155 observers have sorted this set of test images in the order of their visual appearance comparing them to undistorted original. The new metric, PSNR-HVS-M has outperformed other well-known reference based quality metrics and demonstrated high correlation with the results of subjective experiments (Spearman correlation is 0.984, Kendall correlation is 0.948).

1. INTRODUCTION

In the last decades, the scientific community made a great effort to develop image and video quality assessment methods incorporating perceptual measures. Many of the quality metrics proposed were based on properties of HVS, such as CSF and luminance masking [1, 2]. Some of subjective quality models are focused on subband decomposition separating the visual stimulus into different spatial and temporal bands. Initially the Discrete Cosine Transform (DCT) has often been utilized in contrast masking due to its suitability for certain applications and accuracy in modeling the cortical neurons [3]. In the DCT domain there are different approaches to model the contrast sensitivity masking in order to compute a visually optimal quantization matrix for a given image [4]. Although in literature the mutual interrelation between DCT coefficients has been analyzed [5], it does not always match with subjective quality assessment. In this paper, we propose an efficient and simple model that does not require any additional data except an image itself.

Section 2 contains description of the proposed model. In Section 3 it is shown how this model can be taken into the derivation of PSNR. The used set of test images and noise parameters are described in Section 4. Section 5 describes the experiment intended on verification of the proposed model efficiency. Experimental results and data analysis are given in Section 6.

2. DESCRIPTION OF PROPOSED MODEL

As stated in [5], each DCT coefficient X_{ij} of an image block in some degree masks any other block coefficients, except DC coefficient with the index 0,0 that corresponds to the block mean luminance. In our model, we rely on assumption that masking degree of each coefficient X_{ij} depends upon its square value (power) and human eye sensitivity to this DCT basis function determined by means of CSF. Several basis functions can jointly mask one or few other basis functions. Then, their masking effect value depends upon a sum of their weighted powers.

Let us denote a weighted energy of DCT coefficients of an image block 8x8 as $E_w(X)$:

$$E_w(X) = \sum_{i=0}^{7}\sum_{j=0}^{7} X_{ij}^{2} C_{ij} \,, \tag{1}$$

where X_{ij} is a DCT coefficient with indices i,j, C_{ij} is a correcting factor determined by the CSF.

The DCT coefficients X and Y are visually undistinguished if $E_w(X\text{-}Y) < \max(E_w(X)/16, E_w(Y)/16)$, where $E_w(X)/16$ is a masking effect E_m of DCT coefficients X (normalizing factor 16 has been selected experimentally).

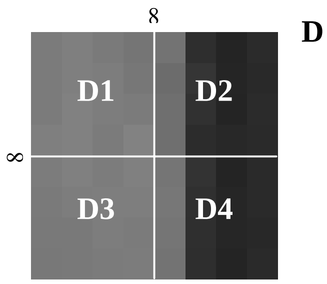


Fig.1. Block D of an image contains an edge

The value of masking effect (1) can be too high if an image block belongs to an edge (see Fig. 1). In such a case we propose to reduce a masking effect for a block **D** proportionally to the local variances V(.) in blocks D1, D2, D3, D4 in comparison to the entire block:

$$E_m(D) = E_w(D)\delta(D)/16, \qquad (2)$$

where $\delta(D) = (V(D1)+V(D2)+V(D3)+V(D4))/4V(D)$, V(D) is the variance of the pixel values in block D.

Table 1 presents the calculated values of C_{ij}. While obtaining them, we have used the quantization table for the color component Y of JPEG [6] that has been also obtained on the basis of CSF. Note that the values in quantization table JPEG have been normalized and then squared.

Table 1. Values of C_{ij}

i\j	0	1	2	3	4	5	6	7
0	0	0.8264	1.0000	0.3906	0.1736	0.0625	0.0384	0.0269
1	0.6944	0.6944	0.5102	0.2770	0.1479	0.0297	0.0278	0.0331
2	0.5102	0.5917	0.3906	0.1736	0.0625	0.0308	0.0210	0.0319
3	0.5102	0.3460	0.2066	0.1189	0.0384	0.0132	0.0156	0.0260
4	0.3086	0.2066	0.0730	0.0319	0.0216	0.0084	0.0094	0.0169
5	0.1736	0.0816	0.0331	0.0244	0.0152	0.0092	0.0078	0.0118
6	0.0416	0.0244	0.0164	0.0132	0.0094	0.0068	0.0069	0.0098
7	0.0193	0.0118	0.0111	0.0104	0.0080	0.0100	0.0094	0.0102

Using this proposed model, it is possible to evaluate a masking effect for each image block. In the next Section we show how this can be used for assessment of image visual quality.

3. MODIFICATION OF PSNR USING A NEW MASKING MODEL

A basis of the proposed metric is a PSNR-HVS [7]. The modified metric, PSNR-HVS-M that takes into account the proposed masking model can be calculated for each image block as it is shown in Fig. 2. Here MSE_H is the MSE taking into account CSF [7].

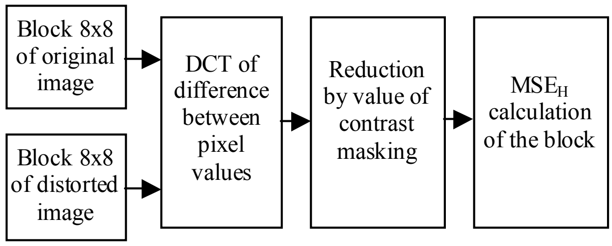


Fig. 2. Flow-chart of PSNR-HVS-M calculation

Reduction by value of contrast masking in accordance to the proposed model is carried out in the following manner. First, the maximal masking effect E_{max} is calculated as $max(E_m(X_e), E_m(X_d))$ where X_e and X_d are the DCT coefficients of a original image block and a distorted image block, respectively. Then, the visible difference between X_e and X_d is determined as

$$X_{\Delta ij} = \begin{cases} X_{eij} - X_{dij}, & i=0, j=0 \\ 0, & \left|X_{eij} - X_{dij}\right| \le E_{norm}/C_{ij} \\ X_{eij} - X_{dij} - E_{norm}/C_{ij}, & X_{eij} - X_{dij} > E_{norm}/C_{ij} \\ X_{eij} - X_{dij} + E_{norm}/C_{ij}, & \text{otherwise} \end{cases}, \qquad (3)$$

where E_{norm} is $\sqrt{E_{max}/64}$. The Matlab source of the proposed PSNR-HVS-M is available from [8]. Similarly to [7], PSNR-HVS-M can be calculated in both non-overlapping and overlapping blocks (in experimental part we used the former one).

4. TEST IMAGE SET

In our experiments we have used a set of test images that contains 19 images available from [8]. Original image synthesized by us is shown on Fig. 3.

Table 2 gives an information on the test image set. We have considered additive i.i.d. Gaussian noise (G) and spatially correlated additive Gaussian noise (SC) with three different intensities. Besides, we have considered three different cases of noise spatial location:

1) Uniformly through entire image (without masking) (U);
2) Mostly in regions possessing a high masking effect (H);
3) Mostly in regions possessing a low masking effect (L).

This Table also contains the values for all analyzed and compared metrics and the result (averaged positions of each test image in the ordered test set) of subjective experiments RSE (see Section 5). DT denotes results for DCTune 2.0 software [11] that realizes metrics proposed in [5].

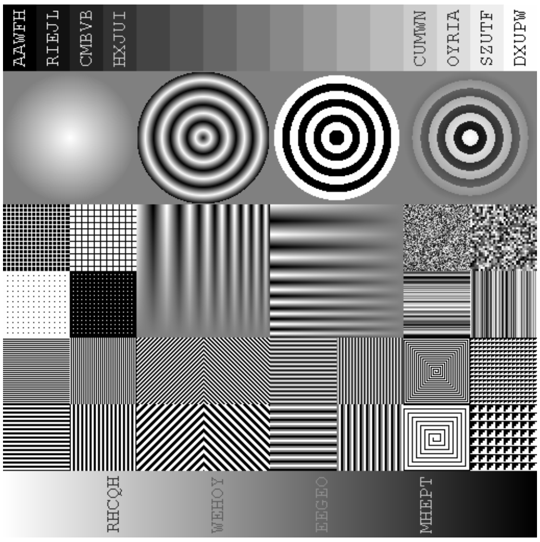


Fig 3. Original test image

Table 2. Obtained characteristics for different metrics

№	Noise	Pre-sence	PSNR	PSNR-HVS [7]	UQI [9]	MSSIM [10]	DT [5] [11]	PSNR HVS-M	RSE
1	G	U	28.60	28.23	0.73	0.80	24.9	33.20	3.6
2	G	H	28.60	27.51	0.78	0.91	13.7	35.47	1.6
3	G	L	28.60	28.55	0.71	0.76	37.2	30.64	6.7
4	SC	U	28.58	23.86	0.73	0.82	36.2	26.68	12.7
5	SC	H	28.58	26.46	0.78	0.93	17.1	32.49	5.3
6	SC	L	28.59	23.79	0.71	0.79	43.5	25.27	13
7	G	U	27.55	27.19	0.72	0.78	29.7	32.07	5.6
8	G	H	27.55	26.63	0.75	0.85	19.0	33.65	2.8
9	G	L	27.51	27.46	0.71	0.74	41.2	29.56	8.4
10	SC	U	27.52	22.81	0.72	0.80	42.5	25.50	14.7
11	SC	H	27.54	24.73	0.75	0.87	24.5	29.31	10.3
12	SC	L	27.56	22.82	0.71	0.77	47.2	24.39	14.9
13	G	U	26.06	25.71	0.70	0.75	34.6	30.40	8.0
14	G	H	26.02	25.26	0.72	0.79	27.2	31.29	5.4
15	G	L	26.04	25.96	0.69	0.72	48.5	28.06	10.0
16	SC	U	26.05	21.34	0.71	0.77	49.0	23.85	16.3
17	SC	H	26.01	22.56	0.72	0.81	35.0	26.09	14.5
18	SC	L	26.01	21.29	0.69	0.75	54.1	22.91	17.1

5. SUBJECTIVE EXPERIMENTS

In carrying out subjective experiments, 155 observers (45 from Finland, 43 from Italy, 67 from Ukraine) have participated. In each experiment, an observer has to choose from a pair of distorted images one which is "closer" to the sample (undistorted) image. In this way, each observer formed a sorted sequence of distorted images in the order of increased visual appearance. Totally, 8192 comparisons of visual appearance of test images have been performed (on average 53, for each observer).

A monitor brightness, illumination and distance from an observer's eyes to the monitor varied in wide limits. The only fixed parameter in our experiments was the monitor resolution, 1152x864 pixels. There were 128 experiments carried out using CRT monitors, and 27 experiments using LCD monitors.

By analyzing the obtained sequences for all observers, averaged orders (ordered sequences) of distorted test images have been obtained. It is important that the cross correlation factors for different groups of observers were very high (see data in Table 3).

Table 3. Cross correlation factors:

Group of observers	Spearman correlation	Kendall correlation
Finland – Italy	0.996	0.895
Finland – Ukraine	0.996	0.935
Italy - Ukraine	0.997	0.961
CRT - LCD	0.998	0.922

These data evidence in favor of high confidence of experimental results and applicability of conclusions to both CRT and LCD monitors.

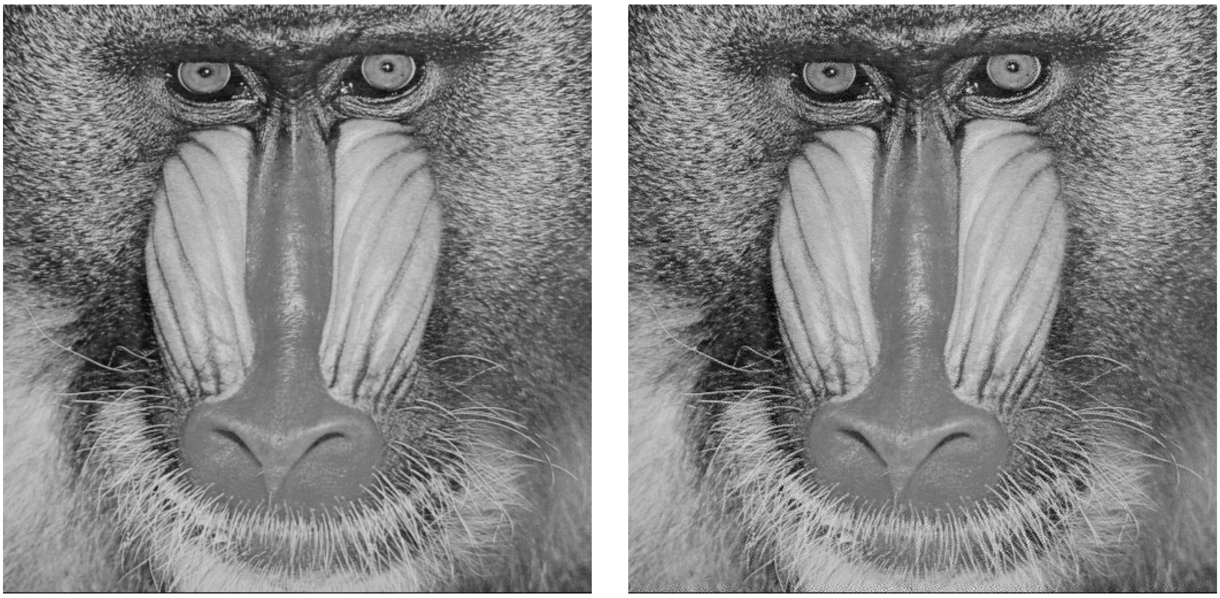
6. EXPERIMENTAL RESULTS

The result of subjective experiments consists in getting the ordered set of distorted test images. For analysis of adequacy of the considered metrics we have used Spearman and Kendall correlations that can be exploited for determination of correlation between sorted data. Kendall correlation usually produces more "careful" results than Spearman correlation (it approaches to unity more slowly). Table 4 presents the correlation data for the considered metrics with data of subjective experiments.

Table 4. Correlations for the considered metrics

Measure	Spearman correlation	Kendall correlation
PSNR	0.537	0.359
PSNR-HVS [7]	0.895	0.712
UQI [9]	0.550	0.438
MSSIM [10]	0.406	0.358
DCTune [5, 11]	0.829	0.712
PSNR-HVS-M	0.984	0.948

The data presented in Table 4 allow concluding the following. First, the popular metrics UQI and MSSIM as well as the standard PSNR do not relate to human perception well. The metrics PSNR-HVS and DCTune suit visual perception considerably better although cross-correlation is not too high (Kendall correlation is equal to 0.712). Finally, the proposed PSNR-HVS-M outperforms all other considered metrics and demonstrates an appropriate correspondence to human perception.



a) b)

Fig 4. Image Baboon (a) and the image with masked noise (b), PSNR=26.18 dB, PSNR-HVS=34.43 dB, PSNR-HVS-M=51.67 dB (both images are available from [8])

An example of a distortion masking in image Baboon in accordance with the proposed model is given in Fig. 4.

7. CONCLUSIONS

In this paper, a simple model of between-coefficient masking of DCT basis functions is proposed and the modifications of PSNR that take into account this model are put forward. The new measure PSNR-HVS-M has shown its higher efficiency (adequacy) in comparison to known metrics.

One more advantage of the new metric is that it is expressed in dB. Therefore, for people who are got used to exploit and analyze standard PSNR, the new metric could be convenient and understandable.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. G. Boato for running subjective experiments at the University of Trento, Italy.

REFERENCES

[1] A.B. Watson and J.A. Solomon, "Model of visual contrast gain control and pattern masking", *J. Opt. Soc. Am. A*, 14(9), pp. 2379–2391, 1997.

[2] Foley, J.M. "Human Luminance Pattern-Vision Mechanisms-Masking Experiments Require a New Model", *Journal of the Optical Society of America a-Optics Image Science and Vision*, 11 (6), 1710-1719, 1994.

[3] W. Zhou, R.S. Hamid, and A.C. Bovik. *The handbook of video databases: Design and Applications*, CRC Press, 2003.

[4] A. B. Watson, "Perceptual optimization of DCT colour quantization matrices", *Image Processing ICIP-94*, 1, pp. 100–104, 1994.

[5] Solomon J. A., Watson A. B., and Ahumada A. "Visibility of DCT basis functions: Effects of contrast masking", *Proceedings of Data Compression Conference*, Snowbird, Utah: IEEE Computer Society Press, 361-370, 1994.

[6] G. Wallace, *The JPEG Still Picture Compression Standard*, Comm. of the ACM, vol. 34, No.4, 1991.

[7] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, M. Carli, New full-reference quality metrics based on HVS, *CD-ROM Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, Scottsdale, USA, 2006, 4 p.

[8] www.cs.tut.fi/~ponom/psnrhvsm.htm, PSNR-HVS-M page.

[9] Z. Wang, A. Bovik, A universal image quality index, *IEEE Signal Processing Letters*, vol. 9, pp. 81–84, March, 2002.

[10] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Transactions on Image Processing*, vol. 13, issue 4, pp. 600-612, April, 2004.

[11] http://vision.arc.nasa.gov/dctune/ DCTune 2.0 page.