

# Modified Image Visual Quality Metrics for Contrast Change and Mean Shift Accounting

Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Marco Carli

**Abstract – Graphical information as color images is widely used in CAD and telecommunication systems. Several factors can contribute to impair the quality of an image. This paper deals with image visual quality assessment using objective metrics. Experimental results show that the two modified quality metrics outperform existing ones for a wide set of possible distortions.**

**Keywords – Visual quality, distorted images, metrics.**

## I. INTRODUCTION

Digital images are widely used for sharing information in many applications as telecommunications, CAD, diagnostic, remote sensing systems [1], etc. Since any imaging system is not perfect, acquired images are often corrupted by noise, blur and other factors [2]. Besides, when images are delivered through communication networks [3], additional distortion can occur due to compression, coding/decoding errors, etc.

Visual quality is of prime importance especially if the final user of the image communication/sharing process is a human being. Because of this, image visual quality assessment has been a topic of intensive research in recent years [4]. Many objective visual quality metrics have been proposed (see [4-9] and references therein). These metrics (also called quality indices) are widely used in image processing applications as image lossy compression, denoising, rendering and others [5-10].

To adequately assess efficiency of filtering for color images acquired by digital cameras the quality metric should be able to adequately incorporate peculiarities of human visual system (HVS). The main problem is that till now HVS has not been completely analyzed and only a limited number features are taken into account in design of modern quality metrics.

Due to this fact, most of existing metrics are heuristic and with limited performances.

Examples of specialized set of distorted images to be used as test in metrics performances are LIVE [9] and TID2008 [8, 11]. Such databases serve several goals.

First, they allow verifying and comparing different metrics [8, 9, 12-14]. Each database contains a set of reference (noise/distortion-free) images and their versions corrupted by noise or other types of distortions. The values of Mean Opinion Scores (MOS) for the distorted images are also

---

Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin – Dept of Transmitters, Receivers and Signal Processing, National Aerospace University, 17, Chkalova Str., Kharkiv, 61070, UKRAINE, E-mails: [nikolay@ponomarenko.info](mailto:nikolay@ponomarenko.info), [ol.ieremeiev@mail.ru](mailto:ol.ieremeiev@mail.ru), [lukin@ai.kharkov.com](mailto:lukin@ai.kharkov.com)

Karen Egiazarian – Institute for Signal Processing, Tampere University of Technology, P.O. Box-553, Tampere, FIN-33101, FINLAND, E-mail [karen@cs.tut.fi](mailto:karen@cs.tut.fi)

Marco Carli – Applied Electronics Dept, University of Roma TRE, via della Vasca Navale, 84, 00146, Rome, Italy, E-mail [carli@uniroma3.it](mailto:carli@uniroma3.it)

reported. This allows establishing links between objective quality metrics and MOS obtained by averaging subjective judgments. The most typical measures characterizing this relationship are rank order correlation coefficients, Spearman and, more rarely, Kendall ones [13] (SROCC and KRACC, respectively). The use of rank order correlation avoids data fitting or visual quality metric linearization [13]. Note that from the viewpoint of maximal number of distortion types and observer judgements carried out for MOS obtaining, the database TID2008 [9] is the largest. It contains 1700 distorted images (25 reference images corrupted by 17 types of distortions with four levels each). More than 800 observers from three countries (Ukraine, Italy and Finland) participated in experiments and this provided high accuracy of MOS evaluation.

Second, these databases allow determining drawbacks and malfunctions of the existing metrics. This has led to design of modifications for several metrics [16] that consider the local importance of regions that attract more attention of humans in images. A useful approach in analysis of metrics' performance and their comparisons using TID2008 is that it is possible to use subsets of distortions and to calculate SROCC between a given metric and MOS for each subset. It often occurs that whilst a given metric is the best for one subset, another metric that takes into account some other peculiarities of HVS is preferable for another subset.

For example, according to the results of the study in [17], the metric MSSIM [18] provides the largest SRC for entire set of distorted images in TID2008 (SROCC is equal to 0.853). Meanwhile, for the subset called *Actual* that includes noise of different types and distortions due to lossy compression and filtering, MSSIM is outperformed by several other metrics (SROCC for MSSIM is 0.868 whilst it equals to 0.929 for the metric PSNR-HVS-M [19, 20]). Analysis carried out for other subsets as *Exotic* shows that the metric PSNR-HVS-M has very small SROCC when a given subset contains images with contrast and mean brightness distortions. Thus, this allows determining the drawbacks of this metric and potential ways to improve its performance.

This paper deals with improving the performance of the potentially perspective metrics PSNR-HVS [21] and the aforementioned PSNR-HVS-M. The introduced modifications described relate to accounting for peculiarities of human perception of contrast and mean brightness distortions and their use in metric design. To explain necessity of accounting just for these types of distortions, we briefly describe the database TID2008 and distortion subsets. We also revisit the earlier obtained results (Section II). Then, the modified metrics are presented and the rank correlation factor values for them are given and considered (Section III). The conclusions drawn from this analysis are presented.

TABLE I

DISTORTION TYPES AND CONSIDERED SUBSETS OF TID2008

No	Type of distortion	Noise	Noise2	Noise3	Safe	Hard	Simple	JPEG	Exotic	Exotic2	Exotic3	Actual	Full
1	Additive Gaussian noise	+	+	+	+	-	+	-	-	-	-	+	+
2	Different additive noise in color components	-	+	-	-	-	-	-	-	-	-	-	+
3	Spatially correlated noise	+	+	+	+	+	-	-	-	-	-	+	+
4	Masked noise	-	+	-	-	+	-	-	-	-	-	-	+
5	High frequency noise	+	+	+	+	-	-	-	-	-	-	-	+
6	Impulse noise	+	+	+	+	-	-	-	-	-	+	+	+
7	Quantization noise	+	+	-	-	+	-	-	-	-	-	+	+
8	Gaussian blur	+	+	+	+	+	+	-	-	-	-	+	+
9	Image denoising	+	-	+	-	+	-	-	-	-	-	+	+
10	JPEG compression	-	-	-	+	-	+	+	-	-	-	+	+
11	JPEG2000 compression	-	-	-	+	-	+	+	-	-	-	+	+
12	JPEG transmission errors	-	-	-	-	+	-	-	-	+	-	-	+
13	JPEG2000 transmission errors	-	-	-	-	+	-	-	-	+	-	-	+
14	Non eccentricity pattern noise	-	-	-	-	+	-	-	+	+	+	-	+
15	Local block-wise distortions of different intensity	-	-	-	-	-	-	-	+	+	+	-	+
16	Mean shift (intensity shift)	-	-	-	-	-	-	-	+	+	-	-	+
17	Contrast change	-	-	-	-	-	-	-	+	+	-	-	+

## II. BRIEF REVIEW OF TID2008

As already mentioned, the database TID2008 contains 25 color reference images (24 of them were taken from Kodak image database and the 25<sup>th</sup> was artificially created). All images are of equal size (512x384 pixels) to avoid influence of image size on evaluation of its visual quality in experiments [22]. For each reference image, its 68 distorted versions were created by simulating 17 types of distortions (see the leftmost column of Table I) with four different levels of these distortions. Thus, in aggregate, there are 1700 distorted images.

Visual quality of images has been analyzed by volunteers (observers). Each subject was given one group of 68 distorted images obtained by the same original image (reference). Tri-stimulus analysis has been carried out by simultaneous presentation of the reference and two distorted images. The subject has to select the best between the two distorted images. Comparison results could be processed in different ways. In particular, it was possible to determine MOS by processing all 1700 distorted images (this case is denoted as Full). Note that MOS values in our experiment setup have varied from 0 (very low visual quality) to 9 (practically ideal or definitely the best visual quality).

It was also possible to calculate averaged metric values and MOS for particular subsets created according to some specific types of noise and distortions. A typical example can be the subset called JPEG that includes data only for JPEG and JPEG2000 lossy compression. The analyzed subsets are listed in the upper row of Table I, the signs “+” below in the corresponding columns mark distortion types that are taken into account for a given subset. For example, the subset Noise2 includes the images corrupted by additive white Gaussian noise (distortion type # 1), different intensity spatially uncorrelated noise (additive noise in color components is more intensive than additive noise in the

luminance component, distortion type # 2), spatially correlated noise (distortion type # 3), masked noise (distortion type # 4), high frequency noise (distortion type # 5), impulse noise (distortion type # 6), quantization noise (distortion type # 7), and Gaussian blur (distortion type # 8). As it is seen, the subset includes most typical distortion types met in acquired images before their processing (filtering, compression, etc.).

Note that for this subset our metrics PSNR-HVS and PSNR-HVS-M perform almost perfectly providing SROCC equal to 0.933 and 0.930, respectively. Other good metrics as WSNR, VSNR, and MSSIM produce SROCC equal to 0.908, 0.896, and 0.850, respectively (see data in Table II). Thus, even if further improvement of metric performance is desirable, it is not so crucial.

From analysis of data in Table II it can be also noticed that the metrics PSNR-HVS and PSNR-HVS-M perform well for the subsets *Noise*, *Noise3*, *Safe*, *Simple*, *JPEG*, and *Actual*, i.e. for such subsets that in different combinations include different noise types and distortions due to quantization, compression and filtering (see Table 1). In turn, PSNR-HVS and PSNR-HVS-M poorly perform for the subsets *Exotic* and *Exotic2* (see Table 2) where both subsets include images distorted by Mean (intensity) shift (distortion type # 16) and Contrast change (distortion type # 17).

To further analyze the obtained results, it is useful to detail this type of distortions. The contrast change is illustrated in Fig. 1.a (larger contrast change) and Fig. 1.b (smaller contrast change). As can be noticed, the first distorted image looks better. Contrast change (stretching) is often used in image processing performed by image processing tools or by operator at image visualization stage. It can also be performed “indirectly” by different gamma correction and other types of nonlinear operations widely used in practice. Similar motivations are behind image mean shift. Thus, the distortion

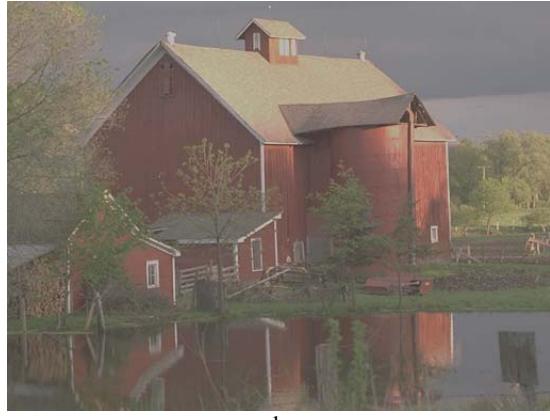
TABLE II

SPEARMAN CORRELATIONS FOR THE CONSIDERED METRICS

No	Metric	Noise	Noise2	Noise3	Safe	Hard	Simple	JPEG	Exotic	Exotic2	Exotic3	Actual	Full
1	PSNR-HA	<b>0.922</b>	<b>0.917</b>	<b>0.930</b>	<b>0.933</b>	<b>0.819</b>	<b>0.939</b>	<b>0.966</b>	<b>0.803</b>	<b>0.821</b>	0.622	<b>0.924</b>	<b>0.868</b>
2	MSSIM	0.813	0.850	0.830	0.849	<b>0.874</b>	0.898	0.957	<b>0.728</b>	<b>0.811</b>	<b>0.673</b>	0.868	<b>0.853</b>
3	PSNR-HMA	<b>0.932</b>	0.912	<b>0.936</b>	<b>0.940</b>	0.811	<b>0.943</b>	<b>0.973</b>	<b>0.822</b>	<b>0.814</b>	0.591	<b>0.934</b>	<b>0.846</b>
4	VIF	0.820	0.900	0.835	0.908	<b>0.844</b>	0.935	0.956	0.531	0.671	0.045	0.841	0.750
5	VSNR	0.857	0.896	0.859	0.888	0.735	0.906	0.930	0.554	0.597	0.490	0.869	0.705
6	SSIM	0.562	0.637	0.570	0.632	0.812	0.769	0.901	0.385	0.594	0.163	0.726	0.645
7	NQM	0.865	0.887	0.865	0.896	0.733	0.903	0.932	0.602	0.432	0.517	0.874	0.624
8	UQI	0.526	0.599	0.485	0.638	0.759	0.784	0.860	0.292	0.546	0.156	0.677	0.600
9	PSNR-HVS	0.917	<b>0.933</b>	<b>0.930</b>	0.932	0.791	<b>0.939</b>	<b>0.966</b>	0.275	0.324	0.541	0.920	0.594
10	XYZ	0.848	0.834	0.872	0.822	0.791	0.820	0.815	0.155	0.188	<b>0.679</b>	0.829	0.577
11	IFC	0.663	0.743	0.673	0.775	0.736	0.817	0.898	-0.269	0.276	-0.075	0.732	0.569
12	PSNRHVS-M	<b>0.918</b>	<b>0.930</b>	0.922	<b>0.936</b>	0.783	<b>0.942</b>	<b>0.971</b>	0.274	0.287	0.518	<b>0.929</b>	0.559
13	SNR	0.712	0.687	0.698	0.699	0.646	0.794	0.805	0.227	0.290	0.561	0.760	0.523
14	PSNR	0.704	0.612	0.698	0.689	0.697	0.799	0.877	0.248	0.308	<b>0.671</b>	0.794	0.525
15	WSNR	0.897	0.908	0.892	0.921	0.776	0.931	0.949	0.157	0.059	0.544	0.900	0.488
16	DCTUNE	0.864	0.881	0.868	0.877	0.703	0.902	0.933	0.529	0.260	0.556	0.860	0.476



a



b

Fig.1 Reference image # 22 with change to larger contrast (a) and smaller contrast (b)

types # 16 and 17 are not so rare in practice and it is worth accounting them in metric design.

In the following details concerning PSNR-HVS and PSNR-HVS-M metrics are reported. PSNR-HVS-M is an extension of PSNR-HVS able to cope with an effect of HVS known as visual masking [23]. The outcome of this effect is that distortions in spatial frequencies that have small absolute

values of amplitudes are practically not seen if in a given fragment (block) there exist one or two spatial frequency components that are considerably more intensive (have essentially larger amplitudes). Due to this, a PSNR-HVS-M (both are expressed in dB) for a given image is always not smaller than PSNR-HVS. Both metrics take into account the effect of different sensitivity of HVS to distortions in different spatial frequencies. The considered metrics are calculated as

$$PSNR - HVS = 10 \log_{10} (255^2 / MSE_{HVS})$$

$$PSNR - HVS - M = 10 \log_{10} (255^2 / MSE_{HVS-M})$$

for each 8-bit color component (the results are averaged then) where  $MSE_{HVS}$  and  $MSE_{HVS-M}$  are mean square errors (MSEs) weighted in DCT domain with specific weights (see details in [19, 21]). Any contrast change and mean shift lead to non-zero MSEs and, respectively, non-zero values of the weighted versions  $MSE_{HVS}$  and  $MSE_{HVS-M}$ . Larger contrast change and mean shift result in larger  $MSE_{HVS}$  and  $MSE_{HVS-M}$  and, as a result, smaller PSNR-HVS and PSNR-HVS-M treated as worse visual quality of considered images.

Recall that the metric MSSIM also called image quality index that varies in the limits from 0 (very bad quality) to 1 (perfect quality) allows taking into account several aspects of HVS including such psycho-visual feature that image stretching and mean changing (in some limits) do not considerably influence human perception of images (MSSIM code can be freely downloaded from [24]). Due to this, it performs considerably better (produces larger SROCC, see Table 2) for the subsets *Exotic* and *Exotic2* than PSNR-HVS and PSNR-HVS-M. Therefore, it is reasonable to take advantage of MSSIM properties and operation principles in design of modifications of PSNR-HVS and PSNR-HVS-M further called PSNR-HA and PSNR-HMA, respectively. The ways to do this are proposed in the next Section.

TABLE III

KENDALL CORRELATIONS FOR THE CONSIDERED METRICS

Nº	Metric	Noise	Noise2	Noise3	Safe	Hard	Simple	JPEG	Exotic	Exotic2	Exotic3	Actual	Full
1	PSNR-HA	<b>0.757</b>	<b>0.751</b>	<b>0.762</b>	<b>0.775</b>	<b>0.643</b>	<b>0.788</b>	<b>0.837</b>	<b>0.596</b>	<b>0.624</b>	0.451	<b>0.757</b>	<b>0.689</b>
2	PSNR-HMA	<b>0.770</b>	0.742	<b>0.770</b>	<b>0.785</b>	0.632	<b>0.792</b>	<b>0.857</b>	<b>0.616</b>	<b>0.614</b>	0.418	<b>0.773</b>	<b>0.660</b>
3	MSSIM	0.609	0.650	0.631	0.649	<b>0.676</b>	0.719	0.818	<b>0.522</b>	<b>0.604</b>	<b>0.478</b>	0.675	<b>0.654</b>
4	VIF	0.634	0.729	0.645	0.742	<b>0.660</b>	0.776	0.814	0.370	0.499	0.092	0.657	0.586
5	VSNR	0.665	0.713	0.663	0.701	0.546	0.725	0.764	0.377	0.418	0.372	0.677	0.534
6	PSNR-HVS	0.751	<b>0.780</b>	<b>0.766</b>	0.772	0.614	0.785	<b>0.837</b>	0.195	0.238	0.385	0.750	0.476
7	SSIM	0.388	0.450	0.388	0.437	0.618	0.564	0.718	0.266	0.431	0.139	0.527	0.468
8	NQM	0.673	0.704	0.677	0.713	0.541	0.720	0.766	0.428	0.288	0.349	0.678	0.461
9	PSNRHVS <sub>M</sub>	<b>0.752</b>	<b>0.771</b>	0.755	<b>0.778</b>	0.606	<b>0.789</b>	<b>0.847</b>	0.194	0.210	0.364	<b>0.765</b>	0.449
10	UQI	0.363	0.420	0.330	0.454	0.565	0.587	0.666	0.196	0.389	0.115	0.489	0.435
11	XYZ	0.654	0.641	0.677	0.631	0.594	0.638	0.633	0.104	0.138	<b>0.480</b>	0.638	0.434
12	IFC	0.477	0.547	0.482	0.581	0.552	0.624	0.714	-0.156	0.208	0.004	0.542	0.426
13	WSNR	0.714	0.736	0.712	0.753	0.586	0.766	0.797	0.107	0.047	0.379	0.715	0.393
14	SNR	0.512	0.492	0.498	0.497	0.464	0.593	0.604	0.154	0.205	0.396	0.558	0.374
15	DCTUNE	0.683	0.711	0.690	0.701	0.527	0.735	0.791	0.357	0.170	0.379	0.676	0.372
16	PSNR	0.501	0.424	0.490	0.486	0.516	0.598	0.692	0.178	0.225	<b>0.488</b>	0.593	0.369

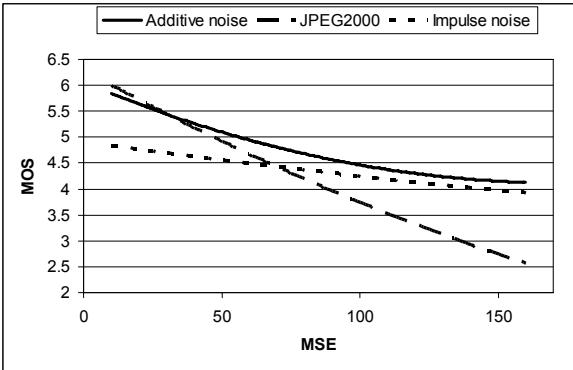


Fig.2 Dependence of MOS on MSE for images of TID2008 corrupted by AWGN, distortions due to JPEG2000 and impulse noise

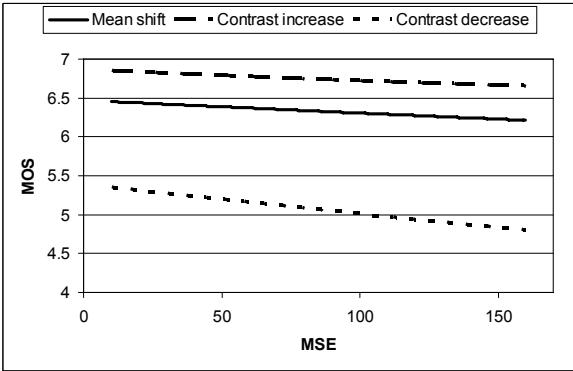


Fig.3 Dependence of MOS on MSE for mean shift, contrast increase and reduction.

### III. PROPOSED MODIFICATIONS AND THEIR PERFORMANCE ANALYSIS

The studies carried out in [17, 25] have demonstrated that HVS is considerably less sensitive to mean brightness and contrast change distortions than to other types of distortions as, e.g., additive white Gaussian noise (AWGN). Note that for the same MSE there are many distortion types resulting more annoying than AWGN. To demonstrate this

phenomenon, Fig.2 presents dependences of MOS on MSE for AWGN, lossy compression by JPEG2000 [26] and impulse noise. Keeping in mind that larger MOS corresponds to better visual quality, it can be concluded that for small MSE values impulse noise is the most annoying whilst for large MSE values the distortions due to lossy JPEG2000 lead to the worst visual quality of distorted images.

Fig. 3 presents the dependences of MOS on MSE for mean brightness distortions as well as for contrast increase and reduction. These plots are also obtained using the database TID2008. As it can be noticed, HVS is the least sensitive to contrast increase (MOS is the largest for all considered values of MSE). HVS is more sensitive to mean shift and the most sensitive to contrast reduction. HVS sensitivity to contrast reduction is practically the same as sensitivity to AWGN (compare the corresponding curves in Figures 2 and 3).

Analysis of dependences in Figures 2 and 3 also shows that the standard metric MSE is not adequate to HVS (peculiarities of image perception by humans) since for the same MSE values there is considerable difference in MOS. Fig. 4 presents similar dependences for the metric PSNR-HVS-M. This metric match HVS scores considerably better since for the same PSNR-HVS-M the difference between MOS values is sufficiently smaller (although it is not equal to zero as well). However, since PSNR-HVS-M in no way takes into account mean shift and contrast change, the plots of MOS vs. MSE in Fig. 5 are practically the same as in Fig. 3.

It is possible to notice that the plots in Fig. 5 can be converted to plots in Fig. 4 using first order polynomial approximation as  $p1*x + p2$  where  $p1$  and  $p2$  are correcting factors and  $x$  denotes argument. The main problem is to detect situations when mean shift and contrast change have occurred and corrections to calculation of MSE are to be introduced.

The proposed modifications of the metrics PSNR-HVS (the modified version is PSNR-HA) and PSNR-HVS-M (the modified version is PSNR-HMA) are the following:

1. For a given reference image **A** and the corresponding analyzed distorted image **B**, calculate difference of image

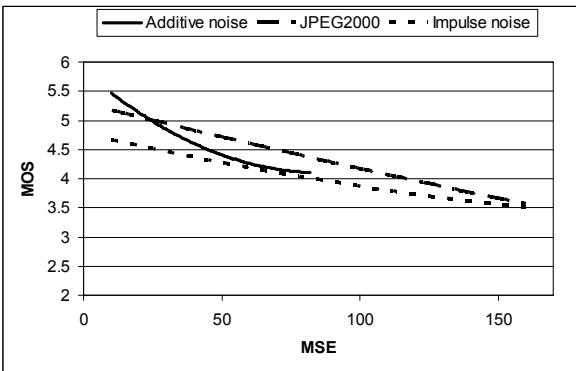


Fig. 4 Dependence of MOS on  $MSE_{HVS-M}$  for AWGN, distortions due to JPEG2000 and impulse noise

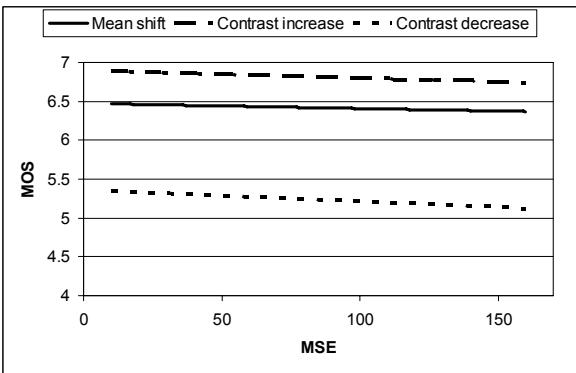


Fig. 5 Dependence of MOS on  $MSE_{HVS-M}$  for mean shift, contrast increase and reduction

means as  $Delt = \bar{A} - \bar{B}$  where  $\bar{A}$  and  $\bar{B}$  denote image mean values.

2. Obtain the image **C** with corrected mean as  $\mathbf{C} = \mathbf{B} + Delt$ .

3. Calculate the correcting factor  $Popr$  for correcting possible contrast change as:

$$Popr = \frac{\sum(\mathbf{A} - \bar{A})(\mathbf{C} - \bar{C})}{\sum(\mathbf{C} - \bar{C})^2}.$$

4. Obtain the corrected image **D** with corrected contrast as  $\mathbf{D} = \mathbf{C} * Popr$  (this operation guarantees minimally possible MSE between the images **A** and **D** [27]).

5. Determine  $MSE_{HVS}$  for the metric PSNR-HVS for the images **A** and **C** and denote it as  $M1$  (if the metric PSNR-HVS-M is to be calculated, then determine  $MSE_{HVS-M}$  for these images, other operations are the same as below).

6. Determine  $MSE_{HVS}$  for the images **A** and **D** (denote it as  $M2$ ).

7. If  $M1 > M2$ , then we deal with contrast change and  $M1$  is corrected as

$$M1 = M2 + \begin{cases} (M1 - M2)Coef1, & Popr < 1 \\ (M1 - M2)Coef2, & Popr \geq 1 \end{cases}.$$

8. The final MSE is calculated as

$$M = M1 + Delt^2 * Coef3.$$

9. Then, using the obtained value if  $M$ , the modified metric is calculated as (and similarly for PSNR-HMA)

$$PSNR-HA = 10 \log_{10} \left( \frac{255^2}{M} \right).$$

The correcting factors  $Coef1$  and  $Coef2$  are used for accounting for errors induced by image contrast increase or reduction, respectively. The factor  $Coef3$  takes into account errors due to image mean shift.

Dealing with color images both reference and distorted images are first transformed to color space YCbCr. Then the error  $M$  is computed separately for each component and the errors  $MY$ ,  $MCb$  and  $MCr$  are aggregated as  $(MY + MCb * Coef4 + MCr * Coef4) / (1 + 2 * Coef4)$  where  $Coef4$  is the weighting factor for color components.

To obtain the correcting factors  $Coef1 \dots Coef4$ , numerical optimization methods have been used (Monte-Carlo and table method) and test images of the database TID2008. For each set of correcting coefficients, SROCC of the obtained metric and MOS has been calculated. As optimization result, the SROCC value was to be maximized. The following optimal correcting factors have been obtained:  $Coef1=0.002$ ,  $Coef2=0.25$ ,  $Coef3=0.04$ ,  $Coef4=0.5$ , the same for both metrics.

To verify the proposed modifications PSNR-HA and PSNR-HMA the Spearman and Kendall rank order correlation coefficients have been determined for TID2008 (see Tables II and III). For comparison purposes, the SROCC and KROCC values have been also calculated for many other existing metrics: PSNR-HVS, PSNR-HVS-M (abbreviated as PSNRHVSM), MSSIM, VIF [28], VSNR [29], SSIM [7], NQM [30], UQI [31], XYZ [32], IFC [33], WSNR [34], DCTUNE [35], SNR and PSNR. Codes for most of them are available [24]. In Table III, we give the KROCC values for the full set of distorted images and the subsets.

First of all, the analysis of data in Tables I and III shows similar behavior. The only difference is that the values of KROCC are smaller than the corresponding values of SROCC. Second, due to accounting for contrast change and mean shift (the degradation types included into subsets *Exotic* and *Exotic2*) the metrics PSNR-HA and PSNR-HMA considerably outperform the corresponding prototypes PSNR-HVS and PSNR-HVS-M (see data for the subsets *Exotic* and *Exotic2*). The results for the subset *Exotic3* have improved as well. Meanwhile, the values of SROCC for PSNR-HVS and PSNR-HA for the subsets that do not include contrast change and mean shift distortions (e.g., the subsets *Noise* and *Noise2*) practically have not changed. The same relates to the metrics PSNR-HVS-M and PSNR-HMA. Similar effects are observed for KROCC. In aggregate, this has lead that SROCC and KROCC values for the metrics PSNR-HA and PSNR-HMA have become the largest for all images of the database TID2008. The metric PSNR-HA is the most adequate to HVS according to both SROCC and KROCC for entire database whilst PSNR-HMA is the best for the subset *Actual*. This subset includes distortions most typical for practice.

For an effective metric, a scatter-plot of MOS depending upon this metric should have a compact form without outliers and with a tendency to monotonous behavior, either increasing or decreasing. Fig. 6 shows the scatter-plot of MOS vs. PSNR-HVS. As it can be noticed a compact region

with some outliers is present. Fig. 7 presents the scatter-plot for the proposed metric PSNR-HA. Comparison of these plots shows that the percentage of outliers decreased. This means that PSNR-HA is more adequate to human perception of distorted images.

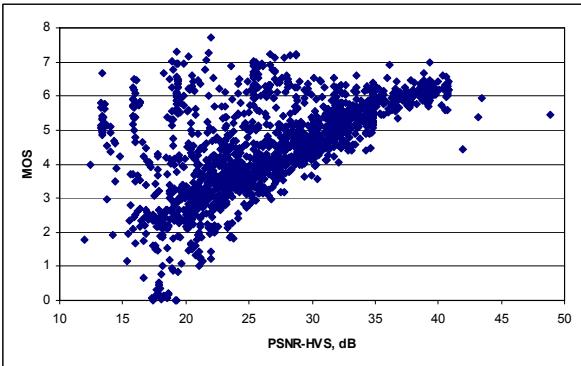


Fig. 6 The scatter-plot of MOS and PSNR-HVS

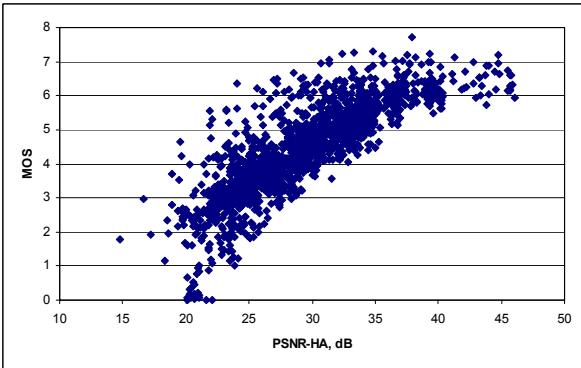


Fig. 7 The scatter-plot of MOS and PSNR-HA

#### IV. CONCLUSIONS

It is demonstrated that by introducing modifications into the metrics PSNR-HVS and PSNR-HVS-M it is possible to improve their performance. The modifications provide less sensitivity to specific types of distortions as mean level shift and contrast stretching. Analysis carried out for a large number of distorted color images of the database TID 2008 shows that for the modified metrics their rank correlation with MOS has been considerably increased and the modified metrics are now in the top-three of the known visual quality metrics.

#### REFERENCES

- [1] W.K. Pratt, "Digital Image Processing, Fourth Edition," *Wiley-Interscience*, NY, USA, 2007.
- [2] M. Elad, "Sparse and Redundant Representations. From Theory to Applications in Signal and Image Processing," *Springer Science+Business Media*, LLC, 2010, p. 376.
- [3] D. Salomon, "Data Compression: The Complete Reference," *Springer-Verlag*, 2004.
- [4] B.W. Keelan, "Handbook of Image Quality," *Marcel Dekker*, New York, USA, 2002, 516 p.
- [5] A. Bovik, "Handbook on Image and Video Processing," *Academic Press*, USA, 2000.
- [6] F. De Simone, D. Ticca, F. Dufaux, M. Ansorge, T. Ebrahimi, "A Comparative study of color image compression standards using perceptually driven quality metrics," *Proceedings of SPIE Conference Applications of Digital Image Processing XXXI*, San Diego, USA, August 2008, 10 p.
- [7] 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, Apr. 2004, pp. 600-612.
- [8] N. Ponomarenko, V. Lukin, K. Egiazarian, J. Astola, M. Carli, F. Battisti, "Color Image Database For Evaluation Of Image Quality Metrics", *IEEE Signal Processing Society 2008 International Workshop on Multimedia Signal Processing*, Australia, Sept. 2008, pp.403-408.
- [9] H.R. Sheikh, M.F. Sabir, F.C. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, 2006, pp. 3441-3452.
- [10] W. Zeng, S. Daly, S. Lei, "An overview of the visual optimization tools in JPEG 2000," *Signal Processing: Image Communication*, Volume 17, Number 1, January 2002, pp. 85-104.
- [11] TID2008 page: <http://ponomarenko.info/tid2008.htm>
- [12] J. Redi, H. Liu, H. Alers, R. Zunino, I. Heynderickx, "Comparing subjective image quality measurement methods for the creation of public databases," *Proceedings of SPIE Conference on Image Quality and System Performance VII*, SPIE vol. 7529, 2010, 1 p.
- [13] K. Okarma, "Combined full-reference image quality metric linearly correlated with subjective assessment," *Springer Lecture Notes on Artificial Intelligence*, Vol. LNAI 6113, 2010, pp. 539-546.
- [14] F. Zhang, S. Li, L. Ma, K.N. Ngan, "Limitations and challenges of image quality measurement," *Proceedings of SPIE*, Vol. 7744, 2010, 8 p.
- [15] M.G. Kendall, "The advanced theory of statistics," *Charles Griffin & Company Ltd*, London, UK, Vol. 1, 1945, 457 p.
- [16] N. Ponomarenko, S. Krivenko, K. Egiazarian, V. Lukin, J. Astola, "Weighted mean square error for estimation of visual quality of image denoising methods," *CD ROM Proceedings of VPQM*, Scottsdale, USA, 2010, 5 p.
- [17] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, "Metrics Performance Comparison for Color Image Database," *CD-ROM Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, Scottsdale, USA, 2009, 6 p.
- [18] Z. Wang, E.P. Simoncelli, A.C. Bovik, "Multi-scale structural similarity for image quality assessment," *Proceedings of IEEE ACSSC*, San Francisco, USA, 9-12 Nov. 2003, pp. 1398-1402.
- [19] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, V. Lukin, "On between-coefficient contrast masking of DCT basis functions," *CD-ROM Proc. of the Third Int. Workshop on Video Proc. and Quality Metrics*, USA, 2007, 4 p.
- [20] PSNR-HVS-M page: <http://ponomarenko.info/psnrhvsmt.htm>
- [21] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, M. Carli, "New full-reference quality metrics

- based on HVS," *Proc. of the Second Int. Workshop on Video Proc. and Quality Metrics*, Scottsdale, USA, Jan. 2006, 4 p.
- [22] N. Ponomarenko, O. Eremeev, K. Egiazarian, V. Lukin, "Statistical evaluation of no-reference image visual quality metrics," *Proceedings of EUVIP*, Paris, France, 2010, 5 p.
- [23] J.A. Solomon, A.B. Watson, A. Ahumada, "Visibility of DCT basis functions: Effects of contrast masking," *Proceedings of Data Compression Conference*, Snowbird, Utah: IEEE Computer Society Press, 1994, pp. 361-370.
- [24] M. Gaubatz, "Metrix MUX Visual Quality Assessment Package": [http://foulard.ece.cornell.edu/gaubatz/metrix\\_mux/](http://foulard.ece.cornell.edu/gaubatz/metrix_mux/)
- [25] N. Ponomarenko, M. Carli, V. Lukin, K. Egiazarian, J. Astola, F. Battisti, "Color Image Database for Evaluation of Image Quality Metrics," *Proceedings of International Workshop on Multimedia Signal Processing*, Cairns, Australia, 8-10 Oct. 2008, pp. 403-408.
- [26] D. Taubman, M. Marcellin, "JPEG 2000: Image Compression Fundamentals, Standards and Practice," Boston, USA: Kluwer, 2002, 773 p.
- [27] N. Ponomarenko, K. Egiazarian, V. Lukin, "Acceleration of fractal image compression by correlation trees, Interdisciplinary Applications of Fractal and Chaos Theory," *Editura Academiei Romane*, 2004, pp. 75-84.
- [28] H.R. Sheikh, A.V. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, Vol.15, no.2, 2006, pp. 430-444.
- [29] D.M. Chandler, S.S. Hemami, "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images," *IEEE Transactions on Image Processing*, Vol. 16 (9), 2007, pp. 2284-2298.
- [30] N. Damera-Venkata, T. Kite, W. Geisler, B. Evans, A. Bovik, "Image Quality Assessment Based on a Degradation Model," *IEEE Trans. on Image Processing*, Vol. 9, 2000, pp. 636-650.
- [31] Z. Wang, A. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, Vol. 9, 2000, pp. 81-84.
- [32] B.W. Kolpatzick, C.A. Bouman, "Optimized Universal Color Palette Design for Error Diffusion," *Journal Electronic Imaging*, Vol. 4, 1995, pp. 131-143.
- [33] H.R. Sheikh, A.C. Bovik, G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, Vol. 14, 2005, pp. 2117-2128.
- [34] T. Mitsa, K. Varkur, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithm," *Proceedings of ICASSP*, Minneapolis, USA, April 1993, pp. 301-304.
- [35] A.B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images," *Soc. Inf. Display Dig. Tech. Papers*, Vol. XXIV, 1993, pp. 946-949.