METRICS PERFORMANCE COMPARISON FOR COLOR IMAGE DATABASE

Nikolay Ponomarenko (*), Federica Battisti (**), Karen Egiazarian (***), 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 exploit a new database of distorted test images TID2008 for verification of full-reference metrics of image visual quality. A comparative analysis of TID20008 and its nearest analog LIVE Database is presented. For a wide variety of known metrics, their correspondence to human visual system is evaluated. The values of rank correlations of Spearman and Kendall with the considered metrics and Mean Opinion Score (MOS) obtained by exploiting TID2008 in experiments are presented. The metrics are verified for both full set of distorted test images in TID2008 (1700 distorted images, 17 types of distortions) and for particular subsets of TID2008 that include distortions most important for digital image processing applications.

1. INTRODUCTION

Recently the scientific community has done great efforts to develop and test image and video quality assessment methods incorporating perceptual measures. Many of the quality metrics proposed were based on properties of human vision system (HVS) [1,2]. However, till now there are no such quality metrics that are able to take into account all peculiarities of HVS. There are several reasons for that [1,3]. First, HVS is not well understood yet. Second, it is not clear how to model all possible image distortion types and levels. Third, people use different image databases to carry out testing of existing and new quality metrics. Fourth, experiments with appropriate number of volunteers are needed for assessment of image visual quality and reliable testing of image visual quality metrics.

Conditions to carry out experiments, methodology to process the results of these experiments, what is a necessary number of participants, etc., are other questions yet to be answered. Note also that quality metrics can be application-aided. In this case, there is no need to invent some “universal” quality metric, but just to select the “best suited” one among existing ones for a given particular application.

We have created a color image database TID2008 [4] that allows to alleviate some of shortcomings mentioned above and to make a comparison of different quality metrics integrally for all considered types of distortions and specially for particular groups (subsets) of distortion types. This database provides, among others, the following opportunities:

1) It allows testing new quality metrics (if their implementation codes are available);
2) The database provides means for calculation of mean opinion score for already performed experiments and adding data of new experiments;
3) It is possible to consider applicability of quality metrics for particular applications by grouping experiment data for a given type of distortion(s) and analyzing Spearman or Kendall correlations for the tested quality metrics;
4) The database allows determining types of distortions for which a given quality metric performs poorly, thus showing its drawbacks and, probably, indicating what should be taken into account for improvement of metric’s performance.

TID2008 (original test images, distorted ones and MOS) is freely available for downloading from [5].

In this paper we report new results of studies based on TID2008 exploiting, after the preliminary research data analysis given in [4]. In addition to nine quality metrics considered in the paper [4] we have obtained the results for eight other visual quality metrics, namely, SNR [6], VSNR [7,6], WSNR [8,6], MSSIM [9,6], NQM [10,6], VIF [11,6], VIFP [11,6], and IFC [12]. We have also considered four new subsets of TID2008 that relate to important practical applications and have performed brief analysis of data obtained for these new subsets.

2. SHORT REVIEW OF TID2008

At the moment, TID2008 [4] is the largest database of distorted images intended for verification of full-reference quality metrics (indices). Table 1 presents main parameters and characteristics allowing the comparison of TID2008 to its nearest analog LIVE Database [3].

Table 1. Comparison characteristics of LIVE Database and TID2008 Database

N	Main characteristics	Test image database	
		LIVE Database	TID2008
1	Number of distorted images	779	1700
2	Number of different types of distortions	5	17
3	Number of experiments carried out	161 (all USA)	Totally 838 (437 - Ukraine, 251 - Finland, 150 - Italy)
4	Methodology of visual quality evaluation	Evaluation using five level scale (Excellent, Good, Fair, Poor, Bad)	Pair-wise sorting (choosing the best that visually differs less from original between two considered)
5	Number of elementary evaluations of image visual quality in experiments	25000	256428
6	Scale of obtained estimates of MOS	0..100 (stretched from the scale 1..5)	0..9
7	Variance of estimates of MOS	250	0.63
8	Normalized variance of estimates of MOS	0.083	0.031

The main advantage of TID2008 with respect to LIVE Database is that TID2008 accounts for 17 different types of distortions and, thus, covers more practical applications and known peculiarities of human visual system (HVS). LIVE Database deals with only five types of distortions for which most known metrics tested using LIVE Database commonly have quite good correspondence to HVS. In turn, TID2008 allows carrying out more thorough analysis of quality metrics indicating their drawbacks (not observed yet) and demonstrating prospective ways of further investigations and design.

Another peculiarity of TID2008 is that there are some changes introduced in conventional (standard) methodology of estimating MOS. For conventional methodology, an observer evaluates image visual quality using some scale (for LIVE Database the 5-level scale is used). This conventional methodology has several drawbacks [13,14] that lead to problems arising in carrying out experiments. First, it is quite difficult for a human to assess image visual quality in units (marks). Second, it leads to often met situations when a human has no imagination about quality of images met in a database and, after putting some marks at initial stage of assessment, this human runs into cases when he/she has no possibility to put a better or worse mark to a given considered image although he/she is willing to do so.

Psychologically, it is simpler for a given pair of distorted images to choose a better (worse) one or to point an image that visually differs less from the corresponding original (undistorted) image. Almost all participants of our experiments voted in favor of the latter methodology. The authors of [13,14] also state that assigning marks is not the only possible methodology to carry out visual quality experiment.

This methodology has been used by us in carrying out experiments using TID2008. In each comparison, an image in a considered pair that visually differs less from original one was getting one point. Each distorted image participated in nine comparisons during each experiment. Units were summed up for each distorted image and, thus, each distorted image could get from 0 to 9 points in one experiment. For comparison, images from the same "point" group were selected similarly as it is done in Swiss system chess competition. Note that by experiment we mean that one observer is carrying out comparisons for a given "family" of distorted images that have been generated for a given original image (there are 25 such original images in TID2008).

As the result, for approximately the same number of quality assessments for each original image group in TID2008 and LIVE Database, we have almost three times smaller variance of MOS estimates for experiments with TID2008. The data for LIVE Database given in Table 1 have been obtained by processing the results of 60 experiments carried out for the first original image.

It is worth noting that for the methodology [3] used in experiments with LIVE Database the relative number of abnormal experiments was about 10%; for TID2008 and our methodology we had about 1.5% of such abnormal experiments. This also evidences in favor of the pair-wise comparison methodology as being simpler and convenient for observers. Also note that Spearman correlation for MOS obtained for the first original image family (68 distorted images) for both methodologies of experiment carrying out is equal to 0.97. This correlation factor is very high and it demonstrates that both methodologies lead to similar results. This shows that both methodologies can be used and they differ only by convenience for observers and by the provided accuracy of MOS estimation.

3. COMPARATIVE ANALYSIS OF QUALITY METRICS

The types of distortions of TID2008 are presented in Table 2. 17 types of distortions simulated with four intensity levels have been simulated. The different types of distortions can be grouped in subsets (distortions that belong to a given subset are marked by +). In addition to seven subsets analyzed in [3], in this paper we consider four new subsets that are important for the most intensively used and studied image processing applications, namely:

- the subset called "JPEG" that includes distortions introduced due to compression for two image compression standards JPEG and JPEG2000;
- the subset called "Noise3" that includes distortions due to different types of noise, blur and denoising;
- the subset called "Exotic3" that includes types of distortions that occurred to be the most "complex" for accounting by the considered quality metrics;
- the subset called "Actual" that includes types of distortions most typical for actual tasks of image filtering and lossy compression.

Table 2. Distortion types and considered subsets of TID2008

№	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	-	-	-	-	-	-	-	+	+	-	-	+

Table 3. Spearman correlations for the considered metrics

№	Metric	Noise	Noise2	Noise3	Safe	Hard	Simple	JPEG	Exotic	Exotic2	Exotic3	Actual	Full
-	HVS	*0.991*	*0.991*	*0.991*	*0.993*	*0.994*	*0.994*	*0.996*	*0.994*	*0.994*	*0.985*	*0.994*	*0.994*
1	MSSIM	0.813	0.850	0.830	0.849	***0.874***	0.898	***0.957***	***0.728***	***0.811***	***0.673***	0.868	***0.853***
2	VIF	0.820	0.900	0.835	0.908	***0.844***	***0.935***	0.956	0.531	***0.671***	0.045	0.841	***0.750***
3	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***
4	VIFP	0.734	0.749	0.729	0.814	0.785	0.909	0.949	0.523	***0.651***	0.033	0.821	0.655
5	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
6	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
7	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
8	PSNRHVS	***0.917***	***0.933***	***0.930***	***0.932***	0.791	***0.939***	***0.966***	0.275	0.324	0.541	***0.920***	0.594
9	XYZ	0.848	0.834	0.872	0.822	0.791	0.820	0.815	0.155	0.188	***0.679***	0.829	0.577
10	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
11	PSNRHVSM	***0.918***	***0.930***	***0.922***	***0.936***	0.783	***0.942***	***0.971***	0.274	0.287	0.518	***0.929***	0.559
12	PSNRY	0.752	0.723	0.730	0.744	0.690	0.851	0.866	0.242	0.313	0.630	0.810	0.553
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	MSE	0.704	0.612	0.698	0.689	0.697	0.799	0.877	0.248	0.308	***0.671***	0.794	0.525
15	PSNR	0.704	0.612	0.698	0.689	0.697	0.799	0.877	0.248	0.308	***0.671***	0.794	0.525
16	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
17	LINLAB	0.839	0.853	0.847	0.859	0.761	0.877	0.906	0.135	0.033	0.604	0.847	0.487
18	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

Table 4. Kendall correlations for the considered metrics

№		Noise	Noise2	Noise3	Safe	Hard	Simple	JPEG	Exotic	Exotic2	Exotic3	Actual	Full
-	HVS	*0.921*	*0.919*	*0.919*	*0.931*	*0.930*	*0.938*	*0.947*	*0.932*	*0.935*	*0.902*	*0.933*	*0.935*
1	MSSIM	0.609	0.650	0.631	0.649	***0.676***	0.719	*0.818*	***0.522***	***0.604***	0.478	0.675	***0.654***
2	VIF	0.634	0.729	0.645	0.742	***0.660***	0.776	0.814	0.370	***0.499***	0.092	0.657	***0.586***
3	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***
4	VIFP	0.536	0.554	0.527	0.618	0.597	0.738	0.806	0.364	***0.481***	0.082	0.631	0.495
5	PSNRHVS	***0.751***	***0.780***	***0.766***	***0.772***	0.614	***0.785***	***0.837***	0.195	0.238	0.385	***0.750***	0.476
6	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
7	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
8	PSNRHVSM	***0.752***	***0.771***	***0.755***	***0.778***	0.606	***0.789***	***0.847***	0.194	0.210	0.364	***0.765***	0.449
9	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
10	XYZ	0.654	0.641	0.677	0.631	0.594	0.638	0.633	0.104	0.138	***0.480***	0.638	0.434
11	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
12	PSNRY	0.549	0.521	0.522	0.536	0.504	0.657	0.670	0.172	0.229	0.452	0.609	0.402
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	LINLAB	0.652	0.671	0.659	0.682	0.569	0.715	0.758	0.084	0.026	0.422	0.665	0.381
15	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
16	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
17	MSE	0.501	0.424	0.490	0.486	0.516	0.598	0.692	0.178	0.225	***0.488***	0.593	0.369
18	PSNR	0.501	0.424	0.490	0.486	0.516	0.598	0.692	0.178	0.225	***0.488***	0.593	0.369

We have evaluated correspondence of HVS to the following 18 metrics (quality indices): MSSIM [9, 6], VIF [11, 6], a pixel based version of VIF (VIFP) [11, 6], VSNR [7, 6], PSNR-HVS (PSNRHVS) [15, 17], PSNR-HVS-M (PSNRHVSM) [16, 17], SSIM [18], NQM [10, 6], UQI [19], XYZ [20], LINLAB [21], IFC [12, 6], WSNR [8, 6], DCTUNE [22], SNR [6], MSE [6], PSNR [6] and PSNR calculated for only brightness (intensity) component of color images (PSNRY). Table 3 presents the values of Spearman correlation for the considered 18 metrics and the subsets used in TID2008 (see Table 2). Similarly, Table 4 contains the corresponding values of Kendall correlation factors. The first row of both Tables presents correlations between obtained MOS and "ideal" MOS that could be provided if the number of experiments approaches to infinity.

Here we would like to emphasize the following. Most metrics analyzed in this paper are oriented on applying for grayscale images. Thus, they have been calculated with respect to intensity images of color images used in TID2008. Meanwhile, TID2008 contains three types of distortions (namely, numbers 2, 10, and 12) that are not uniformly distributed between color (RGB) components. Thus, TID2008 can be used to verify both types of metrics, those that take and those that not take into consideration color information. Three best metrics producing the greatest correlations for each subset are marked by bold in Tables 3 and 4.

We would like to draw readers' attention to the fact that Spearman correlation values for the metrics PSNR and MSE are equal to each other. If the conventional Pearson correlation is used, then, without fitting, the correlation factor for these metrics might be not equal to unity although they are absolutely strictly connected. Then, for increasing correlation of metrics one needs their fitting [3]. Because of this, it is preferable to employ rank correlation that avoids necessity of fitting in the considered analysis. This is also important because a quality of fitting commonly influences accuracy of obtained results.

The data presented in Tables 3 and 4 for the whole image database TID2008 (the set marked as "Full") show that the widely used metrics PSNR and MSE have very low correlation with human perception (correlation factors are about 0.5). Even the best among considered metric MSSIM has correlation with HVS of the order 0.85 whilst it is desirable to provide a Spearman correlation value around 0.99.

For the subset "JPEG", the best Spearman correlation (SC) between MOS and analyzed metrics is provided by our metrics PSNR-HVS and PSNR-HVS-M [16, 17] (SC is about 0.97), slightly smaller SCs are observed for the metrics MSSIM and VIF (about 0.96). Applicability of these metrics for lossy image compression has been also recently pointed in [23].

For the subset "Noise3", the largest values of both Spearman and Kendall correlations (about 0.93 and 0.76, respectively) have been observed for the metrics PSNR-HVS and PSNR-HVS-M as well. The metric Weighted SNR (WSNR) performs for this set rather well (SC is about 0.90 and Kendall correlation equals to 0.72).

For "Exotic3" subset (that, in fact, includes different versions of impulsive distortions) even the best metric MSSIM exhibits low values of SC and KC which are the smallest between the considered subsets MOS (SC=0.679, KC=0.488). Surprisingly, just for this subset the metrics

MSE and PSNR perform better than other metrics (according to Kendall correlation). This indirectly shows that till now practically no attention in metric design and analysis has been paid to consider distortions collected in the subset "Exotic3".

For the subset "Actual" that collects the most widely met types of distortions in the area of color image processing, the first two places are occupied by our metrics PSNR-HVS and PSNR-HVS-M. Their correlations with MOS are not ideal but they are, at least, larger than for other metrics. This allows recommending them for evaluating efficiency of image filtering and lossy compression. Matlab code for the metric PSNR-HVS-M is freely available from [17].

4. EXAMPLES OF APPLYING METRICS FOR PARTICULAR TYPES OF DISTORTIONS

Fig. 1 shows dependence of MOS on PSNR for image compression applications using JPEG and JPEG2000. MOS equal to 68 shows that all observers put a given image to the first place in the sorted sequence of distorted images.

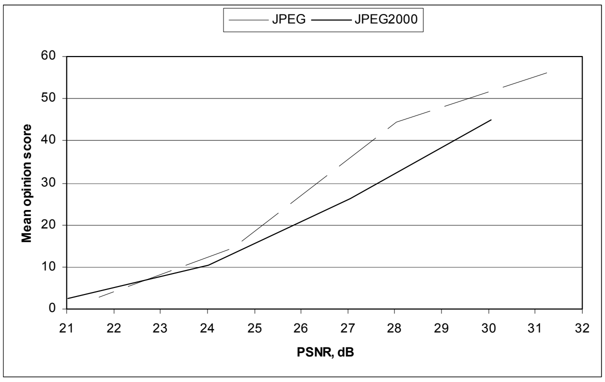

Fig. 1. Dependence of MOS on PSNR for JPEG and JPEG2000

As seen, for large PSNR, the same visual quality (equal MOS values) is observed for JPEG2000 and JPEG when PSNR for JPEG2000 is about 2 dB larger than for JPEG. Note that most often JPEG and JPEG2000 are compared in terms of PSNR and, based on this fact, the conclusion that JPEG2000 is better is drawn. The mentioned difference in visual quality of images compressed by JPEG and JPEG2000 deals with the fact that standard JPEG uses non-uniform quantization of DCT coefficients (i.e. takes HVS into account) whilst the conventional JPEG2000 coder Kakadu (used in TID2008 distorted image creation) uses uniform quantization of wavelet coefficients. Since these coders are commonly used just for compressing images subject to visual analysis, it is not correct to employ PSNR for comparison of their performance. We recommend to use the following metrics: PSNR-HVS-M, WSNR, NQM, VSNR, or MSSIM.

Fig. 2 addresses another important application – filtering of noisy images. The plots of MOS vs PSNR for different noise types and for output images obtained for the case of additive i.i.d. noise in the input images and applying 3D DCT filter [24, 13] are presented there.

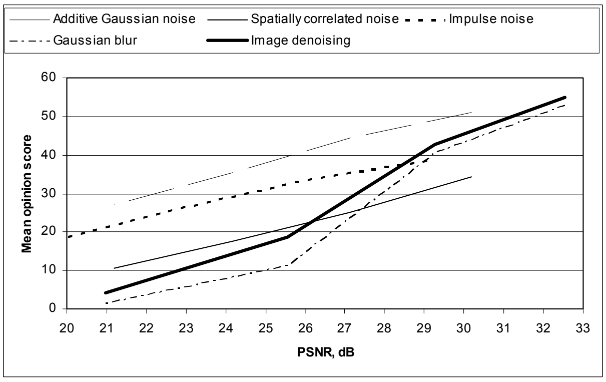

Fig. 2. MOS vs PSNR for different types of distortions

It is important to understand two aspects here. First, one would like to know what kind of distortions visually correspond to "good" filtering results. Second, it is desirable to answer the question what increase in PSNR (provided due to filtering) leads to actual improvement of image visual quality.

It is interesting to observe that the curves for filtered (denoised) images and gaussian blur behave similarly although, for the same PSNR, MOS values for the filtered images are slightly larger than for blurred ones. This shows two things. First, filtering often results in similar effects as blurring (people who deal with filtering and blurring know this well). Second, such filters can be considered good that provide smaller distortions in regions where blur is the most visible, i.e. in the neighborhoods of sharp transitions in images.

Concerning the second question, a good filter should provide PSNR increase of, at least, 2 dB or larger (if PSNR for noisy image is over 30 dB) and of, at least, 6 dB (if PSNR of an original noisy image is about 21 dB). This is an important conclusion that allows stating that filters leading to PSNR increase less than 2 dB cannot be efficient in the sense of image visual quality increase.

In general, similarly to the case of analyzing image compression application, for studying efficiency of filtering it is reasonable to apply such adequate metrics as, for example, PSNR-HVS-M.

5. CONCLUSIONS

In this paper, we have compared the two largest existing databases of distorted images and carried out performance analysis for many known image visual quality metrics for the color image database TID2008. It is demonstrated that for the considered wide range of possible distortion types no existing metric performs well enough. In aggregate, the best results are provided by MSSIM. Analysis has been also performed for particular subsets of distortion types. This analysis has shown that for most typical practical applications like image filtering and compression the metrics PSNR-HVS and PSNR-HVS-M produce reasonably good results.

REFERENCES

[1] Pratt W.K. *Digital Image Processing (4th Edition)*, Wiley-Interscience, - 2007. - 782 p.

[2] Keelan B.W., *Handbook of Image Quality*, Marcel Dekker, Inc.: New York, USA, 2002.

[3] Sheikh H. R., Sabir M. F., Bovik A. C. “A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms”, *IEEE Transactions on Image Proc.*, vol. 15, no. 11, 2006, pp. 3441-3452.

[4] Ponomarenko N., Carli M., Lukin V., Egiazarian K., Astola J., Battisti F. “Color Image Database for Evaluation of Image Quality Metrics, *Proc. of Intern. Workshop on Multimedia Signal Processing*, Australia, Oct. 2008, pp. 403-408.

[5] TID2008 page: http://www.ponomarenko.info/tid2008.htm

[6] Gaubatz M. "Metrix MUX Visual Quality Assessment Package": http://foulard.ece.cornell.edu/gaubatz/metrix_mux/

[7] Chandler D.M., Hemami S.S. "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images", *IEEE Transactions on Image Proc.*, Vol. 16 (9), 2007, pp. 2284-2298.

[8] Mitsa T., Varkur K. "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms", *ICASSP '93-V*, pp. 301-304.

[9] Wang Z., Simoncelli E. P., Bovik A. C. "Multi-scale structural similarity for image quality assessment," *IEEE Asilomar Conf. on Signals, Systems and Computers*, 2003.

[10] Damera-Venkata N., Kite T., Geisler W., Evans B., Bovik A. "Image Quality Assessment Based on a Degradation Model", *IEEE Trans. on Image Processing*, Vol. 9, 2000, pp. 636-650.

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

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

[13] Vansteenkiste E., Van der Weken D., Philips W., Kerre E. “Psycho-visual quality assessment of state-of-the-art denoising schemes”, *CD-ROM proceedings of EUSIPCO*, 2006, 5 p.

[14] Vansteenkiste E. “Quantitative Analysis of Ultrasound Images of the Preterm Brain*”, Thesis for the degree of Doctor of Technology*, Universiteit Gent, Belgium, 2007, 224 p.

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

[16] Ponomarenko N., Silvestri F., Egiazarian K., Carli M., Astola J., Lukin V. "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.

[17] PSNR-HVS-M page: http://ponomarenko.info/psnrhvsm.htm

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

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

[20] Kolpatzik B. W., Bouman C. A. "Optimized Universal Color Palette Design for Error Diffusion", *Journal Electronic Imaging*, vol. 4, 1995, pp. 131-143.

[21] Kolpatzik B.and Bouman C. "Optimized Error Diffusion for High Quality Image Display", Journal Electronic Imaging, pp. 277-292, 1992.

[22] Watson A. B. "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.

[23] De Simone F., Ticca D., Dufaux F., Ansorge M., Ebrahimi T. “A Comparative study of color image compression standards using perceptually driven quality metrics”, *Proc. of SPIE Conf. Applications of Digital Image Proc. XXXI*, San Diego. 2008.

[24] Dabov K., Foi A., Katkovnik V., Egiazarian K. “Image denoising by sparse 3-D transform-domain collaborative filtering”, *IEEE Transactions On Image Processing*, vol. 16, issue 8, 2007, pp. 2080-2095.