Color Image Database for Evaluation of Image Quality Metrics

N. Ponomarenko[#1], V. Lukin[#2], K. Egiazarian[*3], *Senior Member, IEEE*, J. Astola[*4], *Fellow IEEE*, M. Carli[+5], *Senior Member, IEEE* and F. Battisti[+6]

[#]*Dept of Transmitters, Receivers and Signal Processing, National Aerospace University, Kharkov, Ukraine*

[1]`uagames@mail.ru`

[2]`lukin@xai.kharkov.ua`

**Institute of Signal Processing, Tampere University of Technology, Tampere, Finland*

[3]`karen@cs.tut.fi`

[4]`jta@cs.tut.fi`

[+]*Dept of Applied Electronics, University of Rome "Roma TRE", Rome, Italy*

[5]`carli@uniroma3.it`

[6]`fbattisti@uniroma3.it`

***Abstract*— In this contribution, a new image database for testing full-reference image quality assessment metrics is presented. It is based on 1700 test images (25 reference images, 17 types of distortions for each reference image, 4 levels for each type of distortion). Using this image database, 654 observers from three different countries (Finland, Italy, and Ukraine) have carried out about 400000 individual human quality judgments (more than 200 judgments for each distorted image). The obtained mean opinion scores for the considered images can be used for evaluating the performances of visual quality metrics as well as for comparison and for the design of new metrics. The database, with testing results, is freely available.**

***Index Terms*—Visual quality metrics, HVS, test image databases**

I. Introduction

Quality evaluation of digital images is critical in all applications of image processing. Each stage of processing, storing, compression, and enhancement, may introduce perceivable distortions [1][2]. The visibility and annoyance of these impairments are directly related to the quality of the received/processed data. It is fundamental the possibility of measuring the overall perceived quality to maintain, control, or to enhance the quality of the digital data.

Many efforts have been directed during the last two decades by the scientific community to the design of quality metrics. The choice of an adequate metric usually depends on the requirements of the considered application. They can be distinguished in objective and subjective metrics. In objective measurements of the performances of an imaging system, image quality and quality losses are determined by evaluating some parameters based on a given general mathematical, physical or psycho-psychological model. While in subjective tests, the digital image quality is determined from the performance of test-persons in subjective psychological tests.

Objective quality metrics can be classified according to the amount of side information required to compute a given quality measurement. Using this criterion, three generic classes of objective metrics can be classified as Full Reference (FR) when the original and the impaired data are available, Reduced Reference (RR) when some side information regarding the original media can be used, and No-Reference (NR) if only the impaired image is available. The most used FR objective metrics are the Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR). They have low computational cost, physical meanings, and are mathematically easy to deal with for optimization purposes. However, they have been widely criticized for not being well correlated with perceived quality measurement [3][4].

To overcome such problems, recently HVS inspired objective quality metrics have been introduced. The main difference among these metrics and the mathematical ones (MSE, PSNR) is that they are more heuristic. It is more difficult to perform a mathematical comparison of their performances. Thus, to adequately evaluate the quality of such metrics statistical experiments are needed [8][9]. To this purpose, a large database of distorted test images is usually prepared, and the Mean Opinion Score (MOS) from a large number of human observers is collected. Then, the subjective results are compared with the objective scores of the tested metrics to identify the metric more tuned to the subjective scores. However some drawbacks are to be considered: usually the size of the test database is not big enough [10], the number of different distortions is limited [11][12], and methodological errors in planning and execution of the experiments can occur.

Since in most applications humans are the ultimate receivers of the digital data, the most accurate way to determine its quality is to measure it directly using psychophysical experiments with human subjects. Unfortunately, these subjective tests are too expensive and time-consuming.

One of the most intensive studies in this field has been carried out by the Video Quality Expert Group (VQEG)[13]. New metrics of image visual quality have been designed and the largest database of distorted test images (982 distorted images: 29 reference images, 5 types of distortions, 5-7 distortion levels) has been created, LIVE database [14]. However, to our opinion, the LIVE database, as well as the other ones, does not allow to adequately evaluate metrics of image visual quality. This is due to the limited number of the modeled types of distortions, for LIVE namely the ones induced by JPEG and JPEG2000 compression, arising from transmission errors for JPEG2000, modeled by white noise and Gaussian blur. Among them only the distortion caused by JPEG compression allows the evaluation of the correspondence of the tested metrics to one feature of HVS. In [9], the conclusion (based on using LIVE database) is that VIF [15] is the best among tested metrics. At the same time, based on data of subjective experiments carried out in [10] (in which the authors took into account peculiarities of HVS as CSF and contrast masking), the values of Spearman and Kendall correlations for VIF and MOS are 0.377 and 0.255, respectively. Meanwhile, for the best metrics considered the correlations are equal to 0.984 and 0.948, respectively. This means that LIVE database has not allowed to emphasize the poor accounting of CSF and of contrast masking by VIF.

In the design of the proposed database, we have tried to overcome the drawbacks of existing test image databases while maintaining their positive aspects. Briefly, this image database contains 17 types of distortions related to the most important currently known peculiarities of HVS and valuable (wide-spread) practical situations of image processing.

For assessing the human perceived quality, a huge number of participants (654) has been enrolled in the performed experiments for providing reliability of the obtained MOS estimates. Using this database, it is possible to more accurately evaluate the correspondence of metrics to the human visual judgment. We also have analyzed some results, in particular, human perception of noisy and filtered images.

The paper is organized as follows. The requirements to image databases used for testing full-reference quality metrics are considered in Section II. Section III is devoted to the description of the proposed image database. In Section IV we present the performed experiments. Analysis of the obtained results is also given. Finally, Section V conclusions are drawn.

II. Requirements To Image Databases Used For Testing Full-Reference Quality Metrics

The image databases to be used for the considered application have to satisfy several requirements as reflecting the HVS peculiarities and containing non-trivial images. Based on previous experience, we can summarize some guideline for designing a test image database:

- it should include images with considerably different characteristics: percentage of homogeneous regions, details and textures, various texture characteristics, etc.;
- for each HVS feature, the database has to contain, at least, one distortion type that allows to estimate how this feature influences image visual quality;
- it is desirable that the database will contain image distortions typical for practice that originate due to compression, denoising, data transmission errors, etc;
- the images in the database should not be too simple for visual quality estimation: 1) the number of distortion levels should not be large, 2) the number of situations when all metrics evidence in favor of a given image should not be large. Fig. 1 shows three undesirable situations in testing. Quality of the image in Fig. 1 (a) is worse than one in Fig. 1 (b). Then, a majority of tested quality metrics will indicate preference of the latter image quality. The presence of relatively large number of such pairs of compared image combinations in database might result in overestimated effectiveness of all considered metrics. Images represented in Figures 1 (c) and 1 (d) relate to the same type of distortions (impulse noise). The image in Fig. 1 (c) is characterized by a sufficiently higher level of distortions. In this case, most metrics will evidence in favor of the better quality of the image in Fig. 1 (d). In general, this is the correct decision and the property to clearly "recognize" such simple situations has to be provided for quality metrics. However a large number of such combinations leads to increasing correlation of the analyzed metric and MOS. It is also undesirable to use many images in a database for which distortions are unperceived (Figures 1 (e) and 1 (f)).

III. Description Of The Proposed Image Database

The proposed test database (TID2008) contains color images with different textural characteristics, various percentages of homogeneous regions, edges and details. The images are from the Kodak test set [16] that can be considered as a good trade off between the abovementioned requirements. Besides, we have synthesized and added one artificial image that has different texture fragments and objects with various characteristics. All images are of size 512x384 pixels. Table I presents the distortions modeled in our image database.

As can be noticed several distortions have been considered. For example, masked noise and high frequency noises are types of distortions that allow analyzing metrics' adequateness with respect to local contrast sensitivity and spatial frequency sensitivity of HVS. Such artifacts are typically introduced by lossy image compression or digital watermarking [17][18]. Other important type of distortions studied recently [20] are residual distortions resulting from denoising. It is a common result that the PSNR for a filtered image is by 2-3 dB better than an original (noisy) one, but, at the same time, visually a processed image looks worse than the corresponding noisy original. Thus, we have included into our database images for which original additive i.i.d. Gaussian noise is suppressed by one of the state-of-the-art filter [21]. In Fig. 2, an example of the original and of its version corrupted by Gaussian additive noise is reported. As it can be seen, although the processed image is characterized by a larger PSNR, residual noise after filtering and distortions introduced by filtering leads to perceivable artifacts.

a) c) e)

b) d) f)

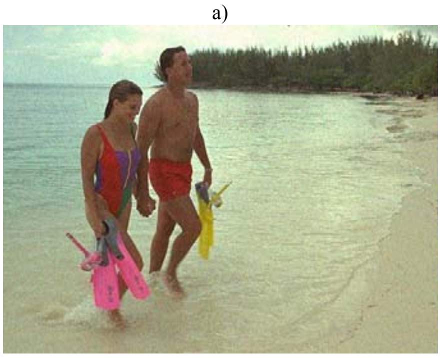
Fig 1. Three examples of too simple cases of visual quality assessment that have to be met quite seldom

TABLE I
TYPES OF DISTORTIONS USED IN OUR IMAGE DATABASE

№	Type of distortion (four levels for each distortion)	Correspondence to practical situation	Accounted HVS peculiarities
1	Additive Gaussian noise	Image acquisition	Adaptivity, robustness
2	Additive noise in color components	Image acquisition	Color sensitivity
3	Spatially correlated noise	Digital photography	Spatial frequency sensitivity
4	Masked noise	Image compression, watermarking	Local contrast sensitivity
5	High frequency noise	Image compression, watermarking	Spatial frequency sensitivity
6	Impulse noise	Image acquisition	Robustness
7	Quantization noise	Image registration, gamma correction	Color, local contrast, spatial frequency
8	Gaussian blur	Image registration	Spatial frequency sensitivity
9	Image denoising	Image denoising	Spatial frequency, local contrast
10	JPEG compression	JPEG compression	Color, spatial frequency sensitivity
11	JPEG2000 compression	JPEG2000 compression	Spatial frequency sensitivity
12	JPEG transmission errors	Data transmission	Eccentricity
13	JPEG2000 transm. errors	Data transmission	Eccentricity
14	Non eccentricity pattern noise	Image compression, watermarking	Eccentricity
15	Local block-wise distortions of different intensity	Image acquisition, inpainting	Evenness of distortions
16	Mean shift (intensity shift)	Image acquisition	Light level sensitivity
17	Contrast change	Image acquisition, gamma correction	Light level, local contrast sensitivity

Another distortion we considered is caused by compression and transmission oven noisy packet channels. We have included into our database the images compressed by JPEG or JPEG2000 and decoded with errors in data transmission channels. Quite often it is not easy to notice distortions induced by such errors since they are almost not seen (visible) due to their non-eccentricity. Fig. 3 presents two examples of distortions due to transmission/decoding errors. Distorted fragments might occur to be similar to original texture and/or color of surrounding fragments and due to peculiarities of HVS a human might not notice (pay attention to) such distortions. To our opinion, the use of images for which the considered distortions are modeled will allow to get some imagination concerning ability of the tested quality metrics to take this feature of HVS into account. Another distortion we introduced in the test set is the so called local block-wise distortions of different intensity. An idea that we would like to verify consists in the following. We suppose that in case of compact impulse-like distortions HVS reacts not to distortion values pixel by pixel but to area (percentage of pixels) that is a subject to (occupied by) distortions. Distortions have been modeled in such a way that blocks of size 32x32 pixels that have arbitrary random color have been placed in an image randomly but mainly in places where there is important information (content). For the first level of distortions, 16 blocks with color slightly differing from mean color of replaced fragment have been added. For the second level of distortions, the amount of such blocks was 8 but their color differs from mean color of replaced fragment more. For the third and fourth levels, four and two blocks have been replaced, respectively. However, for these blocks their color differs more essentially from the mean colors of the corresponding replaced fragments. Trials show that the image corrupted by two blocks is perceived as having better visual quality (although it has smaller PSNR) than the image distorted by 16 blocks. Finally, we have added into our database images for which mean shift and contrast change distortions have been modeled [7].

a)



b)

Fig. 2. Comparison of visual quality a) after filtering out additive noise, PSNR=28.19 dB, b) original noisy image corrupted by additive i.i.d. Gaussian noise, PSNR=26.99 dB

a)

b)

Fig. 3. Image decoded with errors due to unreliable data transmission line: a) for the standard JPEG, PSNR=24.05 dB, b) for the standard JPEG2000, PSNR=23.98 dB

We have set four levels for all types of distortions. For almost all types of distortions, the corresponding levels of PSNR are of about 30 dB, 27 dB, 24 dB, and 21 dB (very good quality, good quality, poor quality, and bad quality). On one hand, such number of distortion levels for 25 reference images allows to “reliably cover” all range of subjective quality of distorted images from “excellent” to “very bad”. On the other hand, four levels do not create too many simple combinations of image pairs at their quality comparison stage (see Section II). Table II gives some additional details concerning generation of distorted images for all types. Note that all color images have been represented as RGB with 8 bits in each component.

TABLE 2

SOME DETAILS OF DISTORTED IMAGE GENERATION

№	Type of distortion	Four levels of distortions
1	Additive Gaussian noise	Variance=64, 130, 260, 525
2	Different additive noise in color components	PSNR=30 dB, 27 dB, 24 dB, 21 dB
3	Spatially correlated noise	Variance=64, 130, 260, 525
4	Masked noise	PSNR=30 dB, 27 dB, 24 dB, 21 dB
5	High frequency noise	PSNR=30 dB, 25 dB, 20 dB, 15 dB
6	Impulse noise	Pimp=0.85%, 1.7%, 3.4%, 6.8%
7	Quantization noise	QS (quantization step)=27, 39, 55, 76
8	Gaussian blur	Window size is 11, Sigma (parameter of blur) = 0.65, 1, 1.7, 4
9	Image denoising	Variance of additive noise before filtering = 144, 484, 1764, 8100
10	JPEG compression	quality levels with parameters of compression equal to 60, 23, 8, 4 (100 - max quality, 0 - min quality)
11	JPEG2000 compression	PSNR=30 dB, 27 dB, 24 dB, 21 dB
12	JPEG transmission errors	PSNR=30 dB, 27 dB, 24 dB, 21 dB
13	JPEG2000 transmission errors	PSNR=30 dB, 27 dB, 24 dB, 21 dB
14	Non eccentricity pattern noise	PSNR=30 dB, 27 dB, 24 dB, 21 dB
15	Local block-wise distortions of different intensity	16, 8, 4 and 2 blocks
16	Mean shift (intensity shift)	Value of the shift is +10,-20,+30,-40
17	Contrast change	x1.2, x0.75, x1.45, x0.5

IV. EXPERIMENTS DESCRIPTION

There are different methodologies that can be used to evaluate the quality of an image [9][22]. In [23] we tested the following approach to evaluate the image visual quality. The basic idea is that is easier for a subject to select the image of higher quality between two than to rank one image with a scale. The experiment performed is organized in two phases. In the first one the observers have been asked to sort the images from the test set according to their visual quality through a pair-wise comparisons. For each couple of distorted images the observer had to decide which one, between the two, was less distorted when compared to the original one. Then, a quantitative evaluation of the image quality is determined by its position in the obtained ordered set. In the second phase of the experiment, to produce the MOS, the observers were asked to rate the annoyance of possible impairments in the test set using a continuous scale [0,100], where ‘100’ corresponds to the case where no distortions are detected (i.e. the highest quality) and ‘0’ to the case where very annoying impairments are present. At the end of the experiments, the subjects have marked that it was simpler the first stage of quality evaluation although for sorting the test set they had to perform more judgments. Moreover, the Pearson correlation between the MOS obtained from the quantitative evaluation of image quality (second stage in the experiments), and the MOS derived by averaging the image positions in the sorted samples (first stage in the experiments) was equal to 0.99. This means that both approaches have led to the same results.

A conventional way to measure the correspondence between the HVS and the visual quality is to find the correlation between these metric values calculated for the considered test image database and the MOS values obtained for the same database. Different correlations can be used: the standard Pearson correlation and rank correlations proposed by Spearman and Kendall [24]. In our tests we have used the Spearman and Kendall correlation. In fact the use of Pearson correlation requires a preliminary data fitting [9] whilst the rank correlations of Spearman and Kendall can be derived without such pre-processing.

According to [22], the execution time of one experiment by each observer should not exceed 30 minutes. In our case database of test images contains 1700 images. The full sorting of this image database will require about 1700 x $\log_2$ (1700) decisions for each observer. Supposing that each comparison takes approximately 2-3 seconds, the total time for each subject experiment will vary from 10 up to 15 hours. This is unrealistic. Thus, experiments for each reference image have been performed separately. The averaged time needed by each subject for one reference image is 13.5 minutes.

Totally 654 experiments have been carried out in three countries: Finland, Italy, and Ukraine. 251 experiments have been carried out in Finland, 150 in Italy, and 253 in Ukraine. In Italy and Ukraine the experiments have been performed in off-line mode; in Finland, in on-line mode via Internet.

Our experiments have been performed on LCD and TFT monitors with screen sizes 17 or 19 inches. The monitor brightness, illumination and distance from the observer varied in wide limits. The only fixed parameter in our experiments was the monitor resolution, 1152x864 pixels.

It is useful to note that the designed database is intended for verification of visual quality metrics in a priori unknown conditions. Each experiment consisted of 9 cycles. During the first cycle all the 68 distorted images of a given set that correspond to a selected reference image have been randomly divided into 34 pairs. The “winners” of each pair (an image that has better visual quality according to the observer opinion) got one point, the “losers” got no points. In each cycle those images have been randomly combined to pairs that had equal or almost equal number of points. Thus, each image had a chance to be compared to any other image in the set, but images that have approximately equal quality have been compared more frequently to provide high quality ordering.

Each observer during one experiment has carried out 306 comparisons of image visual quality (612 evaluations of relative visual quality of distorted images). The observers have performed experiments on a variable number of reference images (1-3). Totally, 654 observers have performed 200124 comparisons of visual quality of distorted images or 400248 evaluations of relative visual quality in image pairs. Each image from the total set of 1700 distorted images finally got, on the average, 235 estimates of relative visual quality. As the result, the quantitative estimates of image visual quality have been obtained (the numbers of got points) as well as rank estimates of quality (average place in the ordered sample).

Spearman correlation between the MOS for both variants is 0.998 whilst Kendall correlation is equal to 0.961. These values evidence high statistical confidence of the obtained MOS. Another evidence of reliability of the MOS is the high correlation present in the data independently obtained in the three different countries where the experiments have been carried out (0.93-0.96).

The validity of the subjective test results was verified by a screening of the results performed according to Annex 2 [22].

The average MOS for all the 25 reference images is given in Fig. 4. In each section, corresponding to one distortion, there are four dots that mark the MOS for the four distortion levels; the leftmost dot corresponds to the first level of distortions. This plot allows drawing some interesting conclusions. First, none of the considered distortions produces for all levels a worse image visual quality than for other types of distortions. Similarly, there is not a distortion providing for all levels a better visual quality than the one obtained for other types of distortions. This shows that the database has been properly designed. Third, we would like to highlight the interesting phenomenon that has been also emphasized in [15]. Among images with modified contrast the highest visual quality is assigned to images with slightly enhanced (about 1.2 times) contrast. This confirms the hypothesis that subjectively such images are perceived by observers as even having higher quality than the corresponding reference images. At the same time, the over-contrasted images (with contrast enhanced by 1.45 times) occur to be visually perceived as having worse quality than images subject to some other type of distortions.

It is also interesting to note that distortions induced by filtering occur to be very similar to distortions due to JPEG image compression. The range of visual quality of images for these two distortions is the widest. We would like to highlight the dependences of the visual quality on the PSNR for distortions of the 1st (Additive Gaussian noise), 9th (DCT 3D denoising), 10th (JPEG), and 11th (JPEG2000) types.

These dependences are shown in Fig. 5. These plots prove one more time that the PSNR is not suitable to characterize the visual quality of distorted images, especially filtered ones. If the PSNR after filtering increases by 1-2 dB, the visual quality of the filtered image can results even worse. To be sure that the filtering operation leads to an improvement in the image visual quality, its PSNR should increase by, at least, 2.5-3 dB if the PSNR of the noisy image was of about 29-30 dB, or by, at least, 7-8 dB if noisy image PSNR has been about 20-22 dB. Another comment by the analysis of Fig. 5 is that for images presenting a similar PSNR, the visual quality of images compressed by using the JPEG standard is preferable in comparison to images compressed by using the standard JPEG2000 (as in LIVE, we used the freely available coder

Kakadu [6]). JPEG produces worse visual quality of images than JPEG2000 for PSNR around 22-23 dB [5].

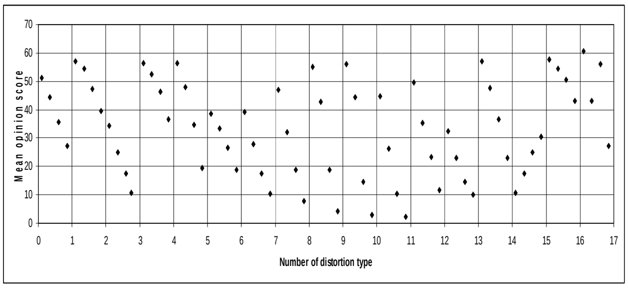


Fig. 4. MOS averaged for all reference images

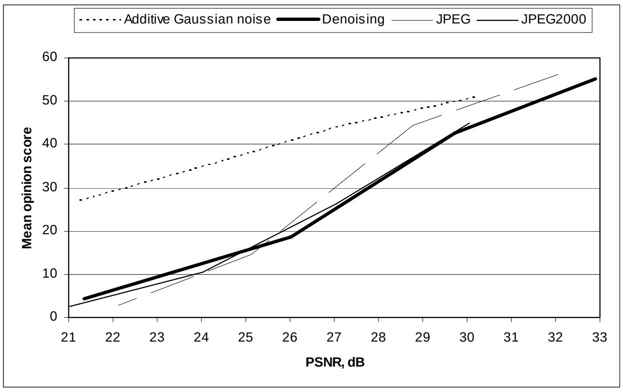


Fig 5. Dependence of the opinion score on PSNR for some types of distortions

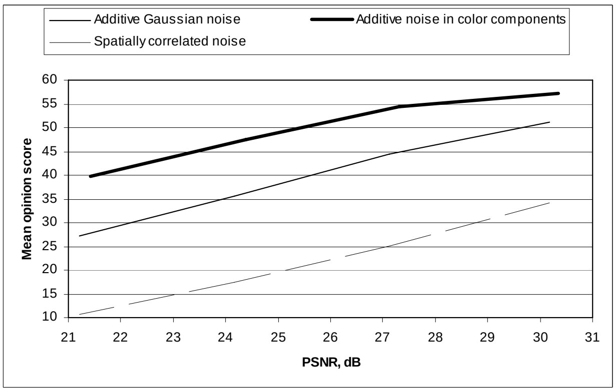


Fig. 6. Dependences of the MOS on PSNR for different spatial frequency and color distributions of additive noise

The obtained results confirm a typical phenomenon for HVS: higher sensitivity to spatially correlated noise and less sensitivity to noise in color components (see plots in Fig. 6). As seen, images corrupted by spatially correlated noise have worse visual quality than those corrupted by i.i.d. Gaussian noise that show a smaller PSNR (6 dB less).

V. ACCESS TO TID2008, CONCLUSIONS

For obtaining TID2008, please send your request by e-mail Karen.egiazarian@tut.fi. This archive includes image files, the file containing the MOS values, the program for calculation of Spearman and Kendall correlations, the readme file that explains how to exploit the database. TID2008 is free of charge for usage with scientific purposes. In case of publishing results obtained by means of TID2008, it is necessary to refer to this paper. Finally, we would like to stress once more the main advantage of TID2008: it satisfies main requirements for image quality testing, containing many different types of distortion related to various peculiarities of HVS.

REFERENCES

[1] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, vol. 70, pp. 247-78, Oct. 1998.

[2] *ITU Recommendation P.930*, "Principles of a reference impairment system for video," 1996.

[3] A. B. Watson, Ed., *Digital Images and Human Vision*, Cambridge, MA: MIT Press, 1993, 224p.

[4] Z. Wang, A. C. Bovik, and L. Lu, “Why is image quality assessment so difficult,” *in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 4, Orlando, FL, pp. 3313–3316, May 2002.

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

[6] D. Taubman, M. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*. Boston: Kluwer, 2002.

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

[8] I. Avcibas¸ B. Sankur, and K. Sayood, “Statistical evaluation of image quality measures,” *J. Electron. Imag.*, vol.11, no.2, pp.206–23, 2002.

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

[10] 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 International Workshop on Video Processing and Quality Metrics*. - USA, 2007. - 4 p.

[11] R. de Freitas Zampolo, R. Seara, “A Comparison of Image Quality Metric Performances Under Practical Conditions”, *Proc. of IEEE International Conf. on Image Processing*, 2005, vol. 3, pp. 1192-1195.

[12] M. Miyahara, K. Kotani, V.R. Algazi, ”Objective picture quality scale (PQS) for image coding”, *IEEE Transactions on Communications*, vol. 46, issue 9, pp. 1215-1226, 1998.

[13] VQEG: The Video Quality Experts Group, http://www.vqeg.org/

[14] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, LIVE Image Quality Assessment Database, Rel. 2 2005 http://live.ece.utexas.edu/research/quality

[15] Sheikh H. R. and Bovik A. C., "Image Information and Visual Quality", IEEE Trans. on Image Processing, vol. 15, February, 2006, pp. 430-444

[16] Kodak Lossless True Color Image Suite. http://r0k.us/graphics/kodak/

[17] I.J. Cox, M.L. Miller, J.A. Bloom, J. Fridrich and T. Kalker, *Digital Watermarking and Steganography* (Second Edition), Morgan Kaufmann, 2008, 619 p.

[18] M. Carli, “Perceptual Aspects in Data Hiding”, Thesis for the degree of Doctor of Technology, Tampere University of Technology, 2008.

[19] K.N. Plataniotis and A.N. Venetsanopoulos, *Color Image Processing and Applications, Springer Verlag*, Berlin, 3-540-66953-1, May 2000.

[20] E. Vansteenkiste, D. Van der Weken, W. Philips, E. Kerre, “Psycho-visual quality assessment of state-of-the-art denoising schemes”, *Proceedings of EISIPCO*, 2006, 5 p.

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

[22] *ITU (2002). Methodology for Subjective Assessment of the Quality of Television Pictures Recommendation BT.500-11*, International Telecommunication Union, Geneva, Switzerland.

[23] 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.

[24] W.H. Kruskal, “Ordinal measures of association”, *J. Amer. Statist. Assoc.*, vol. 53, pp. 814–861, 1958.