Software for automated quality assessment of digitally processed video

Rafał Kłoda[1], Paweł Oczkowski[2], Sabina Żebrowska-Łucyk [3], Anna Ostaszewska-Liżewska[4]

[1] *Warsaw University of Technology, 02-525 Warszawa, ul. św. A. Boboli 8,*
phone: + 48 22 234 8645, fax: +48 22 849 03 95, kloda@mchtr.pw.edu.pl
[2] *Samsung Electronics Poland Sp. z o.o., 02-674 Warszawa, ul. Marynarska 15, p.oczkowski@samsung.pl*
[3] *Warsaw University of Technology, 02-525 Warszawa, ul. św. A. Boboli 8,*
phone: + 48 22 234 8323, fax: +48 22 849 03 95, szl@mchtr.pw.edu.pl
[4] *Warsaw University of Technology, 02-525 Warszawa, ul. św. A. Boboli 8,*
phone: + 48 22 234 8645, fax: +48 22 849 03 95, a.ostaszewska@mchtr.pw.edu.pl

***Abstract*-** The paper presents a novel software tool, which calculates and analyses a large number of metrics applied by different scientists and organizations in order to assess quality of compressed video. Metrics computed for a wide range of diversified video sequences may be used for mapping of objectively measured features into subjective quality space. Moreover, statistical analysis of various metrics would help to uncover relationships that otherwise are not available. This, in turn, may lead to work out new objective measures of video quality that will emulate subjective rating more aptly than currently applied ones.

I. Introduction

Every day millions of people around the world watch videos that come from various sources like digital television, Internet, DVD-Video and Blu-ray discs, computer games and so on. Each video material is exposed to complex digital processing technology. Digital image processing may introduce visible picture quality deterioration. Degree of distortion depends on many factors, such as a video coding format, a type of coder, a method of conversion between standards, scaling, available data transmission bandwidth, parameters of lossy compression, content of the video (dynamics, number of details, contrast, colour), etc.

As the customer satisfaction depends on the technical quality of the video watched, the need for monitoring this aspect becomes significantly important. Therefore many organizations and research institutions are involved in objective video quality metrics development. There are also various dedicated hardware solutions available, known as digital video quality analyzers. The most advanced of them emulate (predict) mean opinion score (MOS) – the score that would be given by the human audience. However, empirical studies conducted with human observers showed that the numerical solutions are far from being satisfactory [4].

The result of many years of theoretical and empirical investigations carried out independently in several research centres in the world is the invention of a substantial set of video quality metrics. Some of them are recommended by international and national standardization organizations. The essential role in this field is played by the International Telecommunication Union (ITU) and the American National Standards Institute (ANSI).

Till now, there was no tool that could be used for computing a large set of video quality metrics, including even those less common, described in scientific papers, but yet hardly used.

Information relating to metrics proposed by various authors, however, is very scattered, often incomplete and sometimes inconsistent, and this is not conducive to their dissemination and application. Probably this is the reason why computer scientists, who work on improvement of standards and video coding algorithms, use easy to implement metrics like PSNR or MSE, although it is well known that there is rather weak correlation between their values and subjective quality evaluation ratings.

These circumstances led researchers from Institute of Metrology and Biomedical Engineering, Warsaw University of Technology, who are also involved in subjective video quality evaluation [1, 2, 3, 4], to develop and implement software, which can calculate a wide set of objective quality metrics at a time.

The use of such tool in autonomous research centers would enable to gather many observations and opinions on sensitivity of individual metrics to specific video artefacts and help to assess the suitability of hitherto known metrics for practice. Calculation of the metrics for a wide range of video sequences (of a different content and methods of coding) would enable examining the correlation of each of the metrics with subjective quality evaluation ratings, as well as to reveal the mutual correlation between individual metrics. Such a data base would be used for objective metrics improvement or would enable to develop completely new ones that perform even better.

II. Objective quality evaluation metrics implemented in the software

Extensive review of the literature on compressed video quality evaluation, as well as the international recommendations, preceded the software implementation. A set of metrics selected for implementation in software cover a wide range of ideas for objective quality evaluation (Table 1). Among them there are both metrics which use the model of human visual system (HVS) like SSIM [10], MSSSIM [11], PSNR-HVS-M [14] and the simplest ones, which can even be calculated in real time such as PSNR or MSE. Among selected metrics there are also others based on uncomplicated algorithms like MAE [7] or AD [8] and metrics for detecting some specific types of impairments, like blurring (BLUR [9]) or blocking artifacts (BLOCK-V, according to the concept of Venkatesh Babu [12] and BLOCK-W of Wang and Bovik [13]). Other metrics describe the video content: its temporal (TI[Fn]) and spatial features (SI[Fn]) [6].
The set of chosen metrics can also be divided into those that use the source material to compare with the processed video (FR, full-reference) and metrics that do not use the original (NR, no-reference).

Tab. 1. The list of metrics implemented in the software along with their classification and the algorithm source

No	metric	NR/FR	HVS	references
1	MSE	FR		Skarbek W. [5]
2	PSNR	FR		Skarbek W. [5]
3	TI[Fn]	NR		Wolf S. [6]
4	SI[Fn]	NR		Wolf S. [6]
5	MAE	FR		Sasi varnan C. [7]
6	AD	FR		Eskicioglu A. M. [8]
7	BLUR	NR		Crete F. [9]
8	SSIM	FR	✓	Wang Z. [10]
9	MSSSIM	FR	✓	Wang Z. [11]
10	BLOCK-V	NR	✓	Venkatesh Babu R. [12]
11	BLOCK-W	NR	✓	Wang Z. [13]
12	PSNR-HVS-M	FR	✓	Ponomarenko [14]

III. VImaQ software

The software, which has been named VImaQ, was implemented in LabVIEW with the use of Vision Development Module – especially designed for machine vision applications.
To analyze video, VImaQ requires a sequence of separate pictures (single frames) saved as bitmaps (BMP) with filenames, which denote number of consecutive video frames. The main advantage of this operation is that it enables quick access to each individual frame for preview and analysis. Besides, it makes possible to analyze video sequences regardless of their resolution and a type of encoder used.
Analysis of material can be performed with all the implemented metrics at once (Table 1), or just some of them, according to user choice (Fig. 1).

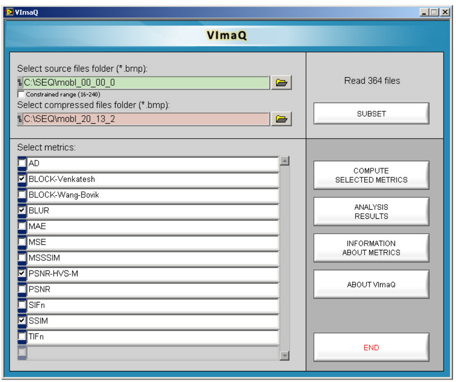

Figure 1. The main window of VImaQ

In case of computing full-reference metrics, it is necessary to specify a sequence of source pictures.

Default analysis use the whole contents of the folder (the whole video split into frames), although there is an option of choosing any part of the video and any set of metrics to be calculated.
The final effect of calculation is the plot which presents waveforms of metrics (Fig. 2).

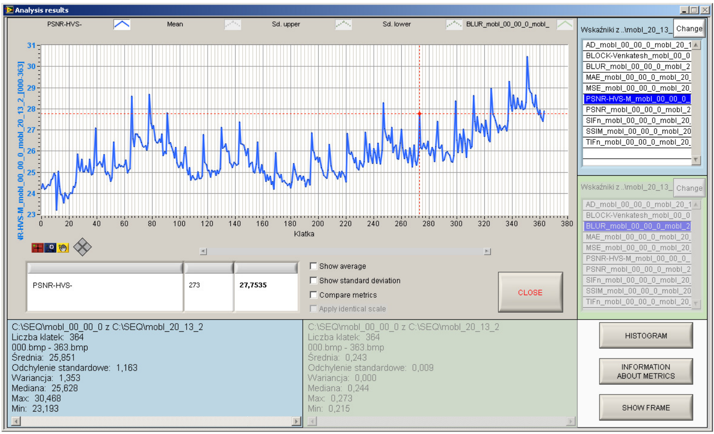

Figure 2. The window with the final result of calculation

Moreover, temporal values of metrics calculated for individual frames are subjected to statistical analysis. Measures of central tendency and dispersion of the measurement data distributions are calculated automatically. They are used to compare metrics sensitivity to the video content and the influence of encoding parameters on the perceived quality of the video.
VImaQ also provides the ability to browse frames for which the metrics were computed (Fig. 3). The preview window also contains a scroll bar and the plot of the actual value of the chosen metric.

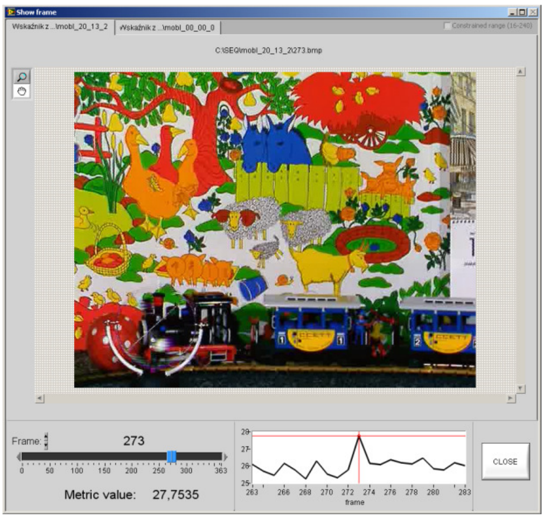

Figure 3. The frame preview window of VImaQ

The final test report contains, in addition to detailed information on each individual frame, the summary of the examination results. The whole analysis is saved in text files on a hard drive of PC used. The file name is a kind of a code, which allows for unambiguous identification of the material that has been tested.
The software has been created in such a way that each metric calculation algorithm is dynamically joined to the main program as a separate module. The advantage of this solution is that the upgrade of new metrics does not require modification of other interacting subprograms. The new module is placed in the specified directory in the application folder structure.
In order to reduce computation time, the application architecture includes support for multi-core processors. This enables more efficient use of computing power, which is particularly useful in case of calculation some complex metrics simultaneously.

IV. Video material analysis

To test the software, a set of video sequences was prepared – they covered a wide variety of encoding parameters for MPEG-2 DVD-Video standard. Afterwards studies were carried out to verify the results obtained with the implemented software.

For the purpose of analysis some video materials were coded in MPEG-2 with 10 bitrate values from 2 to 5 Mbps, which is a typical range for this standard. As the video test sequence, we used the sequence ‘mobl’ (Mobile & Calendar) – a scene with a train and a calendar in the background, because it has high sensitivity to the loss of picture quality caused by reducing the bitrate. The Fig. 4 presents example frame encoded with two bit streams 2 Mb/s and 5 Mb/s.

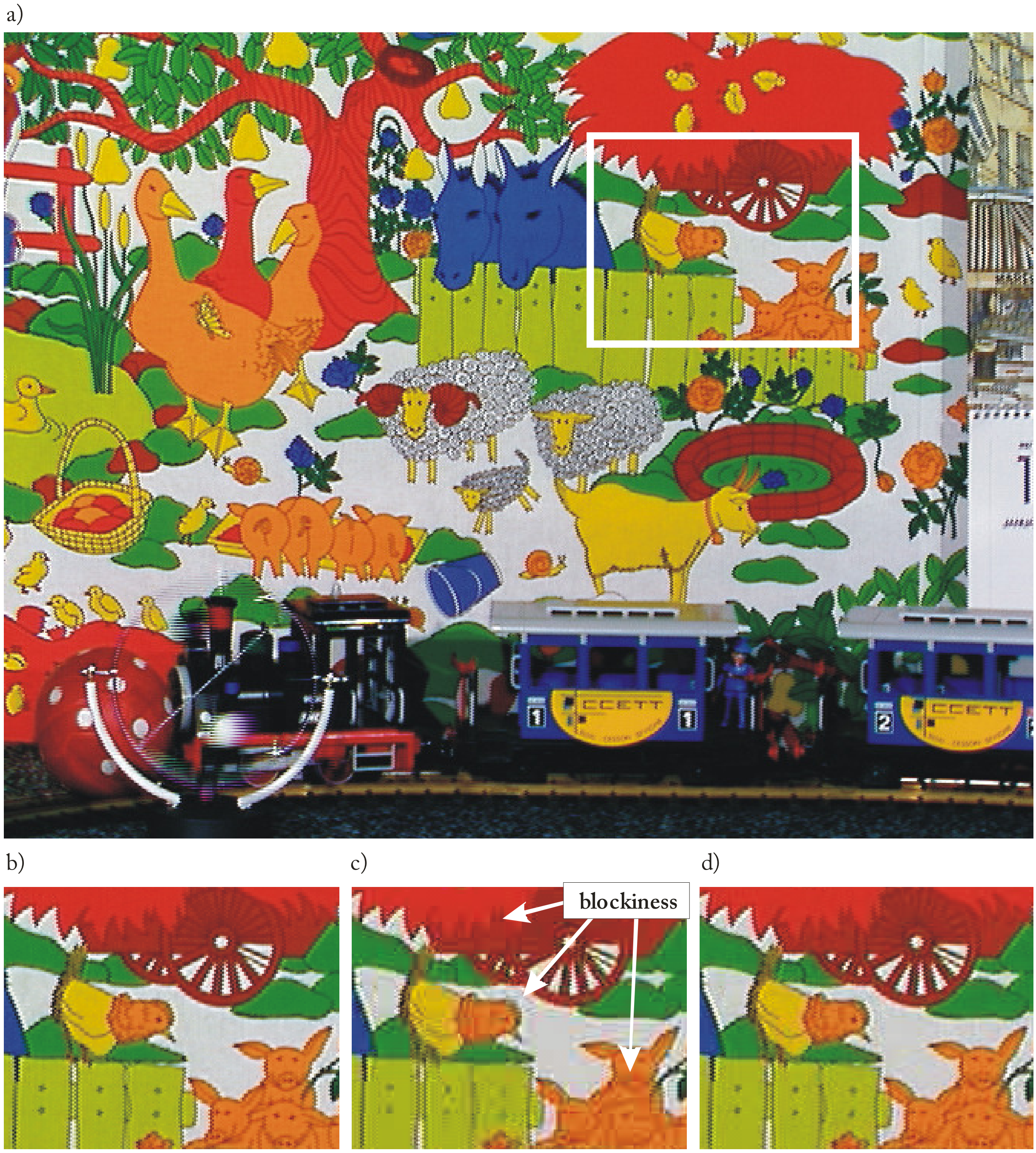

Figure 4. Frame number 273: a) reference frame with marked region (white rectangle) to show in detailed compression artefacts, b) magnified region of reference frame, c) magnified region of frame encoded at $S = 2$ Mb/s, d) magnified region of frame encoded at $S = 5$ Mb/s

As a part of tests, the plot of metric for blockiness artifacts was computed for compressed video (Fig. 5). A metric of Venkatesh Babu was used. This metric integrates several key human visual sensitivity factors such as edge amplitude, edge length, background activity and background luminance to evaluate the effect of block edge impairment on perceived picture quality [12].

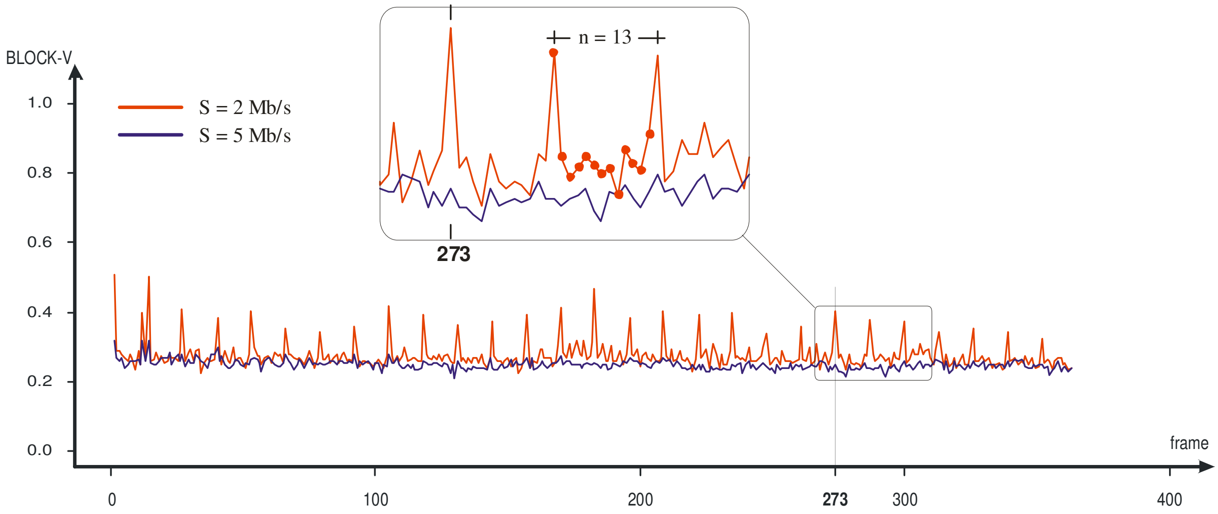

Figure 5. BLOCK-V (blockiness metric) computed for 'mobl' sequence encoded at two bitrates

For this example small peaks may be observed for the whole plot (Fig. 5). They are strictly recurring, as they are directly related to encoding parameters: the bit rate stream S and the Group of Pictures structure (GOP), which is specified for the whole encoding process. In case of the analyzed sequence GOP consists of n = 13 frames with m = 2 B frames (IBBPBBPBBPBBP). The rapid growth of blocking structure is connected with the low bit stream and the beginning of the consecutive GOP structure with I frame, which contains the most information. When the bit stream is increased to 5 Mbps, the mentioned peaks may not be observed (Fig. 5). Similar phenomenon may be observed in case of other metrics, whilst for some of them it does not occur (Fig. 6). In this case we compared two image fidelity metrics: the most popular PSNR and its modification PSNR-HVS-M which takes into account the model of visual inter-coefficient contrast masking of DCT basis functions based on a Human Visual System (HVS) and the contrast sensitivity function [17].

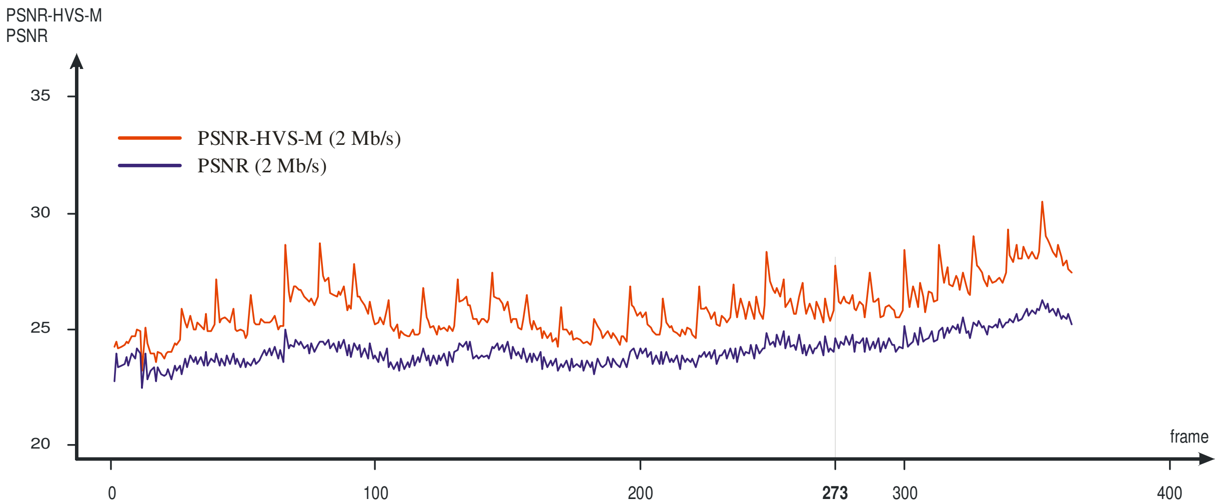

Figure 5. PSNR and PSNR-HVS-M metrics computed for 'mobl' sequence encoded at 2 Mb/s

Such analysis might be done for a set of sequences to select metrics which correlate well with subjective ratings in order to work out a new mathematical model for objective video quality assessment.

V. Summary

Developed at the Institute of Metrology and Biomedical Engineering, Warsaw University of Technology software called VImaQ allows user to calculate twelve parameters developed in various research centers to estimate the degree of distortion of the video (e.g. BLUR, BLOCK-V, BLOCK-W, SSIM) or describing the character of the sequence in terms of susceptibility to distortion (e.g. TI, SI).

The software enables examination encoding (i.e. the type of encoder and the encoding parameters) influence on all implemented metrics just in a single pass.

The results obtained can be used to determine the correlation of the metrics with the results of relevant subjective tests. The application may also be used in the process of improvement or development of new objective visual quality metrics.

Developed software has already been successfully used for educational purposes. At the Faculty of Mechatronics, Warsaw University of Technology, it is used by students of Multimedia Techniques specialization to examine the influence of video encoding parameters on image quality. The possibility of observing temporal changes of metrics in parallel with a frame preview, where some specific impairment may occur, is a big benefit and makes the VImaQ favorite tool both of students and teachers. Authors believe that scientists will appreciate it as well.

References

[1] A. Ostaszewska, R. Kłoda, "Quantifying the amount of spatial and temporal information in video test sequences", *Recent Advances in Mechatronics*, Springer Verlag, pp. 11-15, 2007.

[2] R. Kłoda, S. Żebrowska-Łucyk, A. Ostaszewska, "A dedicated station for MPEG-2 quality evaluation with ACR method", *Fifth International PhD Conference on Mechanical Engineering*, 2007.

[3] P. Oczkowski, „The development of software for video/image quality metrics estimation" (in polish), MSc Thesis under the supervision of S. Żebrowska-Łucyk and R. Kłoda, Warsaw University of Technology, Faculty of Mechatronics, 2011.

[4] A. Ostaszewska-Liżewska, „Examination of subjective continuous quality evaluation of compressed video methods for the purpose of quality monitoring automatization" (in polish). PhD Thesis, Warsaw University of Technology, Faculty of Mechatronics, 2012.

[5] W. Skarbek, *Multimedia - Algorithms and standards of compression (in polish)*, Akademicka Oficyna Wydawnicza PLJ, 1998.

[6] S. Wolf, *Features for automated quality assessment of digitally transmitted video*, Report 90-264, National Telecommunication and Information Administration, 1990.

[7] C. Sasi varnan, A. Jagan, J. Kaur, D. Jyoti, D.S. Rao, "Image Quality Assessment Techniques in Spatial Domain", I*nternational Journal of Computer Science & Technology*, vol. 2, issue 3, 2011.

[8] A.M. Eskicioglu, P.S. Fisher, "Image quality measures and their performance", *IEEE Transactions on Communicatio*n, vol. 43, No. 12, pp. 2959-2965, 1995.

[9] F. Crete, T. Dolmiere, P. Ladret, M. Nicolas, "The blur effect: perception and estimation with a new no-reference perceptual blur metric", *Human Vision and Electronic Imaging* XII, vol. 6492 of Proc. of SPIE, 2007.

[10] Z. Wang, A.C. Bovik, H.R. Sheikh., E.P. Simoncelli, "Image quality assessment: From error measurement to structural similarity", *IEEE Trans. Image Processing,* vol. 13, no. 4, 2004.

[11] Z. Wang, E.P. Simoncelli, A.C. Bovik, "Multi-scale structural similarity for image quality assessment", *IEEE Asilomar Conference on Signals*, Systems and Computers, vol.2, pp. 1398-1402, 2003.

[12] R. Venkatesh Babu., A.S. Bopardikar, A. Perkis, "A Perceptual No-Reference Blockiness Metric for JPEG Images", *Indian Conference on Computer Vision, Graphics and Image Processing*, 2004.

[13] Z. Wang Z., A.C. Bovik, B.L. Evans, "Blind measurement of blocking artifacts in images", *Proc .of International Conference on Image Processing*, vol. 3, pp. 981-984, 2000.

[14] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, V. Lukin: "On between-coefficient contrast masking of DCT basis functions", CD-ROM *Proc. of the Third International Workshop on Video Processing and Quality Metrics*, 2007.