Performance Assessment of Metrics for Video Quality Estimation

Zoran G. Kotevski[1] and Pece J. Mitrevski[2]

Abstract – **There are two types of metrics for measuring the quality of processed digital video: purely mathematically defined (DELTA, MSAD, MSE, SNR and PSNR) where the error is mathematically calculated as a difference between the original and processed pixel, and video quality metrics that have similar characteristics as the Human Visual System (SSIM, NQI, VQM), where the perceptual quality is also considered in the overall video quality estimation. In this paper, experimental comparison of the performance of PSNR, SSIM, NQI and VQM metrics is presented.**

Keywords – **Video quality, PSNR, SSIM, NQI, VQM**

I. Introduction

One of the simplest definitions of video quality is that the video quality is a state of perception by the Human Visual System (HVS) [1]. Hence, the best video quality estimation can be performed by trained human estimators, but in real world situations this is a huge problem and these video quality metrics are practical tool for fulfilling this complex task [9]. Basically there are two types of video quality metrics for measuring the quality of processed digital video. On one side there are purely mathematically defined video quality metrics like: DELTA, MSAD, MSE, SNR and PSNR [6], [11], [14], [15], where the error is mathematically calculated as a difference between the original and processed pixel. These metrics are more technical ones and because visual quality assessment is a task more complex than simple pixel error calculation, many consider their quality estimation to be deficiently accurate. On the other side there are video quality metrics that have similar characteristics as the Human Visual System like: SSIM, NQI, and VQM [3], [4], [5], where the perceptual quality is considered in the overall video quality estimation. This second group of metrics, beside the mathematical error calculation, also calculates the scene structure in the estimation of quality. The SSIM and NQI are known to have similar characteristics, because they calculate the quality by combining three components: luminance, contrast and structure or mutual characteristics. Differently from these two, VQM metrics uses DCT quantization to eliminate the spatial frequencies that are less visible to the human eye. In order to compare the performance of these four metrics, an experiment is conducted in which large number of differently processed video sequences are created and their quality is measured. The results are basic charts that present these metrics dependence to the most common changes in processed video i.e. changes in brightness, contrast, hue, saturation and noise. This paper pinpoints the key characteristics of each metric, gives the conclusion of the better performing one and gives some directions for improvement of objective video quality estimation.

[1]Zoran G. Kotevski is with the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mail: zoran.kotevski@uklo.edu.mk

[2]Pece J. Mitrevski is with the Faculty of Technical Sciences, I.L.Ribar bb, 7000 Bitola, Macedonia, E-mail: pece.mitrevski@uklo.edu.mk

II. Brief Metrics Introduction

A. Peak Signal to Noise Ratio (PSNR)

The PSNR parameter is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. PSNR is usually expressed in terms of the logarithmic decibel scale. Normally, higher PSNR indicates that the reconstruction is of higher quality. In ideal case the value of PSNR would be 100 dB, but in reality, in the field of image processing, typical values for PSNR are between 30 dB and 40 dB [9], [11].

$$PSNR = 10 \cdot \log_{10}\left(\frac{255^2}{MSE}\right) \text{ [dB]} \quad (1)$$

and
$$MSE = \frac{1}{mn}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}\left[f(x,y) - g(x,y)\right]^2, \quad (2)$$

According to the mathematical equations for calculating MSAD, MSE, SNR and PSNR, [6], [14], [15], can be inferred that they represent similar error values i.e. the calculated error is of the same degree. Because of this, PSNR can be considered as an unofficial representative of all the above mentioned video quality metrics and still the most widely used metric for video quality estimation in many video processing systems, especially in video compression systems.

B. Structural Similarity (SSIM)

The Human Visual system (HVS) is highly adapted to extracting the structural information from the area of viewing. SSIM metric uses this characteristic of the HVS in estimation of quality of the processed digital video [4].

Structural information of an image can be defined by those characteristics that represent the structure of the objects in the scene, independently of the mean brightness and contrast [3], [4]. These measurements are based on measurement of three components: luminance comparison, contrast comparison and structure comparison.

$$S(x, y) = f(l(x, y), c(x, y), s(x, y)) \quad (3)$$

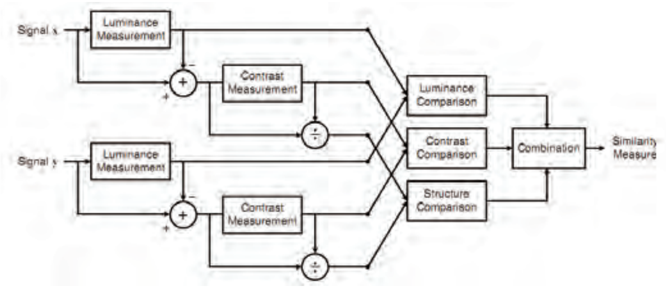

Fig. 1. Diagram of the SSIM measurement system

Structural similarity index is a combination of these separate components (3). The SSIM index can gain values from 0 to 1 where value of 1 represents maximum quality.

C. New Quality Index (NQI)

NQI works in a similar manner as SSIM index. NQI defines picture distortion as a combination of three factors: difference in mutual characteristics, difference in luminance and difference in contrast. Mathematical definition of the NQI can be found at [3].

The values of NQI span from 0 to 1, even though in some cases NQI can be lower than zero. Similar like SSIM the value of 1 represents maximum quality.

D. Video Quality Metrics (VQM)

Human eye sensitivity to spatial-temporal pattern decreases with high spatial and temporal frequency. Based on different sensitivity, high spatial or temporal information can be represented with less data and less precision, while human eyes are more or less insensitive to the loss of this information. This characteristic of HVS is exploited by DCT quantization, which is the base for VQM [5]. The values of VQM start from 0 and in real situations can reach around 12. VQM value of 0 represents minimum distortion and maximum quality. The system diagram of the VQM system is shown in figure 2.

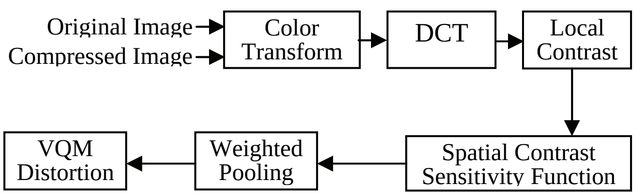

Fig. 2. Diagram of the VQM measurement system

III. Analysis of the Measurements of Modified Video Sequences

For the purpose of this analysis, at first three video sequences were created. In all three videos (Old boat, Sea view and Mountains), static picture in duration of five second is presented. All these three video sequences were made with different scene structure in order to determine if scene structure has some influence in these measurements.

Also for the purpose of creating the diagrams of metrics dependence to the changes in processed video sequences, more than 300 short video sequences were produced, each with different amount of introduced changes and effects, in order to illustrate the influence of more or less visible video deformation to the performance of these metrics. The most common changes that do not highly influence the viewer's quality of experience are changes in brightness, contrast, and saturation. In other video sequences, highly destructive video deformation like Gaussian noise is introduced.

All video sequences were produced with trial version of Sony Vegas Pro v8.0c [16], coded in Main Concept's MPEG-2 coder, main level and profile, with average bit rate of 4MBit/sec. Measurements were performed with the trial version of Elecard Stream Eye Tools v2.9.1 [13].

After the performed measurements, the charts of metrics dependence to different changes were drawn and some of them are presented below. These charts were visually evaluated by dozen estimators and the comments are summarization of their opinion.

Because of the limitations of this paper only the most relevant charts are presented. All the charts and images of differently processed video sequences are publicly available at http://vq.heliohost.org.

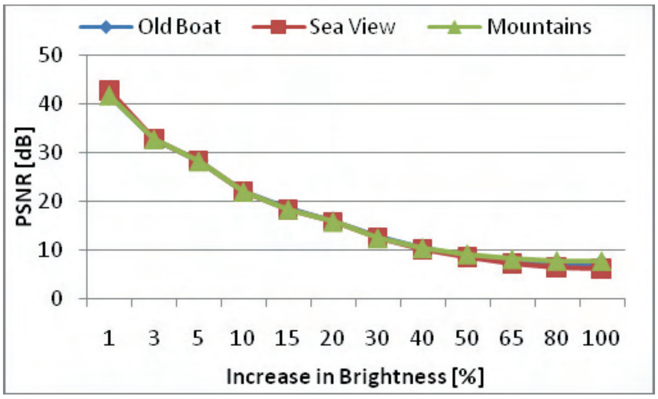

Fig. 3. PSNR decrease due to changes in brightness.

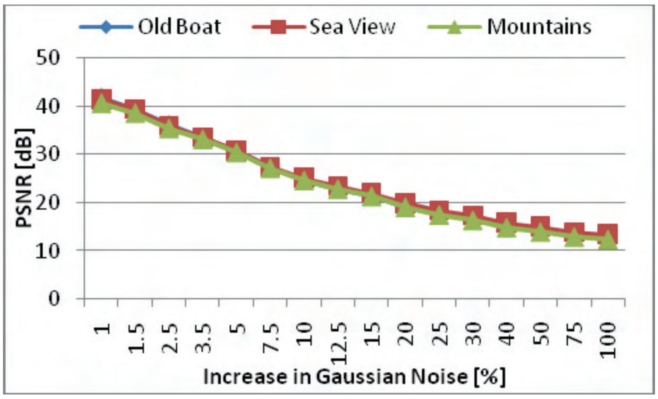

Fig. 4. PSNR decrease due to the amount of Gaussian noise.

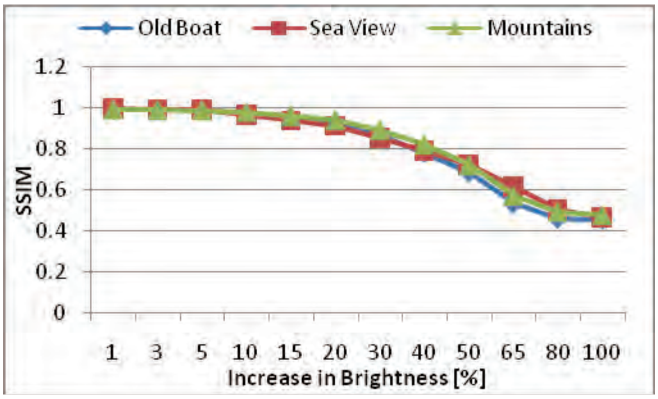

Fig. 5. SSIM decrease due to changes in brightness.

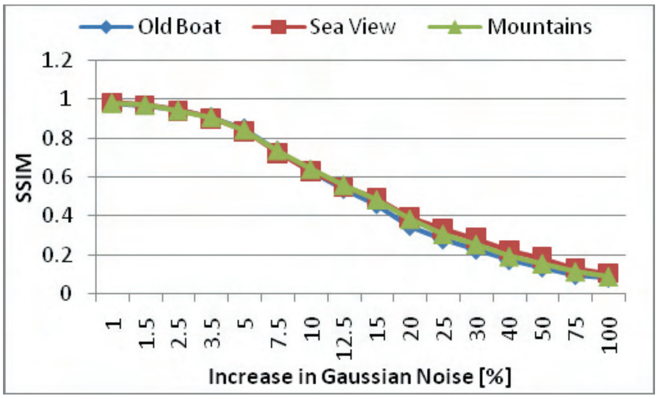

Fig. 6. SSIM decrease due to the amount of Gaussian noise.

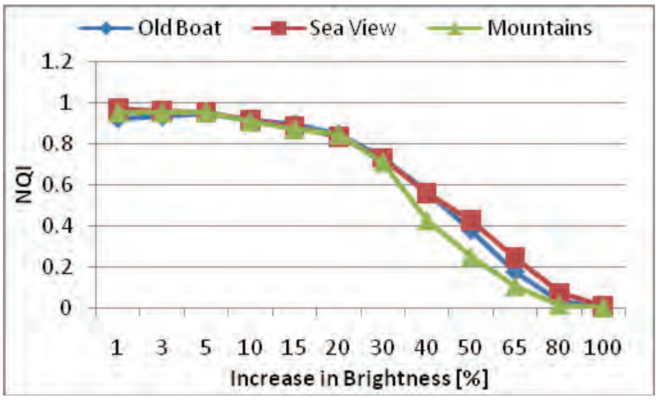

Fig. 7. NQI decrease due to changes in brightness.

After performed analysis of these charts, it can be concluded that the most drastic decreases in PSNR values are due to changes in brightness, as shown in Fig. 3, and combination of changes in brightness and contrast. The second influential factor is the introduced Gaussian noise, as shown in Fig. 4. The changes in hue and saturation have medium effect in decreasing PSNR value. These charts clearly describe the deficiencies of PSNR.

Considering the performance of SSIM, the most drastic decreases in SSIM values are due to the amount of introduced Gaussian noise (Fig. 6). Changes in brightness (Fig. 5), contrast and hue have only mild influence to the overall quality estimation. These characteristics of SSIM speak of certain similarities to the HVS and present solid background for more realistic platform for quality estimation of processed digital video, but it is also obvious that there are some imperfections present.

NQI index reacts most drastically to changes in Hue and increase in Gaussian Noise as shown in Fig. 8. Changes in brightness as shown in Fig. 7, or saturation have even less influence to the NQI index compared to the SSIM index. It is easily noticable that NQI reacts less than SSIM to changes in brightness or saturation, but reacts more drastically to changes in hue or Gaussian noise. These characteristic of NQI clearly speak of some deficiencies that NQI has compared to SSIM index.

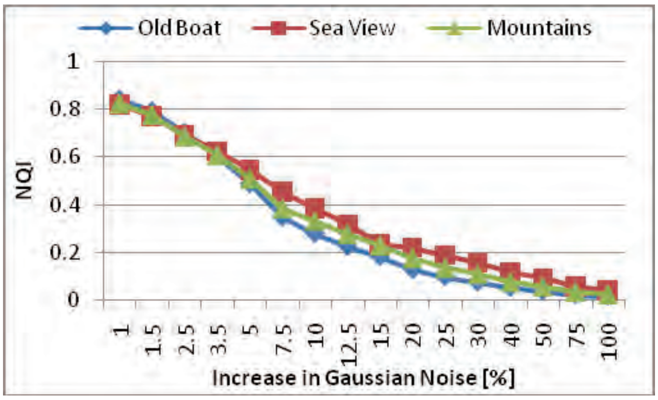

Fig. 8. NQI decrease due to the amount of Gaussian noise.

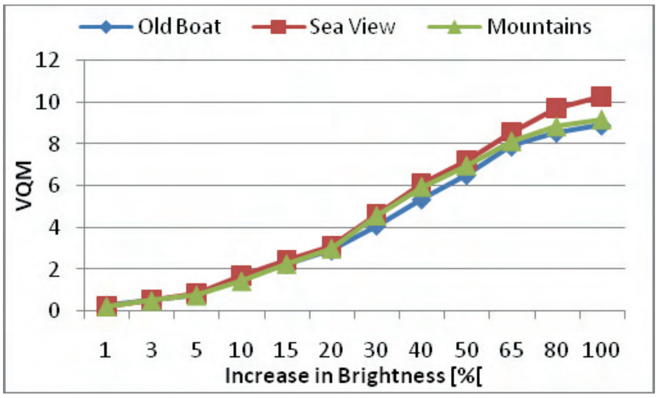

Fig. 9. VQM increase due to changes in brightness

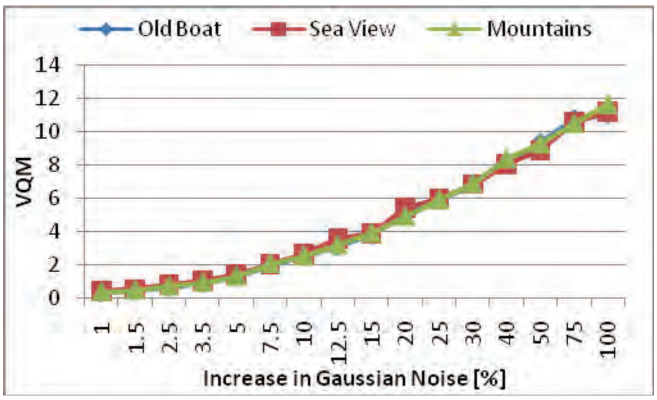

Fig. 10. VQM increase due to the amount of Gaussian noise

VQM index reacts the most to changes in the amount of Gaussian noise as shown in Fig. 10. But, the key concluding elements from these charts are that VQM is not so insensitive to changes in brightness (Fig. 9) or changes in contrast, and opposite to SSIM and NQI shows quite good sensitivity to changes in hue and saturation. SSIM's and NQI's issues of low sensitivity to changes in brightness/contrast and hue contribute to lower performance in quality estimation compared to VQM metric. This advantage of VQM over all other metrics can be observed in most of the examples.

IV. Conclusion

Given the examples, it can easily be concluded that PSNR metric is not valid enough to be used as objective measurement for video quality estimation. There are too many parameters that highly influence the PSNR value that are of minor visual influence to the viewer's perception of quality. Changes in brightness and contrast have high influence to the PSNR that in most cases causes decrease to the performance of this metric. To be more precise, PSNR metric can be taken as valid measurement in some cases if PSNR value is greater than 35 dB. Everything below this degree of PSNR cannot be considered valid because the origin of PSNR decrease is unknown in most cases and the results given by this metric can be misleading.

SSIM metric has quite better performance compared to PSNR and in most cases performs very similar to the Human Visual System. But, imperfections are also present. SSIM is almost insensitive to changes in brightness, contrast and hue that when these changes are bigger SSIM values can become largely inverted. However, many examples indicate good SSIM similarity to HVS and with some small improvements in mentioned areas SSIM performance can be enhanced.

If we compare SSIM to NQI, even though they calculate the quality in a similar manner, they differ in some points quite significantly. From all the examples can be concluded that SSIM metric performs quite better than NQI in most of the situations. Through the analysis of the presented charts it can be concluded that changes in brightness have mild influence to the SSIM index and even milder to the NQI index. Changes in Hue cause steep decrease in NQI but changes in saturation are almost invisible to it. These characteristics of NQI contribute to lower performance compared to SSIM and VQM.

VQM metric on the other hand performs quite better than all previous mentioned metrics in almost all situations. Through the analysis of the presented charts it can be concluded that VQM metric reacts to changes in brightness in a similar manner as HVS which can be considered as VQM's advantage. Changes in hue and saturation are rated in a good manner by VQM, compared to the SSIM and NQI where they are almost invisible to them. The analyzed charts and videos lead to a conclusion of certain imperfections of VQM metric also, but it can be concluded that VQM performs quite well. The overall conclusion is that VQM metric gives best performance of all of these analyzed metrics. The advantage that VQM has is that it exploits the DCT transformation and quantization technique, similar like coding techniques, where high spatial frequencies are omitted as being less visible to the human eye. This characteristic enables VQM to mostly consider the changes that are more noticeable to the human eye, which is the key to creating a better video compression system as well as creating better video quality estimation system.

If better video quality estimation metric is to be created, one must explore HVS behavior first. Drawing similar charts of HVS dependence to changes in brightness, contrast, hue, saturation and noise would be a challenging task, but such charts would be a great foundation for creating video quality metric that would resemble the HVS perception of quality.

Concerning the scene structure and its influence to these measurements it can be concluded that scene composition barely influences these measurements and can be considered as non influential factor.

V. References

[1] Z. Kotevski, *Analysis of Quality and Performance of MPEG-2 Video Compression Techniques*, Master Thesis, Faculty of technical sciences, Bitola, Macedonia, December 2007.

[2] Z. Kotevski, P. Mitrevski, "Experimental Comparison of PSNR and SSIM Metrics for Video Quality Estimation", Proc. of ICT Innovations-09 International Conference, Macedonia, September 2009.

[3] Z. Wang, Student Member, IEEE, A. C. Bovik, Fellow, IEEE, "A Universal Image Quality Index," IEEE Signal Processing Letters, March, 2002.

[4] Z. Wang, Member, IEEE, A.C. Bovik, Fellow, IEEE, H. R. Sheikh, Student Member, IEEE, E.P. Simoncelli, Senior Member, IEEE, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Transactions on Image Processing, April, 2004.

[5] F. Xiao, "DCT Based Video Quality Evaluation," Final Project for EE392J, 2000.

[6] Z. Wang, A.C. Bovik, "Mean Squared Error: Love It or Leave It?," IEEE Signal Processing Magazine, January, 2009.

[7] Z. Wang, X. Shang, "Spatial Pooling Strategies for Perceptual Image Quality Assesment," IEEE International Conference on Image Processing, Atlanta GA, October, 2006.

[8] Z. Wang, Q. Li, "Video Quality Assesment Using a Statistical Model of Human Visual Speed Perception," Journal of the Optical Society of America, 2007.

[9] P. Symes, *Digital Video Compression*, McGraw-Hill, ISBN 0-07-142487-3, 2004.

[10] C. A. Poynton, *A Technical Introduction to Digital Video*, John Wiley & Sons, ISBN 0-471-12253-X, 1996.

[11] A. Bovik, *Handbook of Image and Video Processing*, Academic Press, ISBN 0-12-119790-5, 2000.

[12] C. A. Poynton, *Digital Video and HDTV – Algorithms and Interfaces*, Morgan Kaufmann Publishers, ISBN 1-55860-792-7, 2003.

[13] Elecard, http://www.elecard.com/

[14] Compression Project, http://www.compression.ru/

[15] Compression Links Info, http://www.compression-links.info/

[16] Sony Creative Software, http://www.sonycreativesoftware.com/