# Performance Comparison of Video Quality Metrics

Kotevski Zoran, Mitrevski Pece

Faculty of Technical Sciences, University of "St. Kliment Ohridski", Bitola, R. Macedonia,
zoran.kotevski@uklo.edu.mk, pece.mitrevski@uklo.edu.mk

## ABSTRACT

The development of digital video technology, due to its nature, introduced new approach to the objective video quality estimation. Basically there are two types of metrics for measuring the quality of digital video: purely mathematically defined video quality metrics (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 quality estimation. The metrics from the first group are more technical ones and because the visual quality of perception is more complex than pixel error calculation, many examples show that their video quality estimation is deficiently accurate. The second group of metrics work in a different manner compared to previous, calculating the scene structure in the overall video quality estimation. This paper is concerned with experimental comparison of the performance of Structural Similarity (SSIM) and Video Quality Metric (VQM) metrics for objective video quality estimation. For the purpose of this experiment, more than 300 short video sequences were prepared. The measurements of these video sequences are used to draw the metrics dependence to common changes in processed video sequences. These changes include 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 directions for improvement of objective video quality estimation.

**Keyword list**: Video quality; video coding; PSNR; SSIM; NQI; VQM

## 1. 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]. This means that the best way to estimate the video quality is if it is performed by trained human estimators. But, here comes the problem of availability and affordability. In real world situations, everyday availability of larger number of estimators is a huge problem and these video quality metrics are practical tool for fulfilling this complex task [9]. Many of the present video quality metrics perform in their own manner and their objective quality estimation sometimes differs significantly.

There are two basic types of objective video quality metrics for measuring the quality of processed digital video. The purely mathematically defined video quality metrics like: DELTA, MSAD, MSE, SNR and PSNR [6], [11], [14], [15] belong to the first older type of metrics that express the quality statistically. They estimate the quality by mathematical error calculation as a difference between the original and processed pixel. The HVS works in a more complex manner, and many examples and experiments show that these metrics are not able to offer the needed service level of quality measures [2]. In more recent years, newer video quality metrics are developed, metrics that represent a new type of video quality measurements, like: SSIM, NQI, and VQM [3], [4], [5]. Based on previous definition that video quality is a state of perception by the human visual system, these metrics are constructed to work in a more similar manner to the HVS. They all use some of the characteristics of the HVS performance and their mutual characteristic is that they consider the perceptual quality in the overall video quality estimation. One of the metrics that is used more often everyday is Structural Similarity (SSIM). Compared to PSNR metric [2], SSIM performs video quality estimation much better, but there are also some known issues. Video Quality Metric (VQM), compared to SSIM works in a different manner, but it also exploits some characteristics and functions of the HVS in its objective quality assessment. This paper is concerned with introduction to the basics of the SSIM and VQM calculations and experimental performance comparison of these two video quality metrics.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75460G · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853442

## 2. STRUCTURAL SIMILARITY (SSIM)

The Human Visual system (HVS) is highly adapted to extracting the structural information from the area of viewing. This characteristic of the Human Visual System gives solid information that video quality metric based on extracting the structural information can provide better estimation of quality of the processed digital video [4], [8] in comparison to pure mathematical error calculation metrics like PSNR, even though PSNR is still the most widely used video quality metric, especially for measuring the quality of lossy video compression coders [10], [12].

The luminance of an object that is being observed is a result of the reflected light that hits its surface. Depending on the amount of light that hits the observed object, it can appear brighter or darker, but the structure of the object is totally independent to changes in luminosity. These changes in brightness and contrast are high influential factor to the PSNR and other similar metrics [2] and make video quality estimation deficiently accurate. Because of this, to explore the structural information of an image the influence of the luminosity should be extracted.

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. Structural similarity index is a combination of these three separate components.

$$S(x,y) = f(l(x,y), c(x,y), s(x,y)) \tag{1}$$

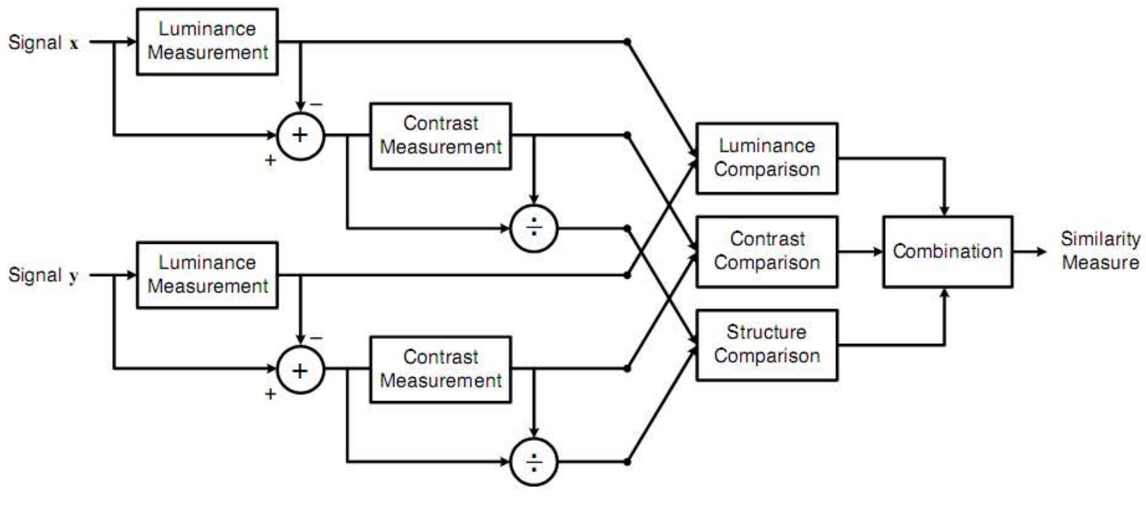


Figure 1. Diagram of the SSIM measurement system

The system diagram of the structural quality assessment system is shown in figure 1. More details about the mathematical equations for calculating the SSIM index can be found at [4]. The SSIM index can gain values from 0 to 1 where value of 1 represents maximum quality.

## 3. VIDEO QUALITY METRIC (VQM)

As a base for all video quality metrics that work similar to the HVS is Human spatial-temporal sensitivity to contrast. 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 [5]. DCT quantization is used in all major image and video coding standards, like (JPEG, MPEG1, MPEG2…) and it is also a basis for VQM. 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.

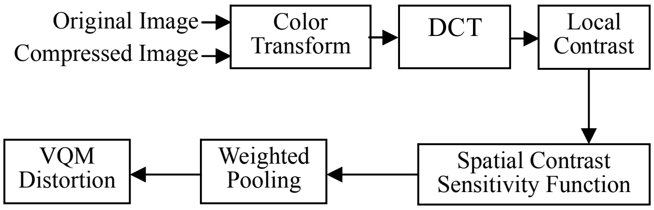

Figure 2. Diagram of the VQM measurement system

## 4. ANALYSIS OF THE SSIM AND VQM MEASUREMENTS

For the purpose of this analysis, at first three video sequences were created. In all three video sequences, static picture in duration of five second is presented. The first video is called *Old boat* with an old boat placed by a rock. The second video is called *Sea view* and has beautiful sea site. The third video is named *Mountains* with a landscape of mountains and sky. The reason that three video sequences were created is 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 SSIM and VQM measurements. 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 created 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].

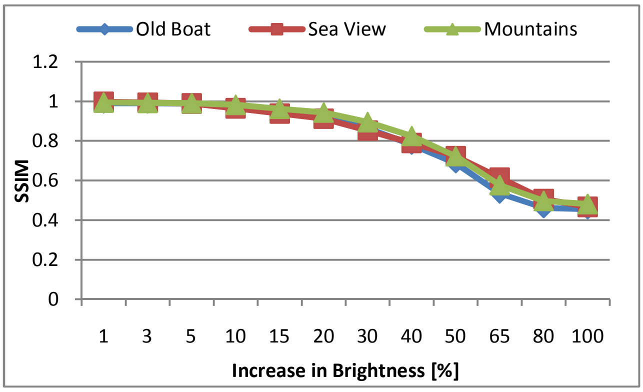

Figure 3. SSIM decrease due to changes in brightness

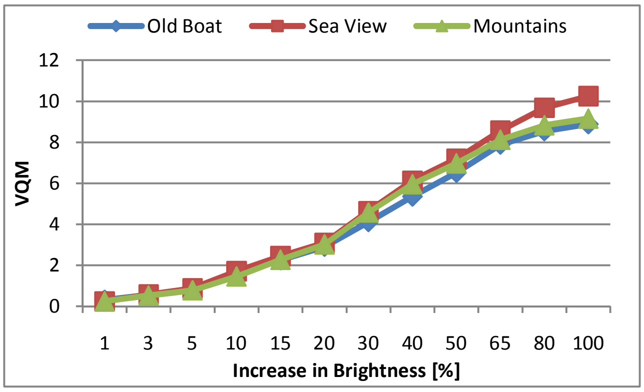

Figure 4. VQM increase due to changes in brightness

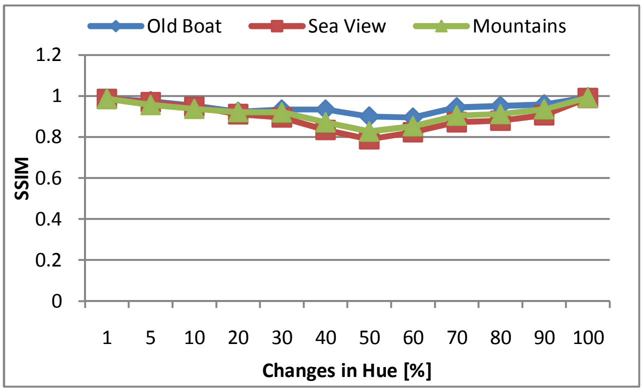

Figure 5. SSIM decrease due to changes in hue

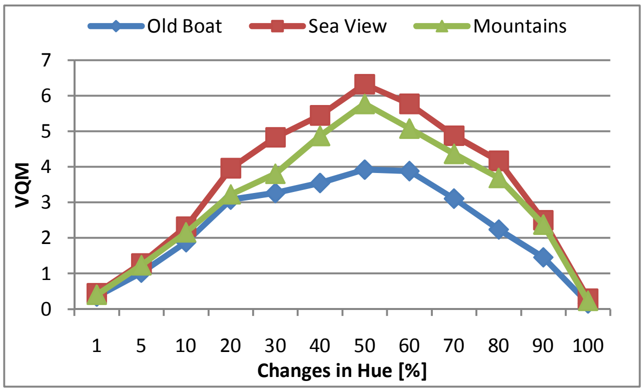

Figure 6. VQM increase due to changes in hue

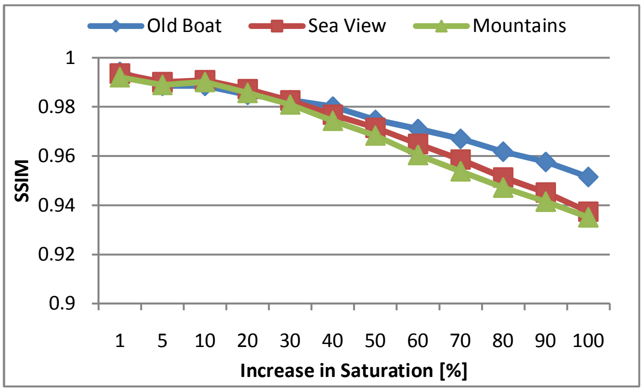


Figure 7. SSIM decrease due to changes in saturation

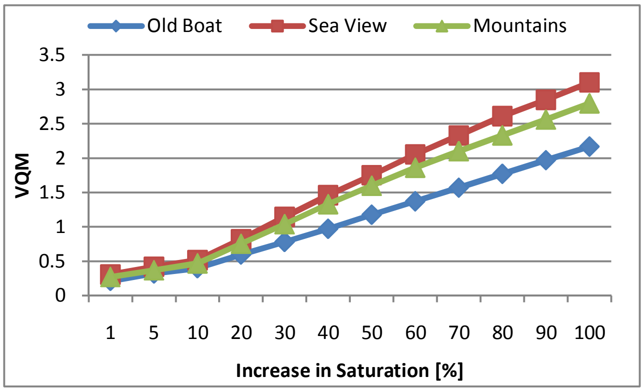


Figure 8. VQM increase due to changes in saturation

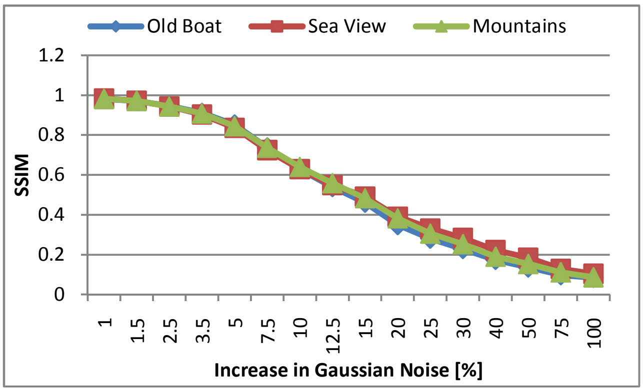


Figure 9. SSIM decrease due to the amount of Gaussian noise

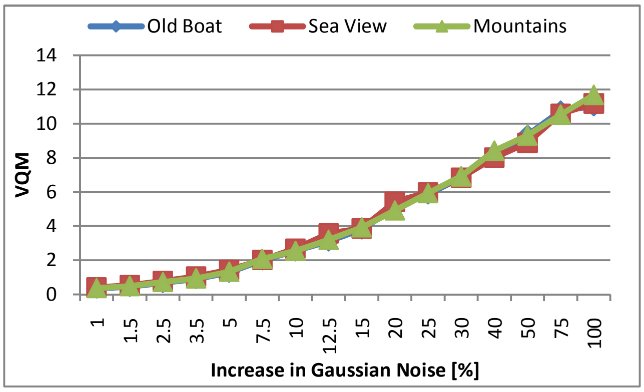


Figure 10. VQM increase due to the amount of Gaussian noise

The analysis of the charts and video sequences[1] shows that SSIM and VQM differ in quality estimation at some key points of image distortion. The most drastic decreases in SSIM values are due to the amount of introduced Gaussian noise as shown in figure 9. Changes in brightness (figure 3) and changes in contrast have only mild influence to the overall quality estimation by SSIM. Increase in brightness or contrast up to 25% barely influences the SSIM index, but introduction of Gaussian noise of only 10% causes quite large SSIM decrease. SSIM is very insensitive to changes in hue (figure 5) and to changes in saturation reacts moderately (figure 7).

On the other hand, the most drastic changes in VQM index, similar to SSIM, are also due to changes in the amount of Gaussian noise as shown in figure 10. But, the key concluding elements from these charts are that VQM is not so insensitive to changes in brightness (figure 4) or changes in contrast, and opposite to SSIM shows quite good sensitivity to changes in hue (figure 6) and saturation (figure 8). These characteristics of VQM clearly speak of some advantages that VQM has compared to the SSIM metric.

These issues of SSIM's low sensitivity to changes in brightness/contrast and hue contribute to lower performance in quality estimation compared to VQM metric. The next two images shown in figure 11 and figure 12 are chosen with similar VQM index. Visual comparison shows that both of them present quite large image distortion. Image presented in figure 12 has medium amount of Gaussian noise and image in figure 11 has a large amount of changes in hue where the colors are almost totally inverted. In this particular example it is hard for any estimator to rate the quality without any hesitation and it is only a matter of taste which image would have the higher quality. This is a subjective quality

[1] Due to the limitations of this paper only some of the charts and images of processed videos are presented. All the charts and images of video sequences are publicly available at http://vq.heliohost.org.

estimation question and the answer is of subjective nature. However, SSIM gives low quality grade to the image shown in figure 12 and extremely high quality grade to the image shown in figure 11, in spite of the highly inverted colors. This SSIM quality estimation with totally different grades clearly shows its performance issue concerning its low sensitivity to hue. Also it's a matter of question whether structure of the scene should play such a huge role in video quality estimation, independently of other changes, as is the case with SSIM metric. VQM metric estimates similar quality for both images shown in figure 11 and figure 12 which is more realistic estimation. This advantage of VQM over SSIM can be observed in most of the examples.

Figure 11. Video sequence with 30% changes in hue
VQM=3,2682 SSIM=0,9344

Figure 12. Video sequence with 12,5% increase in Gaussian noise
VQM=3,1118 SSIM=0,5397

In the second example, opposite to previous, images with similar SSIM index are presented. Comparing them visually leads to another conclusion of previously stated SSIM imperfections [2]. Visual comparison shows quite large difference in the perceived visual quality, with much higher visual quality on the side of the image shown in figure 13. VQM performs quite differently and more likely as any human estimator would rate the quality. According to VQM metric the video quality of the image shown in figure 13 is much higher than the video quality of the image shown in figure 14. This example only confirms the previous conclusion from analyzed charts of SSIM issues of low sensitivity to changes in brightness. Also, from this second example once again can be concluded that VQM metric performs better than SSIM and its performance is closer to visual quality estimation of the Human Visual System.

Figure 13. Video sequence with 7,5% increase in Gaussian noise
SSIM=0,7375 VQM=2,0809

Figure 14. Video sequence with 50% increase in brightness
SSIM=0,7231 VQM=6,9734

## 5. CONCLUSION

From all these charts and examples can be concluded that VQM metric performs quite better than SSIM in almost all situations. Through the analysis of the presented charts it can be concluded that changes in brightness have very mild influence to the SSIM which isn't the case of VQM and can be considered as VQM's advantage. Changes in hue and saturation are rated in a good manner by VQM, but to the SSIM metric they are almost invisible. This analysis clearly shows that VQM metric performance is significantly better than SSIM performance. 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. VQM metric 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.

## REFERENCES


[1] Kotevski, Z., "Analysis of quality and performance of MPEG-2 video compression techniques," Master Thesis, Faculty of technical sciences, Bitola, Macedonia (2007).

[2] Kotevski, Z., Mitrevski, P., "Experimental Comparison of PSNR and SSIM Metrics for Video Quality Estimation", Proc. of ICT Innovations-09 International Conference, Macedonian Society on Information and Communication Technologies, Ohrid, Macedonia (2009).

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

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

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

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

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

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

[9] Richardson, I., [H.264 and MPEG-4 Video Compression], John Wiley & Sons Ltd, ISBN 0-470-84837-5 (2003).

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

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

[12] Poynton, C. A., [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/