AUTHOR(S): CHILUKAMARI, J., KANNANGARA, S. and MAXWELL, G.

TITLE: Investigation of the effectiveness of video quality metrics in video pre-processing.

YEAR: 2013

Publisher citation: CHILUKAMARI, J., KANNANGARA, S. and MAXWELL, G. 2013. Investigation of the effectiveness of video quality metrics in video pre-processing. In Proceedings of the 3rd IEEE international conference on consumer electronics - Berlin (ICCE-Berlin 2013), 9-11 September 2013, Berlin, Germany. New York: IEEE [online], pages 1-5. Available from: https://doi.org/10.1109/ICCE-Berlin.2013.6698005

OpenAIR citation: CHILUKAMARI, J., KANNANGARA, S. and MAXWELL, G. 2013. Investigation of the effectiveness of video quality metrics in video pre-processing. In Proceedings of the 3rd IEEE international conference on consumer electronics - Berlin (ICCE-Berlin 2013), 9-11 September 2013, Berlin, Germany. New York: IEEE, pages 1-5. Held on OpenAIR [online]. Available from: https://openair.rgu.ac.uk

Investigation of the Effectiveness of Video Quality Metrics in Video Pre-Processing

Jayachandra Chilukamari, Sampath Kannangara and Grant Maxwell
School of Engineering, Robert Gordon University, Aberdeen, UK

Abstract—**This paper presents an investigation of the effectiveness of current video quality measurement metrics in measuring variations in perceptual video quality of pre-processed video. The results show that full-reference video quality metrics are not effective in detecting variations in perceptual video quality. However, no reference metrics show better performance when compared to full reference metrics, particularly, Naturalness Image Quality Evaluator (NIQE) is notably better at detecting perceptual quality variations.**

Index Terms— **Objective metric, video, quality, pre-processing, perceptual**

I. Introduction

Perceptual quality optimization algorithms are developed to improve the perceptual video quality of the compressed video. A number of perceptual video quality optimization algorithms employ low-pass pre-processing filters to achieve bandwidth-quality improvements [1-5]. In [1], the authors present an adaptive edge-preserving smoothing and detail enhancement pre-processing filter for perceptual optimization. The results are presented as subjective MOS scores as well as PSNR. However, the results showed that the PSNR results presented do not always correlate with the subjective quality variation. In [2] a pre-processing filter is used to remove spurious noise and insignificant features in video frames. The results indicate improvements in PSNR. In the research by Mancuso and Antonio Borneo [3], the filtering intensity is adjusted according to the amount of noise present in the video sequence to generate perceptually optimized videos. The PSNR is used as the metric to show that their non linear filters achieve higher quality videos. Similarly in [4] and [5] the quality improvement is presented as gains in PSNR and using visual evidence of video frames to highlight the reduction of artefacts. The authors of [6] have used variable Gaussian pre-processing filters which are controlled by a quality map which is used to indicate the distance to the region of interest. They have used PSNR as the objective video quality metric to show that by using a variable number of Gaussian filters there is an improvement in the perceptual quality. However, actual subjective quality results were not presented. In [7], a bilateral filter is dynamically configured depending on the traffic condition of the underlying network. The authors interpret that the filtered surgical video is visually equivalent to non-filtered surgical video for a telesurgery based application with PSNR improvements in regions of interest. De-Frutos-Lopez and his co-authors [8] proposed texture and motion adaptive filtering in which the bilateral filter parameters are estimated based on the motion and texture of the video. PSNR and visual comparisons are used to demonstrate the performance of the algorithm.

A major challenge in developing such pre-processing algorithms is the lack of accurate and repeatable video quality measurement metrics that can be used in video pre-processing applications. Typically, when video frames are pre-processed, the change in pixel values (compared to the original) are significant compared to the actual perceptual quality variations. Therefore, video quality metrics tend to produce inaccurate measurements. Some of the previous research demonstrates improved perceptual video quality vs. bandwidth performance of the developed algorithms using subjective quality evaluations [1]. Subjective video quality assessment (VQA) is an ideal way to validate the developed algorithms. However, its limitations in terms of complexity, time and cost have resulted in some of the visual redundancy based quality optimization algorithms employing objective error based quality measurements such as PSNR to evaluate perceptual quality. PSNR generally provides a degree of correlation with the actual perceptual video quality [9]. Therefore, use of error based measurements, at least at the development stage, is perceived to be justified, given the practical limitations of subjective video quality testing procedures. Currently there is no evidence to determine the suitability of PSNR or other full reference or no-reference perceptual quality metrics for measurement of perceptual video quality variations induced by pre-processing. The objective of this work is to investigate the effectiveness of PSNR and a number of state-of-the-art full reference objective video quality metrics, namely, Structural Similarity Index (SSIM) [10], Multi-Scale Structural Similarity Index (MS-SSIM) [11], Video Quality Metric (VQM) [12] and no reference video quality metrics Blind/Reference less Image Spatial QUality Evaluator (BRISQUE) [13], Blind Image Quality Index (BIQI) [14], Naturalness Image Quality Evaluator (NIQE) [15] and No reference metric for JPEG 2000 [16] in measuring the perceptual quality of pre-processed videos. In this work, we have used a low-pass Gaussian pre-processing filter at varying filter strengths.

II. Experimental Procedure

The experiments are carried out using five different Common Intermediate Format (CIF) resolution video sequences. These video sequences are Coastguard, Soccer, Hall monitor, Crew and Mother-daughter. Screenshots of the video sequences are shown in Fig.1. These video sequences are widely used in video coding research community during the development of perceptual quality optimization algorithms.

Fig.1. Screen shots of the test video sequences

A Gaussian pre-processing filter (kernel size of 3x3) is applied to these video sequences at standard deviations σ = 0.3, 0.5, 0.8 and 1.0 resulting in four different versions of each video sequence. Both the original video sequence and the filtered versions are encoded using High Efficiency Video Coding reference encoder, HEVC 4.0 [17] at four different quantization values (16, 24, 32, 40) resulting in four encoded rate-quality points per each version of the sequence. These rate quality points are chosen in such a way so that the effectiveness of the metrics in detecting the variations in perceptual quality can be studied at both low and high bitrates. This results in total of 100 video sequences. The subjective quality methods recommended by International Telecommunication Union (ITU) [18, 19] are the most widely adopted video quality evaluation strategies. The subjective quality of these videos is evaluated using Absolute Category Method (ACM) because of its ability to obtain repeatable results during the subjective evaluation [20-22]. These tests are carried out in a standard test environment with 60 non-expert viewers so that each of these 100 sequences is subjectively rated by 10 non-expert viewers. The sequences are presented one at a time with an interval of less than 10sec duration for voting time. Precautions are taken to avoid random votes from incoherent observers. Each viewer used an extended 11-point quality rating scale (from Bad to Excellent) to rate the quality of video sequences. An eleven point rating scale gives higher discriminative power that is needed to identify subtle differences in perceptual video quality. Subjective results gathered from subjective quality tests are used as a benchmark to evaluate the effectiveness of the metrics chosen for the investigation.

III. Results and Discussion

Figure 2 shows the Mean Opinion Score (MOS) values from the actual subjective tests against the bit rate for all encoded versions of the Crew sequence. It can be observed that the Gaussian filter with σ = 0.3 produces higher perceptual quality vs. bitrate performance. All other blur levels (standard deviations) generally produce a rate-perceptual quality loss. In Fig.3 the percentage subjective gain/loss is plotted against bitrate at σ =0.3 for all tested video sequences. The graph reveals that almost all the sequences when pre-processed with standard deviation equal to 0.3 achieve higher perceptual quality than the original video sequence (with an exception for soccer sequence at lower bitrates). Note that the actual percentage values are not directly comparable between different metrics due to their unique non-linear algorithms. These subjective results from all the video sequences serve as the benchmark for comparison with the chosen full/no reference metrics to determine the metric that best correlates with subjective perception. Figure 4 show the PSNR vs. Bitrate plots (PSNR calculated with the original video as the reference) for the same Crew video sequence. It is evident that PSNR is steadily decreasing with the increase in standard deviation. Therefore, PSNR does not show the gain in subjective quality that was observed at σ = 0.3. This corresponds to the induced variation in pixel values by the Gaussian filter (i.e. higher filter strength leads to lower PSNR). In Fig.5, the percentage PSNR gain/loss at σ = 0.3 for all the sequences is plotted. Across all the tested video sequences PSNR shows a perceptual loss. This behaviour was observed for all the other full-reference metrics (not shown here because of lack of space). This is because, in full reference metrics (perceptual or PSNR), changes in pixel values have a significant influence on the measured quality. This makes them particularly unsuitable for pre-processing based applications. Table 1 shows the quality results in an abbreviated format for all tested full reference quality metrics for σ = 0.3 and σ = 0.8. The second column indicates whether subjective results showed a gain or a loss in quality. The X or √ under each metric indicates whether that particular metric was able to correctly detect the actual gain or loss.

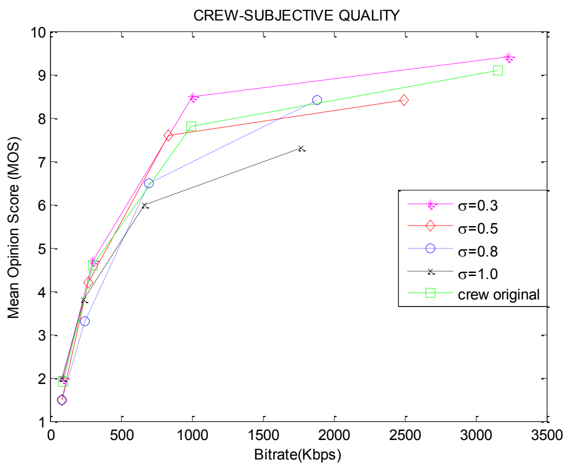

Fig.2. Crew-MOS vs Bitrate(kbps)

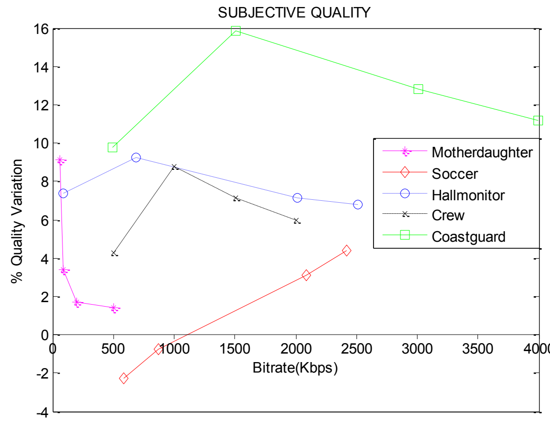

Fig.3. All sequences percentage subjective gain/loss at σ =0.3

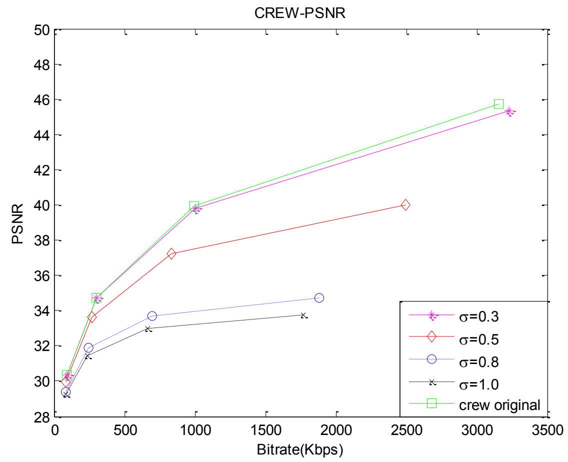

Fig.4. Crew-PSNR vs Bitrate(kbps)

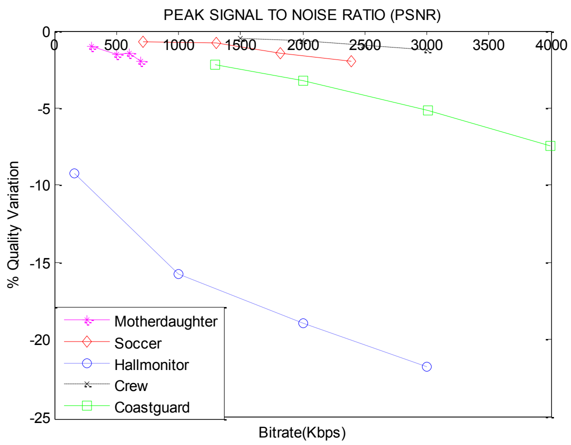

Fig.5. All sequences percentage PSNR gain/loss at σ =0.3

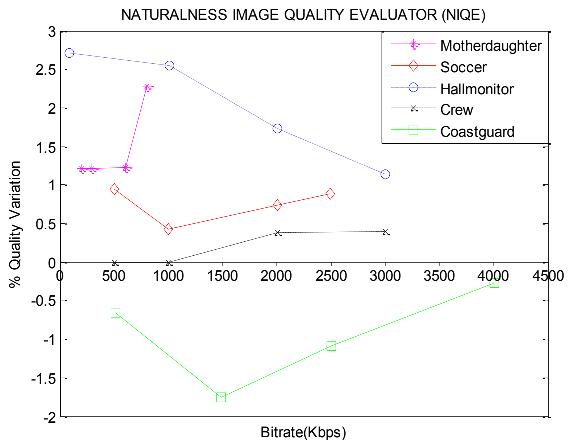

Fig.6. All sequences percentage NIQE gain/loss at σ =0.3

TABLE 1. FULL REFERENCE METRICS

Video Sequence	**Subjective quality**		**PSNR**		**SSIM**		**MS-SSIM**		**VQM**	
	0.3	0.8	0.3	0.8	0.3	0.8	0.3	0.8	0.3	0.8
Mother-daughter	gain	loss	X	✓	X	✓	X	✓	X	✓
Soccer	gain	loss	X	✓	X	✓	X	✓	X	✓
Hall monitor	gain	loss	X	✓	X	✓	X	✓	X	✓
Crew	gain	loss	X	✓	X	✓	X	✓	X	✓
Coastguard	gain	gain	X	X	X	X	X	X	X	X

TABLE 2. NO REFERENCE METRICS

Video Sequence	**Subjective quality**		**BRISQUE**		**BIQI**		**NIQE**		**NR FOR JPEG 2000**	
	0.3	0.8	0.3	0.8	0.3	0.8	0.3	0.8	0.3	0.8
Mother-daughter	gain	loss	X	✓	Partial detection	X	✓	✓	X	✓
Soccer	gain	loss	X	✓	Partial detection	✓	✓	✓	X	✓
Hall monitor	gain	loss	X	✓	X	X	✓	✓	X	✓
Crew	gain	loss	✓	✓	X	✓	✓	✓	X	✓
Coastguard	gain	gain	✓	X	✓	✓	X	X	X	X

In contrast with the full-reference metrics, no-reference video quality metrics apart from NR for JPEG 2000 have shown more promising detection ability. This is because no reference metrics estimate the quality of the video based on the local statistics of the video frames rather than depending on the pixel differences to judge the perceived quality. Their quality detection performance at $\sigma = 0.3$ and $\sigma = 0.8$ across the chosen video sequences can be seen in Table 2. BRISQUE detected a gain in perceptual quality for Crew and Coastguard sequences. BIQI showed partial gains for Soccer and Mother-daughter, and a clear gain for Coastguard sequence similar to subjective quality results. However, BIQI failed to detect the perceptual gain in Hall monitor, crew and the perceptual loss in Hall monitor and mother daughter at $\sigma = 0.8$. Similarly BRISQUE failed to detect perceptual gain in Mother-daughter, Soccer and Hall monitor and the perceptual loss in coastguard. However, NIQE effectively detected perceptual quality variations at both standard deviations $\sigma = 0.3$ and $\sigma = 0.8$ in all the video sequences except in coastguard sequence. Its detection performance at $\sigma = 0.3$ is shown in Figure 6. It can be inferred from Fig.6 that NIQE detects the perceptual gain that is shown by subjective results in four out of five tested video sequences except for coastguard video. Therefore, NIQE shows better detection ability compared to all other objective video quality measurement metrics. Moreover, NIQE apart from contradicting with actual subjective quality in coastguard sequence, its quality detection is not similar to subjective quality at all the bit-rates even in the case of sequences in which it detected the perceptual variations. Therefore, NIQE should be further improved to detect quality variations accurately when videos are pre-processed.

.

IV. CONCLUSIONS

The investigation shows that the full reference metrics (PSNR, SSIM, MS-SSIM, VQM) do not effectively identify the changes in perceptual quality when videos are pre-processed. However, no reference metrics (BRISQUE, BIQI, NR FOR JPEG 2000, and NIQE) show better performance when compared to full reference metrics particularly, Naturalness Image Quality Evaluator (NIQE) is notably better at detecting perceptual quality. Therefore further research has to be carried out to improve the performance of Naturalness Image Quality Evaluator (NIQE), so that it can be used in place of subjective quality testing during the development of perceptual quality optimization algorithms. Furthermore, this investigation indicates that judging the quality based on the local statistics of video frames (no-reference) is an appropriate option when compared to determining the quality based on pixel differences (full-reference) when videos are pre-processed.

REFERENCES

[1] K. Ji-Hye, L. Ji Won, P. Rae-Hong, and P. Min-Ho, "Adaptive edge-preserving smoothing and detail enhancement for video preprocessing of H.263," in *Consumer Electronics (ICCE), 2010 Digest of Technical Papers International Conference on*, 2010, pp. 337-338*)*.

[2] C. A. Segall, P. V. Karunaratne, and A. K. Katsaggelos, "Preprocessing of compressed digital video," in *Photonics West 2001-Electronic Imaging*, 2000, pp. 163-174.

[3] M. Mancuso and A. Borneo, "Advanced pre/post-processing for DCT coded images," *Consumer Electronics, IEEE Transactions on,* vol. 44, pp. 1039-1041, 1998.

[4] J. Xu, R. J. Sclabassi, L. Bing, and M. Sun, "Content-Based Video Preprocessing for Remote Monitoring of Neurosurgery," in *Distributed Diagnosis and Home Healthcare, 2006. D2H2. 1st Transdisciplinary Conference on*, 2006, pp. 67-70.R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev. in press.

[5] M. Li and Z. Xu, "An adaptive preprocessing algorithm for low bitrate video coding," *Journal of Zhejiang University-Science A,* vol. 7, pp. 2057-2062, 2006.

[6] L. S. Karlsson and M. Sjostrom, "Improved ROI video coding using variable Gaussian pre-filters and variance in intensity," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, 2005, pp. II-313-16.

[7] X. Jian, R. J. Sclabassi, L. Qiang, L. F. Chaparro, R. Marchessault, and S. Mingui, "Human Perception based Video Preprocessing for Telesurgery," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, 2007, pp. 3086-3089.

[8] M. de-Frutos-Lopez, H. Medina-Chanca, S. Sanz-Rodriguez, C. Pelaez-Moreno, and F. Diaz-de-Maria, "Perceptually-aware bilateral filtering for quality improvement in low bit rate video coding," in *Picture Coding Symposium (PCS), 2012*, 2012, pp. 477-480.

[9] T. Oelbaum, K. Diepold, and W. Zia, "A generic method to increase the prediction accuracy of visual quality metrics," in *Proceedings of the Picture Coding Symposium (PCS), Lisboa, Portugal*, 2007.

[10] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on,* vol. 13, pp. 600-612, 2004.

[11] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on*, 2003, pp. 1398-1402 Vol.2.

[12] Stephen Wolf and Margaret Pinson, " Video quality measurement techniques," NTIA Report 02-392, June 2002.

[13] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Blind/Referenceless Image Spatial Quality Evaluator," in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, 2011, pp. 723-727.

[14] A. K. Moorthy and A. C. Bovik, "A Two-Step Framework for Constructing Blind Image Quality Indices," *Signal Processing Letters, IEEE,* vol. 17, pp. 513-516, 2010.

[15] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a Completely Blind Image Quality Analyzer," *Signal Processing Letters, IEEE,* vol. 20, pp. 209-212, 2013.

[16] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *Image Processing, IEEE Transactions on,* vol. 14, pp. 1918-1927, 2005.

[17] HEVC. Available: http://hevc.kw.bbc.co.uk/trac/browser

[18] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," Rec. BT.500,Geneva, 2002.

[19] ITU-R, "Subjective assessment methods for image quality in high-definition television," Rec. BT.710, Geneva, 1998.

[20] ITU-T, "Subjective video quality assessment methods for multimedia applications," Rec. P.910, Geneva, 2008.

[21] Q. Huynh-Thu, M. Ghanbari, D. Hands, and M. D. Brotherton, "Subjective video quality evaluation for multimedia applications," in *Electronic Imaging 2006*, 2006, pp. 60571D-60571D-11.

[22] H.T. Quan, M. N. Garcia, F. Speranza, P. Corriveau, and A. Raake, "Study of Rating Scales for Subjective Quality Assessment of High-Definition Video," *Broadcasting, IEEE Transactions on,* vol. 57, pp. 1-14, 2011.