Performance Analysis of a No-Reference Temporal Quality Assessment Metric for Videos Impaired by Frame Freezing Artefacts

Muhammad Arslan Usman, Muhammad Rehan Usman, and Soo Young Shin

Abstract—In This paper, a temporal video quality assessment metric has been studied and evaluated on the basis of its ability to detect frozen video frames in a digitally transmitted video sequence and assess the quality accordingly. The metric is a temporal quality method that is centered around measuring the annoyance of frame freeze duration. This metric uses mean square difference (MSD) value between video frames to mark freeze events and builds a mapping function based on such durations of freeze to estimate the objectivemean opinion score (MOS). Two test scenarios have been designed for the analysis, Single freeze and multiple freeze experiment, keeping in view the recommendations of international telecommunication union (ITU). The analysis has been done with the help of MATLAB simulations and finally some limitations of the temporal quality metric are shown at the end of the paper.

Index Terms—Video quality, no-reference, temporal artefacts, frame freezing.

I. INTRODUCTION

The process of digital video transmission contains a series of steps in order to digitally transmit a video which includes a digital encoder, a channel, a digital decoder and an end user display hardware. All these steps are typically involved in transmitting a video sequence over a channel. During the transmission certain errors can take place which can degrade the digitally encoded data. The errors can be classified into a number of categories. Networks have certain limitations for which it is necessary to maintain acoding bitrate for the transmission of digital videos. For this reason, network operators provide a control mechanism on the encoder side in order to optimize the bitrate. Videos have different temporal and spatial variations depending on the content of the video. In case of high complexity in the content of a video, the network operators have to employ a mechanism on the encoder side in order to maintain the bitrate. The bitrate is normally maintained by fixing a threshold for the bitrate and reducing the frame rate of the video sequence. If the frame rate is not reduced then there is a certain chance that the bitrate will not be maintained causing spatial and temporal impairments such as a jitter or fluctuation in the video sequence. Also, if the complexity in the video content is not controlled then this can result in frame freezing or some frames might be dropped from the original video. Furthermore, an error prone network can result in temporal degradations of the transmitted video as it might introduce some errors in the transmitted video sequence which can result in frame freezing or frame dropping. Multiple number of consecutive frozen frames is called a freeze event. The reception side of the network is thus unable to recover the video in its original form. Some transmission protocols such as transmission control protocol (TCP) help in retransmission of a video if the already transmitted video fails to regain its original shape at the reception side. Retransmission can result in delays in a live video broadcasting scenario or live video streaming which is highly unfavorable for the end user. Network operators make sure that the video transmission is done without any delay or lag in order to have a good rating between the end users.Network operators need to employ different techniques in order to monitor the video quality so that such temporal degradations might be avoided. There are certain algorithms or metrics that help to monitor the quality of the video thus giving a better chance to the network operators in order to provide better Quality of Service (QoS) to the end users. All these concepts have been explained in detail in [1].

The metrics used to monitor the quality of the received videos, use different types of reference techniques. Full reference (FR), reduced reference (RR) and no reference (NR)are the 3 kinds methods used to monitor the received videos on the reception side. These methods have been used and described in details in [2]-[5]. FR method requires the full reference of the transmitted video on the receiving side in order to assess the video quality and reduced referenceuses partial reference or a few statistics of the transmitted video. FR and RR methods are not friendly towards network operators and have limited application as they require higher capital and operational expenditures. NR method, as explained in [4], [5], does not require any reference of the transmitted video sequence on the receiving side and is designed to assess the video quality by the received statistics of the video as shown in Fig. 1. Network operators ensure that the video transmission is monitored so that they can adapt certain measures to make the quality better after receiving feedback from the video quality assessment (VQA) metric.

The authors are with the Department of IT Convergence, Kumoh National Institute of Technology, Gumi, South Korea (e-mail: arslanusman@msn.com, rehanusman@gmail.com, wdragon@kumoh.ac.kr).

The VQA metric in [6] uses the NR method for temporal quality assessment of the received video sequences. This metric has been studied in this paper and it works by calculating the MSD between consecutive frames in order to highlight the freeze events and then builds a mapping function. The mapping function is then used to assess the quality of the video [6]. The metric has been discussed in section III of the paper in order to have a better understanding with the help of figures and mathematical equations. The performance of the metric is tested with the help of different videos which were generated for two scenarios, i.e., single freeze and multiple freeze. The test scenarios have been described in section II of this paper. With the help of the subjective tests and the results from the VQA metric a comparison is made in order to analyze the performance of the metric. Also, some limitations were found in the performance of the metric which are discussed at the end of the paper.
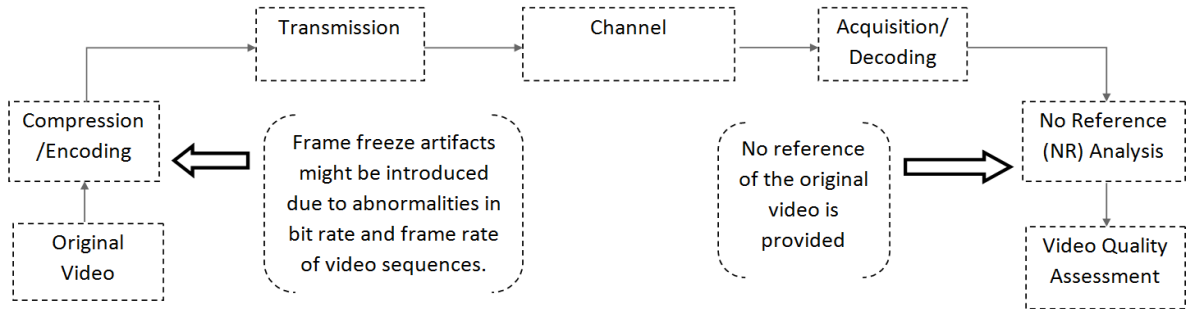

Fig. 1. Block diagram for digital communication process including no-reference method for video quality assessment.

Other related work includes the following. The metric used in [7] recursively aggregates arbitrary freeze distortions as an equivalent freeze distortion in the video under test. The temporal quality metric proposed in [8] accounts for the impact of various frame dropping situations and spatio-temporal luminance variations due to motion. In[9] the authors present estimation of the effects of lost frames on visual quality by analyzing the inter frame correlation present at the output of the rendering application.

The rest of the paper is organized as follows. Section II provides the test stimuli for the VQA subjective tests, Section III describes the VQA metric, Section IV includes the simulations. Finally Section V includes the conclusion.

TABLE I: DURATION OF FRAME FREEZES FOR SINGLE FREEZE

Frame Freezes (seconds)	0.12	0.2	0.52	1	2	3

II. TEST SCENARIO

Subjective tests are very important in order to benchmark any algorithm. These tests are used to make useful comparisons in order to better understand the performance of an algorithm or metric under certain conditions. Parameters such as frame rate, bitrate and frame size impact a lot on the working of a VQA metric. Some studies, such as [10], [11] have investigated the impact of these parameters on subjective testing. The contribution of this paper majorly emphasizes over evaluating the VQA metric under certain conditions. So, we have designed these tests using a number of video sequences and altering them using MATLAB. The frame rate of the video sequences is kept same as the native frame rate, i.e., 30 frames per second (fps). All the videos were kept uncompressed and used same resolution i.e. VGA 640x480, so that the video sequences could emphasize strongly to assess the perceptual impacts of frame freeze. We selected 5 videos with different spatial (SI) and temporal (TI) perceptual information(SI and TI as recommended by ITU-T Rec. P. 910 [12]). The video sequences used are described briefly as follows:

- Ice Skating: A crowd skating on white ice, low SI and TI.
- City View: Camera panning over high buildings, high SI and low TI
- Crew: A Space shuttle crew walking through an aisle, low SI and high TI
- Harbor: Boats moving in the sea near a harbor, high SI and high TI
- Football: A football practice match, high SI and high TI

All the video sequences were of equal duration of 12 seconds. The duration of the video, frame freeze and the number of frame freezes have to be selected carefully as the test are of high importance in order to assess the VQA metric. There are different sets of test conditions available and we have used the ones in [13], which comprises of single freeze experiment and multiple freeze experiments. There is a broad range of test scenarios for different frame freeze durations available but we have selected two of them for our test and they have been described in Table I and Table II. Table I refers to the single freeze conditions and Table II refers to the multiple freeze test conditions. Before performing the tests, the test lab was setup according to the recommendations of ITU-R BT 500-12 [14]. A cubical room with white walls was used in which a laptop was placed on a table. The software used for the subjective tests was configured and installed in the laptop. A total of 18 test subjects were chosen from youth, expert and non expert categories. The test subjects were trained before they were asked to do the assessment on the laptop. After every video was displayed, the test subject was asked automatically by the software to make the assessment of the quality of the video according to the absolute category rating (ACR) scale recommended by ITU-T P. 910 [12]. The test subject was left alone in the room till the test was complete in order to minimize the chance for any distractions. The ACR scale has been described in Table III. The video sequences were presented in a random order so that test subjects would not experience any fatigue during the tests. The tests were designed for a total duration of half an hour.

Both the tests, single and multiple freeze experiments, have been described in the following sections. The subjective test experiments for both the scenarios, multiple freeze and single freeze, can be studied in detail in [15].

TABLE II: DURATION OF FREEZE OCCURRENCES AND TOTAL DURATION OF FREEZE IN TEST STIMULI (MULTIPLE FREEZE)

Number of Freeze Occurrences	Duration of each freeze occurrence(seconds)/ total freeze Occurrence (seconds)					
1	0.08/0.080	0.16/0.16	0.48/0.48	0.96/0.96	1.92/1.92	3.84/3.84
2	0.08/0.160	0.16/0.320	0.24/0.48	0.48/0.96		
3	0.08/0.240	0.16/0.48	0.24/0.72	0.64/1.92		
5	0.08/0.400	0.160.8	0.4/2.0	0.8/4.0		
8	0.08/0.640	0.16/1.280	0.24/1.92	0.48/3.84		

A. Single Freeze Experiment

For the single freeze experiment, a single freeze was introduced using the halt and continuestrategy. Every sixth frame for each video was copied in the following frames of the video sequence in order to create a frame halt or frame freeze. There were in total six different frame freeze durations as described in Table I and for every video, all the frame freeze durations were used [13]. Thus, there were a total of 30 videos for the test subjects to perform the assessments.

The single freeze experiments were no longer than a duration of 20 minutes. The test results were used to generate the results for the single freeze experiment in order to calculate the mean opinion scores (MOS). The MOS from the tests has been used to make the comparison which has been discussed in the Section IV.

B. Multiple Freeze Experiment

For the multiple freeze experiment, same videos were used but this time for a different set of frame freezes and number of freeze even occurrences which have been described in Table II [13]. So, a total of 110 videos were generated for the multiple freeze experiment. These tests were performed on the same test subjects and the total duration of the test was not longer than half an hour approximately.

The MOS for this experiment was recorded in order to make the comparison with the VQA metric and it has been discussed in Section IV of the paper.

III. DESCRIPTION OF THE METRIC USED FOR PERFORMANCE ANALYSIS

The metric in [6] uses temporal information in the video sequences to assess the quality of the video. The algorithm used in this metric calculated the MSD between consecutive frames and if the MSD is found to be zero then it means that the video is frozen at that particular instant. This kind of approach is practically not easy as in real practice there is always a noise factor involved due to an error prone channel. So, even if there are two identical frames, there MSD might not be equal to zero, as one of the frames might have introduced some noise in them. In this algorithm, firstly all the frames in the video sequence are converted to YUV color format as it is recommended to use this format for the purpose of image processing.

For calculating the MSD, two steps are involved as described in [6] and are briefly explained in this paper.

First, MSD is calculated between the consecutive frames and if the MSD is zero then it is considered as a frozen frame. But if there is low complexity in the motion content then it might be considered as a frozen frame which is not favorable. So another step is involved in which the MSD between the frozen frame and the first frozen frame of the freeze event is calculated. So the chance of low complexity motion be considered as a frozen frame can be avoided.

After calculating the MSD with the help of the above steps it can be said a frame 'i' will be considered a frozen frame if the MSD between a frame and its previous frame is zero and also the MSD between the same frame and the first frame of the freeze event is below a certain threshold. This has been explained more thoroughly in [6]. The threshold parameters are defined in [6]. After the information about the freezes and freeze events is collected, then the following equations are used in order to develop a mapping function. The mapping function is then used to calculate the MOS for a given video sequence. The mapping function is derived with the help of following equations.

$$FDF(b) = \frac{FrDur\ (b)}{TotDur}\ 100 \tag{1}$$

$$FTDP(b) = \frac{FrTotDur\ (b)}{TotDur}\ 100 \tag{2}$$

where; 'b' is called a bin and it represents a freeze event in the video sequence. *FrDur*, *TotDur* and *FrTotDur* represent the duration of a single freeze event, total duration of the video and cumulative sum of all the freeze events respectively.

Finally, the mapping function is computed using Eq. (1), (2) which is as follows:

$$T_1(b) = \frac{1}{f2(FTDP(b)).f1(FDP(b))+f3(FTDP(b))} \tag{3}$$

where $f_1(x)$, $f_2(x)$ and $f_3(x)$ are calculated as;

$$f_1(x) = a_1 + b_1 \times log(c_1 \times x + d_1) \tag{4}$$
$$f_2(x) = a_2 \times x^2 + b_2 \tag{5}$$
$$f_3(x) = a_3 \times x^2 + b_3 \tag{6}$$

The values for a_1, b_1, c_1 and d_1 have been empirically calculated with the help of the subjective tests in [6]. Each $T_1(b)$ value is then bound-limited in the range 1-5 according to the ACR scale given in Table III and the following equation is used to suggest the MOS of the video sequence:

$$T_1'(b) = min\ (max(T_1(b),1),5) \tag{7}$$

where in Eq. (7), 5 represents excellent quality video and 1

represents poor quality [6]. Finally, the temporal VQA metric is set to:

$$T_1 = min\ (T_1'(b)) \tag{8}$$

This metric has been implemented in MATLAB and its simulations results are discussed in the next section along with the subjective test results from Section II.
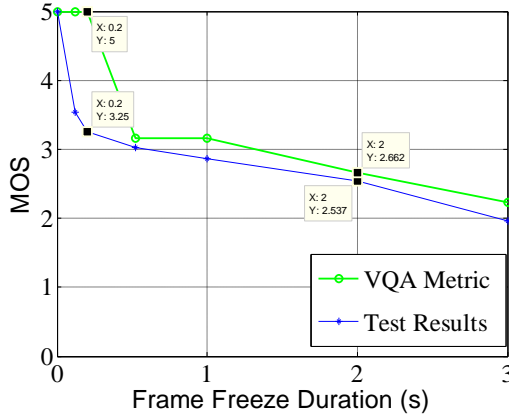

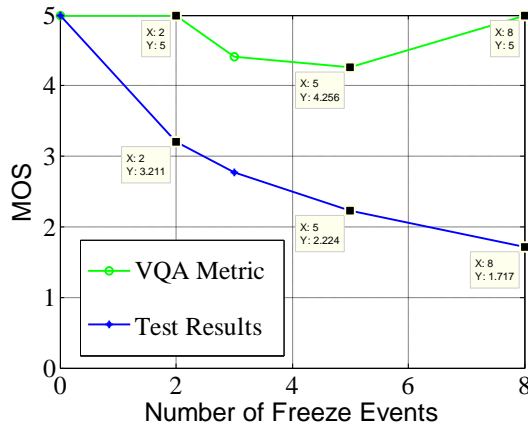Fig. 2. Comparison of the metric and test results (single freeze).

Fig. 3. Comparison of the metric and test results (multiple freeze).

TABLE III: DESCRIPTION OF ABSOLUTE CATEGORY RATING (ACR)

MOS	1	2	3	4	5
QUALITY	BAD	POOR	FAIR	GOOD	EXCELLENT

IV. SIMULATION RESULTS

After performing the subjective tests described in section II, a comparison has been made between the MOS from the tests and the VQA metric. The simulation results have been discussed separately for the single and multiple freeze scenarios.

A. Simulation Results for Single Freeze Scenario

From Fig. 2 it can be seen that the VQA metric has performed remarkably well and the MOS results are much closer to the MOS calculated through the subjective tests. Especially, if we take a look at the MOS for freeze durations of 0.5 and 2 seconds, it seems that the VQA metric gives almost the same results as the subjective tests.But, after careful analysis of the results, it can be noted that the frame freezes which are lesser than 0.52 seconds have not been noticed at all by the VQA metric. The frame freezes with freeze durations of 0.12 and 0.2 seconds have been assessed

as excellent quality and an MOS of 5 is shown by the metric which means that these freeze durations passed the metric undetected.

This limitation has further been studied for the multiple freeze scenario as it is too early to make any conclusions regarding the performance of the metric.

B. Simulation Results for Multiple Freeze Scenario

In Fig. 3, it is clear that the metric does not perform well for videos that contain multiple frame freezes. Here, again it can be noticed that the frame freezes which were lesser than the duration of 0.5 seconds approximately, have passed the VQA metric without any detection.

For a video sequence containing one freeze event and two freeze events the VQA metrichas shown a MOS of 5 (excellent quality)where the subjective test results show a MOS of 3 (fair) as shown in Fig. 4. Similar behavior can be noticed in fig. 3 for the video sequence that contains 8 freeze events, the VQA metric shows a MOS of 5 but the test results show a MOS between 1 (bad) and 2 (poor).But, for three and five number of freeze events, the metric is showing results which are not even closer to the subjective tests. Here, if Table II is noticed, it can be seen that most of the frame freeze durations are lower than 0.5 seconds and the VQA metric is simply unable to detect the frozen frames which are lower than this duration.

After careful study of this metric and running a number of tests with a freeze duration less than 0.5 seconds, it can be said that the VQA metric is unable to detect frozen frames which are lesser than 0.5 seconds (0.48 seconds approximately).

V. CONCLUSION

After going through the simulation results for the single and the multiple freeze scenarios, it can be concluded that the VQA metric performs well for detection of frozen frames and assessing the quality of the videos but under certain parameters and conditions. The performance of the metric for single freeze scenario is better than multiple freeze scenario as shown in Fig. 1 and Fig. 2 respectively. The multiple freeze scenario is more complex and hence requires more careful detection of frozen frames for which the metric does not perform well.

Also we noticed that any frozen frame greater than the duration of 0.48 seconds is detected and the quality of the video is assessed accordingly but, the metric fails to perform for frame freezes lesser than 0.48 seconds which is shown in Fig. 4. Most of the frame freezes for the multiple freeze scenario were designed to be lesser than 0.48 seconds as shown in Table II, so this can be the reason for the metric to perform poorly in this scenario.

Hence, it can be concluded that the VQA metric is limited to perform quality assessment of the videos that contain frame freezes greater than 0.48 seconds. Otherwise it fails to detect any frozen frames in the videos resulting in wrong assessment of the video quality.

REFERENCES

[1] Q. H. Thu and M. Ghanbari, "Temporal aspect of perceived quality in mobile video broadcasting," *IEEE Transactions on Broadcasting*, vol. 54, pp. 641–651, September 2008.

[2] S. Li, L. Ma, and K. N. Ngan, "Full-reference video quality assessment by decoupling detail losses and additive impairments," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1100–1112, July 2012.

[3] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 4, pp. 684–694, April 2013.

[4] M. Shahid, A. Rossholm, and B. Lövström, "A no-reference machine learning basedvideo quality predictor," in *Proc. the 5th International Workshop on Quality of Multimedia Experience*, pp. 176–181, June 2013.

[5] S. Wolf, "A no reference (nr) and reduced reference (rr) metric for detecting dropped video frames," in *Proc. the Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, vol. 2, 2009.

[6] Q. H. Thu and M. Ghanbari, "No-reference temporal quality metric for video impaired by frame freezing artefacts," in *Proc. the Conference on Image Processing IEEE International*, 2009, pp. 2221–2224.

[7] K. Watanabe, J. Okamoto, and T. Kurita, "Objective video quality assessment method forevaluating effects of freeze distortion in arbitrary video scenes," in *Proc. the SPIE-Image Quality and System Performance IV*, vol. 6494, 2007.

[8] R. R. P. Vidal and J. C. Gicquel, "Automatic quality assessment of video fluidity impairments using a no-reference metric," in *Proc. the International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, vol. 2, 2006.

[9] F. Battisti and A. Neri, and M. Carli, "No-reference qualitymetric for color video communication," *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2012.

[10] G. Zhai, J. F. Cai, W. S. Lin, X. K. Yang, W. J. Zhang, and M. Etoh, "Cross-dimensional perceptual quality assessment for low bit-rate videos," *IEEE Transactions on Multimedia*, vol. 10, pp. 1316–1324, November 2008.

[11] M. Shahid, A. K. Singam, A. Rossholm, and B. Lövström, "Subjective quality assessment of H.264/AVC encoded low resolutionvideos," in *Proc. the 5th International Congress on Image and Signal Processing (CISP)*, pp. 63–67, 2012

[12] *Subjective Video Quality Assessment Methods for Multimedia Applications*, ITU-T, Recommendation ITU-R P.910, September 1999.

[13] S. V. Kester, T. Xiao, R. E. Kooij, K. Brunnström, and O. K. Ahmed, "Estimating the impact of single andmultiple freezes on video quality," *The International Society for Optical Engineering*, vol. 7865, pp. 78650–78660, February 2011.

[14] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, ITU-R, Recommendation BT.500-12, September 2009.

[15] M. A. Usman, "Performance evaluation of video quality assessment methods based on frame freezing," M.S. thesis, Dept. Applied Signal Processing, Blekinge Institute of Tech., Karlskrona, Sweden, 2014.

M. Arslan Usman was born in Okara City, Pakistan in 1988. He received his B.S. degree in electrical engineering from COMSATS Institute of Information Technology, Lahore, Pakistan, in 2010 and M.S. degrees in electrical engineering with specialization in signal processing from Blekinge Tekniska Hogkola (BTH), Karlskrona, Sweden, in 2013. His research interests include coexistence among networks, cognitive radio, MIMO, OFDM, NOMA, small cells, audio, visual signal processing and next generation mobile wireless networks (4G/5G).

He has worked in Alcatel-Lucent Nigeria as a wireless pre-sales engineer (4G-LTE campaign for West-Central Africa), from 2013 to 2014, in the department of technical sales. After serving in Alcatel-Lucent for a year he is now a PhD research scholar at WENS Lab in School of Electronics in Kumoh National Institute of Technology since September 2014.

M. Rehan Usman was born in Lahore City, Pakistan in 1986. He received his B.S. degree in electrical engineering from COMSATS Institute of Information Technology, Lahore, Pakistan, in 2008 and He holds two M.S. degrees, M.S. electrical engineering with specialization in telecommunications, M.S. project management and operational development from Blekinge Tekniskac Hogkola (BTH), Karlskrona, Sweden, in 2010 and 2013. His research interests include coexistence among networks, cognitive radio, MIMO, OFDM, NOMA, small cells and next generation mobile wireless networks (4G/5G).

He was a lecturer in electrical engineering department at University of South Asia Pakistan from July 2012 to January 2013. Then he joined as a lecturer in Electrical Engineering Department of Superior University Lahore, Pakistan, from January 2013 to March 2014. After serving in superior University he is now a PhD research scholar at WENS Lab in School of Electronics in Kumoh National Institute of Technology since March 2014.

Soo Young Shin was born in 1975. He received his B.S, M.S, and Ph. D degrees in electrical engineering and computer science from Seoul National University, Korea in 1999, 2001, and 2006, respectively. His research interests include wireless LAN, WPAN, WBAN, wireless mesh network, sensor networks, coexistence among wireless networks, industrial and military network, cognitive radio networks, MIMO, OFDM, mmWave, NOMA and next generation mobile wireless broadband networks (4G/5G).

He was a visiting scholar in FUNLab at University of Washington, US, from July 2006 to June 2007. After 3 years working in WiMAX Design Lab. of Samsung Electronics, he is now an assistant professor in School of Electronics in Kumoh National Institute of Technology since September 2010.