TIME ALIGNED VIDEO QUALITY MEASUREMENTS OF LONG HAUL SKYPE CALLS

Jayasekara J. M. N. D. B.[1], Gunarathne T.[2], Kumara W. G. C. W.[3]

[1]Faculty of Applied Sciences, Rajarata University of Sri Lanka, Sri Lanka. Email: nuwanjayasekara@gmail.com
[2]Department of Mechatronics, Faculty of Engineering, South Asian Institute of Technology and Medicine (SAITM), Sri Lanka. Email: tharanga.saitm@gmail.com
[3]MINE Lab, Department of Computer Science and Information Engineering, National Central University, Taiwan. Email: chinthakawk@gmail.com

ABSTRACT

Quality of the video calls is very important in any kind of short or long haul video communications through the Internet. Even though the Internet is now equipped with high bandwidth intercontinental data channels still general video calls such as Skype suffers from video and audio quality degradations due the best effort nature of the Internet. In this paper an analysis of the long haul Skype video call quality is presented. Considering the video quality, there are two factors which affects the user experience, namely, misalignment of the frames of the video at the receiver with respect to the original sequence, and quality degradations of the respective frame pair of the sender and receiver. Hence, we first found the best matched frame at the receiver and then performed several standard quality measurements. Presented results are much better comparing to the previous findings due to the fact of that realignment.

***Key words*:** video call, VoIP, Quality of user Experience (QoE)

1. INTRODUCTION

Based on Jing Zhu's [1] study on traffic characteristics and video quality of Skype video calls in a LAN simulating delay characteristics of a WAN, Kuamra et al. [2] carried out a similar analysis based on real WAN Skype call. In a video call user experience can be damaged due to two factors as, misalignment of the original frame sequence and quality degradation of each respective frame, between sender and receiver, where in [2] only quality measurements are presented without aligning received frames based on the sender's original sequence of frames. Here in this paper, a detailed analysis of video quality is presented after aligning receiver's frames w.r.t. sender's frames.

Here, objective assessment methods are used which are basically based on algorithms to evaluate the quality. Objective quality assessment then can be divided into three main categories as full reference (FR), reduced reference and no reference (NR), where FR is used in this paper. Full reference compares the received video with the original video, reduced reference compares only some characteristics and no reference use characteristics of the received video only [3].

Following are the seven parameters measured during the experiment in short [4].

- Peak-to-peak Signal-to-Noise Ratio (PSNR),

$$PSNR = 10 \cdot log_{10} \frac{MaxErr^2 \cdot w \cdot h}{\sum_{i=0,j=0}^{w,k} (X_{i,j} - Y_{i,j})^2} \quad (1)$$

  where, *MaxErr* = maximum possible absolute value of color components difference, *w* = video width and *h* = video height, *X* and *Y* are the reference and test video frames. If frames are equal value is 100 and higher values stands for better similarity.
- Structural similarity (SSIM) index is based on measuring of three components (luminance similarity, contrast similarity and structural similarity) and combining them into result value. We used the SSIM (fast) parameter only. Higher values are better and 1 is for equal frames.
- MSU blocking metric measures subjective blocking effect in video sequence where lower values correspond to lower blocking.
- MSU blurring metric compares power of blurring of two images where lower values correspond to higher blurring.
- Delta is the mean difference of the color components in the correspondent points of image.

$$d(X,Y) = \frac{\sum_{i=1,j=1}^{m,n}(X_{i,j} - Y_{i,j})}{mn} \quad (2)$$

where m and n are the number of pixels in width and height of the frame, X and Y are the reference and test video frames. 0 means equal frames, positive and negative values mean deviation, lower absolute values are better.

- Mean Sum of Absolute Difference (MSAD) is the mean absolute difference of the color components in the correspondent points of image.

$$d(X,Y) = \frac{\sum_{i=1,j=1}^{m,n}|X_{i,j} - Y_{i,j}|}{mn} \quad (3)$$

0 means equal frames, lower values are better.

- Mean Squared Error (MSE),

$$d(X,Y) = \frac{\sum_{i=1,j=1}^{m,n}(X_{i,j} - Y_{i,j})^2}{mn} \quad (4)$$

Lower values are better, 0 for equal frames.

2. METHODOLOGY

Two experiments were carried out using 1 sec and 2 sec source videos based on Foreman standard AVI video file with 320×240 spatial resolution, 25 fps, 4:2:0 YUV color space. First, SplitCam[1] software was used to feed the source video to Skype at the sender and played continuously in a loop, where sender was in Taiwan and the receiver was in Sri Lanka through NATs. Evaer[2] software was used to record the received video for 10 s after a 30 s waiting time for network traffic to settle down and five videos were recorded and analyzed for both experiments. At last received video was compared with the original (reference) video using the MSU Video Quality Measurement (VQM) Tool[3].

To compensate the fact that received video frames are misaligned w.r.t the original video sequence, we performed time alignment as in [1] which is sometimes referred as calibration. Reference video (x), recorded video (y) and matched video (z) are the original video at the sender, received video at the receiver and matched video frame by frame w.r.t. original video respectively. Normalized cross correlation was used as follows,

$$z^{(s)} = \arg\max_n (f(x^{(s)}, y^{(n)})), \quad (5)$$

where, $x^{(s)}$ is the s^{th} reference video frame, $y^{(n)}$ is the n^{th} recorded video frame, and $z^{(s)}$ is the index of the recorded video frame that match the s^{th} reference video frame.

$$f(a,b)$$

where, a_{ij} and b_{ij} are the value of luma (brightness) of the $(i,j)^{th}$ pixel in reference and recorded video frames, and A and B are width and height of the reference and recorded videos. Figure 1 shows best matched frame numbers against original frame numbers for 1 sec example.

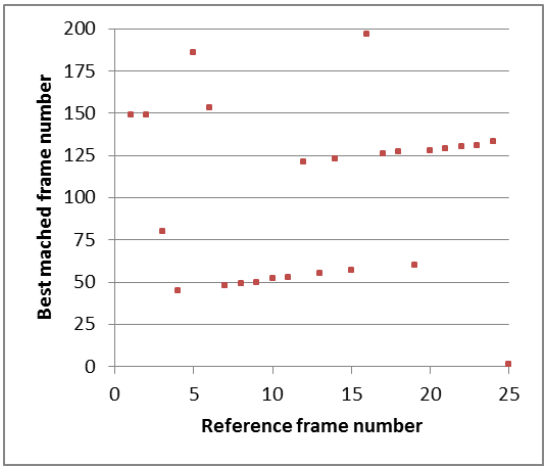


Figure 1: Best matched frame numbers

3. RESULTS

Table 1: Average experimental results

Parameter		[2]	Test 1	Test 2
PSNR		13.52	27.93	26.74
SSIM (fast)		0.38	0.90	0.87
MSU Blocking metric	Source	15.08	12.64	12.47
	Test	14.48	12.31	13.02
MSU Blurring metric	Source	15.08	16.32	16.28
	Test	13.78	14.79	13.88
Delta		0.96	1.04	1.01
MSAD		38.26	5.37	6.57
MSE		2956.5	107.34	145.40

Table 1 provides the average parameter values

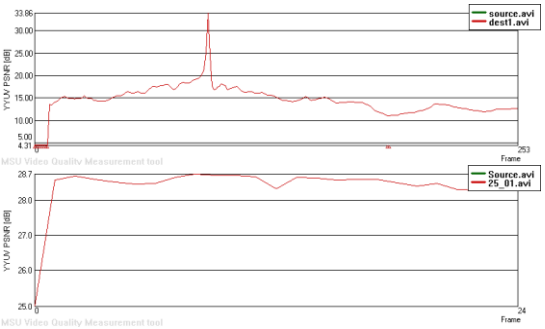
[1] http://www.splitcamera.com

[2] http://www.evaer.com/

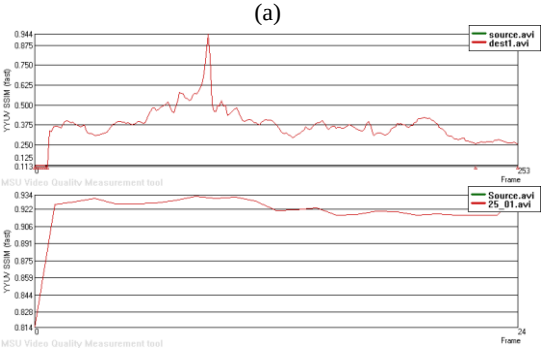
[3] http://compression.ru/video/quality_measure/video_measurement_tool_en.html

provided by the MSU VQM tool during the experiment and values in [2]. Values obtained in both test 1 and 2 are similar and are better compared to the values in [2] after the frame alignment based on the best matches.

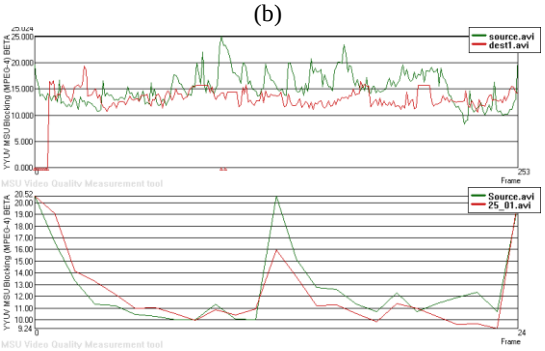
Figure 2 shows the graphs provided by the MSU VQM tool for each parameter in [2] and 1 sec test. First and second figure in (a) shows the PSNR values for each frame for [2] and 1 sec test respectively, where in the top figure a good PSNR peak is achieved only once where the frames are very similar and in the other hand lower figure shows high PSNR values for every frame since frames are aligned and tested against the similar frame only. Similarly, figures (b) to (g) shows SSIM (fast), MSU Blocking metric, MSU Blurring metric, Delta, MSAD and MSE respectively, with better results in the lower figures due to the frame alignment.



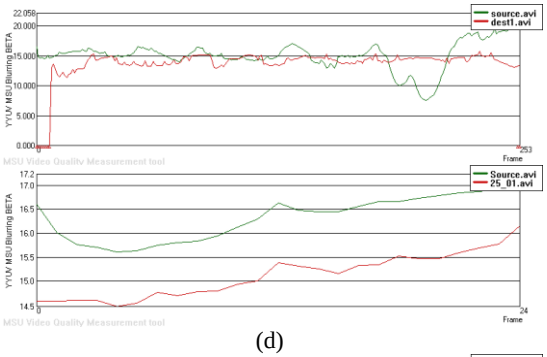
(a)



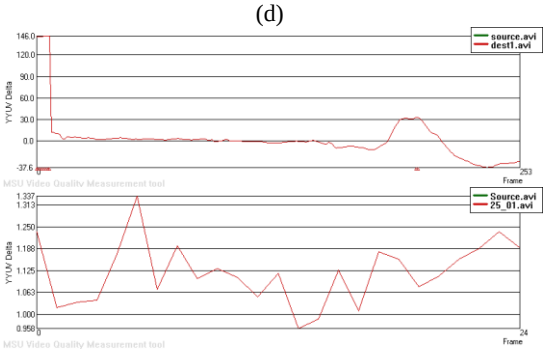
(b)



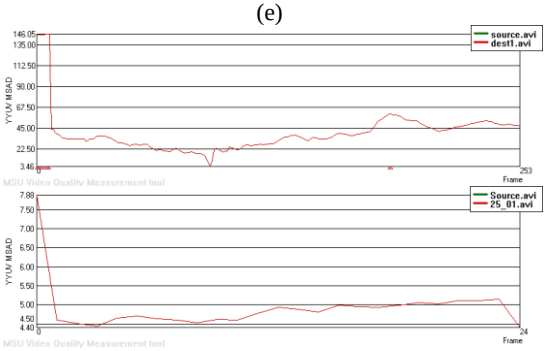
(c)



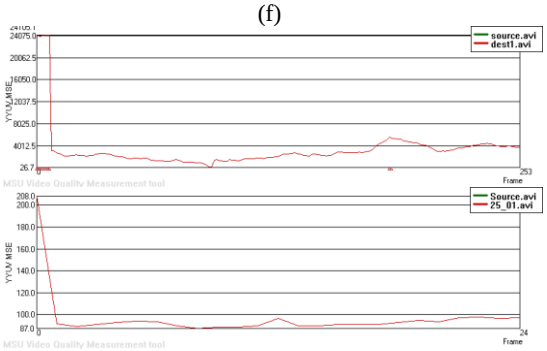
(d)



(e)



(f)



(g)

Figure 2: Performance graphs (a) PSNR (b) SSIM (fast) (c) MSU Blocking metric (d) MSU Blurring metric (e) Delta (f) MSAD (g) MSE. Upper is for [2] and lower is for 1 sec test.

Figure 3 shows several frames extracted from reference video and 1 sec test video to demonstrate that frames are now aligned. Best matches for the frames 0, 13 and 24 in the reference video are found at frames 149, 123 and 1 in the test video respectively.

(a) (b) (c)

Figure 3: Selected frames for PSNR test (top: source frame, bottom: matched frame) (a) 0 and 149 (b) 13 and 123 (c) 24 and 1

4. CONCLUSION

A detailed video quality analysis for a long haul Skype video quality is presented. Best matches for the reference video is first found in the received video based on realignment of the frames using normalized cross correlation between frames. Reference and realigned received video is then tested for seven video quality parameters namely, PSNR, SSIM, MSU Blocking metric, MSU Blurring metric, Delta, MSAD, and MSE. All the tested parameters provided better results comparing to the values received without realignment of the frames. Analysis of the received frames based on the sequence and delay in the received video with respect to the original video to understand the effect on user experience is planned as a future work.

5. REFERENCES

[1] Jing Zhu, "*On traffic characteristics and user experience of Skype video call*," Quality of Service (IWQoS), 2011 IEEE 19th International Workshop on, vol., no., pp.1-3, 6-7 June 2011

[2] Kumara W. G. C. W., Jayasekara J. M. N. D. B. and Gunarathne T., "*Video quality and traffic characteristics of Skype video calls in WAN*", SAITM Research Symposium on Engineering Advancements 2013 (RSEA 2013), Apr 2013

[3] Iain E. Richardson, "*The H.264 Advanced Video Compression Standard*", 2nd Ed., Wiley, 2010, ISBN 0470989289, 9780470989289

[4] *MSU Video Quality Measurement Tool*, http://compression.ru/video/quality_measure/video_measurement_tool_en.html, visited on 8 Jan 2014