Journal of Physics:
Conference Series

PAPER • OPEN ACCESS

Experimental Comparison of Video Streaming Platforms and Devices with Objective Quality of Experience Metrics Towards Reliable Multimedia Applications in Education

To cite this article: Emmanuel Adetiba *et al* 2019 *J. Phys.: Conf. Ser.* **1378** 032087

View the article online for updates and enhancements.

Experimental Comparison of Video Streaming Platforms and Devices with Objective Quality of Experience Metrics Towards Reliable Multimedia Applications in Education

Emmanuel Adetiba[1,3*], Matthew B. Akanle[2], Daniel T. Babalola[2], Victor Akande[2], Peter Sheidu[2], Bolaji Ariyo[2], Surendra Thakur[4], Jules Raymond Kala.[5,]

[1]Department of Electrical and Information Engineering, Covenant University, Canaanland, P.M.B 1023, Ota, Nigeria
[2]Centre for Systems and Information Services (CSIS), Covenant University, Canaanland, P.M.B 1023, Ota, Nigeria
[3]HRA, Institute for Systems Science, Durban University of Technology, P.O. Box 1334, Durban, South Africa
[4]KZN e-Skills CoLab, Durban University of Technology, Durban, South Africa
[5]Department of Information Technology, Durban University of Technology, Durban, South Africa

Abstract
This paper presents the result of experimental comparison of the video quality of streamed videos using streaming devices and streaming platforms. The objective metrics considered for the comparison of these media are the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), Visual Information Fidelity (VIF), Information Fidelity Criterion (IFC), and Visual Signal-to-Noise Ratio (VSNR). To determine the better platform for effective multimedia applications especially for educational purposes, videos were obtained and streamed via Data video, YouTube and Facebook. In the first two scenarios, Data video at two different settings were engaged at the encoding end and PotPlayer was used for decoding and recording of the transmitted videos. In the third and fourth scenarios, Videos were streamed via YouTube and Facebook. The aforementioned objective metrics were implemented in MATLAB R2017b for the experimental analyses and our results showed that YouTube produced streamed videos of the best quality with PNSR of 37.98, SIMM of 0.98 and VSNR of 33.12.

Key words: Video Quality, Video Streaming, Streaming Platform, Streaming Device, Datavideo, YouTube, Facebook

1. Introduction
Today, there has been an upsurge in the use of video services across different sectors of human endeavour. The demand for visual data has greatly increased as modern technologies have made transmission at the highest possible resolutions achievable [1] Several technologies have also emerged to aid video transmission across diverse platforms. This has made it possible to stream videos across various devices as well. Video streaming today has found application in business, education, entertainment and several other domains (Cisco Visual Networking Index, 2019). Thus, analysing and ensuring that videos remain at the highest possible quality has become very important.
As videos are streamed or transmitted, they tend to degrade in quality [3]. This is because transmission systems introduce some form of distortion as the video is generated, processed and transmitted, which leads to a loss in video quality. Video quality can be analysed using two major approaches, namely: subjective and objective analysis.

Subjective video quality analysis involves interacting with viewers directly. Here, videos are shown to a group of viewers who then rate the overall quality of the videos. The viewers' ratings are averaged to give a Mean Opinion Score (MOS). Subjective analysis is usually carried out in three major ways, which include i) Single stimulus, ii) Double stimulus and iii) Comparison stimulus. In Single stimulus**,** the distorted or streamed videos are displayed one after the other and are then evaluated by each viewer. The original video is sometimes presented among the other videos without the viewer's knowledge. Double stimulus approach involves the presentation of the streamed/distorted and the original videos to the viewer. He/she is then required to rate the streamed/distorted video relative to the original one. Comparison stimulus is quite similar to the double stimulus. However, in the comparison stimulus, only the streamed/distorted videos are presented to the viewer without any reference to the original. The viewer is then expected to rate the videos relative to one another (Recommendation ITU-R BT,500-13, 2012).

Objective video quality analysis on the other hand tries to estimate user opinion of video quality without any need for user interaction [5]. It involves the use of algorithms and models to evaluate the quality of videos. These models are generated by reading certain video measurable characteristics such as the resolution, aspect ratio and frame rate to name a few [6]

All over the world, education is currently powered using several media to provide seamless learning experience both for on-campus students and for those who may not be able to physically attend a school [7]. This is anchored on recent advances in computer, communication and multimedia technologies [8] Today, it has become very easy to transmit audio, video, text and their combinations across the Internet. As a result, entire courses and even programmes are being delivered online. The study at hand is concerned with experimental evaluation of the quality of videos across various streaming devices and platforms to compare and discover which would be more suited for an efficient and reliable multimedia applications for educational purposes.

2. Methodology

2.1 Streaming Devices, Platforms and Dataset

The videos, which are the dataset for this study were recorded and transmitted using Datavideo NVS-20, YouTube and Facebook. The NVS-20 is a streaming and recording device used to stream live events as well as record them for future editing or analysis. It allows for recording in MP4 or TS file formats and can be used in encoding mode to generate H.264 stream [9]. To configure Datavideo NVS-20, several settings had to be selected in order to determine the parameters through which the videos would be streamed. Two different instances were set up to determine the influence of parameter settings on video quality. The configurations of the NVS-20 in this study are shown in Table 1. The settings under Configuration 1 column indicate an ideal streaming environment while the settings in Configuration 2 column mirrors YouTube's encoding recommendations.

Table 1: Two Different Configurations of Datavideo NVS-20

S/N	Encoder Setup	Datavideo	
		Configuration 1	**Configuration 2**
1	Scale Down	No	No
2	H.264 Encode	Main 3.1	High 4.0
3	GOP Structure	IBP	IBBP
4	GOP Size	14	15
5	Video Bitrate(Kbps)	3,000	3,000
6	Video Rate Mode	VBR	CBR
7	Audio Stereo	Stereo	Stereo
8	Stereo Bitrate(Kbps)	384	384
9	Audio Source	DIG	DIG
10	Analog Audio System	EBU	EBU

YouTube and Facebook provide easy platforms to stream videos. These videos can either be recorded and uploaded or uploaded live. However, to upload a recorded video for live streaming, an Open Broadcaster Software (OBS) is usually engaged. The OBS that was adopted for this study is named Streamlabs. It allows for easy streaming while maintaining very high video quality standards.

2.2 Parameter Tuning for Video Analysis and Processing

An effective video analysis using the streaming devices (Datavideo-NVS-20) and the selected streaming platforms (YouTube and Facebook) involves tuning of relevant technical parameters hereafter described.

2.2.1 H.264 Encoding: This has become one of the most common encoding formats used for video processing today. It was developed by the Video Coding Experts Group (VCEG) in collaboration with the Moving Picture Experts Group (MPEG). H.264 encoding supports several profiles, which are sets of algorithms designed for various applications [9]. The Datavideo NVS-20 supports three encoding profiles, namely Main 3.0, Main 3.1 and High 4.0. The main profile is primarily used for standard-definition TV broadcasts while the high profile is associated with high-definition TV broadcasts. YouTube streaming recommends the H.264 encoding at High 4.1 [9]

2.2.2 Group of Pictures (GOP) Structure: This refers to the order in which frames are arranged. It allows for easier decoding, enhanced video quality and limited data loss. These frames are grouped and arranged in a pre-set order. They include IBP, IBBP and IPPP. The GOP size refers to the distance between two key frames (i.e. I-frame). Datavideo NVS-20 allows either IBP, IBBP or IPPP GOP settings with varying GOP sizes, YouTube’s recommended setting is IBBP for a proper stream while the recommended settings for Facebook is I-frame every 2s.

2.2.3 Video Bitrate: This refers to the amount of data that is processed at a given instance of time[10]. The video bitrate affects the quality of the video. A smaller bitrate reduces the number of bits that are available in one second thereby lowering the video quality but with the advantage of faster transfer rate. Conversely, a higher bitrate inversely increases the number of bits available in a second, leading to a higher video quality but a slower transfer rate (Sayood, K. 2003) Datavideo NVS-20 offers several bitrate options ranging from 800 – 6,000 Kbps. For high quality streaming, YouTube's recommended setting is a bit rate of 3,000 – 6,000 Kbps while the recommended bitrate for Facebook is 4,000Kbps.

2.2.4 Bitrate Encoding Mode: This can either be Variable Bitrate (VBR) or Constant Bitrate (CBR). In VBR mode, the encoder varies the bit rate of the video to maintain a particular video quality selected by the user while at CBR, the encoder maintains the same bit rate not considering the quality of the video. Datavideo supports both encoding modes while YouTube and Facebook recommended setting is CBR.

Table 2 illustrates the various parameters used by both configurations of the Datavideo as well as the settings for YouTube and Facebook.

Table 2: Encoding Parameters Specifications for Datavido, Youtube and Facebook

PARAMETERS	Datavideo NVS-20 (Configuration 1)	Datavideo NVS-20 (Configuration 2)	YOUTUBE (recommended settings)	Facebook (recommended settings)
Video compression type	H.264	H.264	H.264	H.264
Codec profile	MAIN	HIGH	HIGH	-
Codec layer	3.1	4.0	4.1	-
GOP Structure	IBP	IBBP	IBBP	I-frame every 2s
GOP Size	14	15	Half the video frame-rate	-
Video Bitrate (Kbps)	3,000kbps	3,000kbps	3,000 – 6,000kbps	4,000kbps
Bitrate encoding	VBR	CBR	CBR	CBR

2.3 Objective Video Quality Metrics

Rather than using humans to rate a video or image through subjective quality assessment, several mathematical models are often adopted to evaluate video quality based on several parameters, which can be gotten from the images or videos. The objective video quality metrics are grouped into three categories based on how much information is available to be used as a reference, viz; no reference, reduced reference and full reference (C. Sasivarnan, et al., 2011). **No reference metrics** assesse the quality of a streamed video without using the original video as a reference (Y. Fu-zheng et al., 2003). It tends to be less accurate than the full or reduced reference but uses up less computing power. In **reduced reference metrics,** the original or referenced image is partly available. Only certain parts of both videos are compared to give an overall score. They may not be as accurate as the full reference but are more efficient because they use up fewer computing resources like no reference (A.A. Webster et al., 1993). The **Full reference metrics** require a complete reference video. They compute video quality by comparing the original video with the received video. These metrics are usually the most accurate at the expense of higher computational costs. They cannot be used in situations where the original video is absent as illustrated in Fig. 1 [11]. The most common full reference metrics include the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), Visual Information Fidelity (VIF), Information Fidelity Criterion (IFC), and Visual Signal-to-Noise Ratio (VSNR)[21,22,23,24]. These metrics were employed in this study in order to rigorously compare the streaming qualities of the selected streaming devices and platforms.

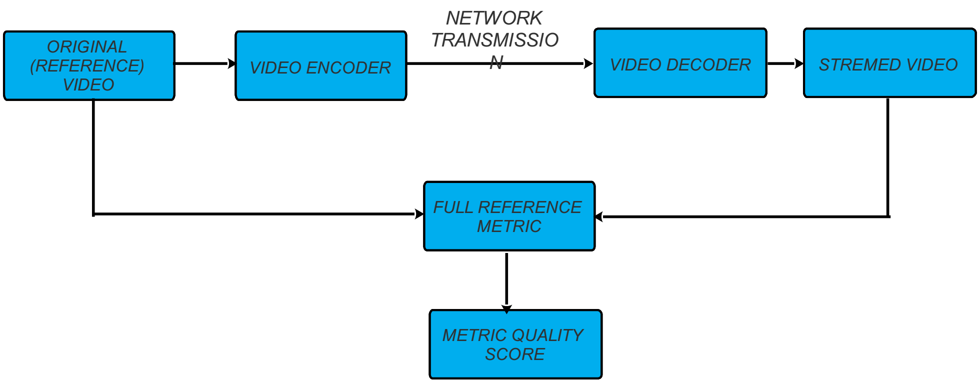

Figure 1: Block diagram of full-reference Video Quality Assessment (VQA)

2.3.1 Peak Signal to Noise Ratio

The Peak Signal to Noise Ratio (PSNR) metric refers to the ratio between a signal and the associated noise introduced in its representation. As it is a full reference metric, the signal refers to the original while the noise is gotten from the compressed or distorted data. It is easily determined via the Mean Squared Error (MSE). The MSE simply measures the average of the squares of the deviations (E. L. Lehmann et al. 1998). A higher PSNR value generally indicates a distorted or compressed image of better quality.

It is given by the equation below:

$$PSNR = 10\log_{10}\frac{(2^n-1)^2}{\sqrt{MSE}} \tag{1}$$

where MSE is defined as:

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(x_{i,j}-y_{i,j})^2 \tag{2}$$

2.3.2 Structural Similarity Measure

The Structural Similarity Measure (SSIM) is a full reference metric used to measure the similarity between two video streams. This implies that the measurement is based on an original video, which serves as the reference. It is currently the best objective method for determining image or video quality. Its value lies from 0 to 1. Due to its performance, SSIM is used ahead of other metrics and is recognized globally as a standard for video quality measurement. It is given by the equation below.

$$SSIM = \frac{(2\times\bar{x}\times\bar{y}+C1)(2\times\sigma_{xy}+C2)}{(\sigma_x^2+\sigma_y^2+C2)\times(\bar{x}^2+\bar{y}^2+C1)} \tag{3}$$

where C1 and C2 are constants.

$\bar{x}$, $\bar{y}$, σ_x^2, σ_y^2 and σ_{xy} are defined as follows:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{4}$$

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{5}$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2 \tag{6}$$

$$\sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})^2 \tag{7}$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y}) \tag{8}$$

2.3.3 Visual Information Fidelity
The Visual Information Fidelity (VIF) measures the loss of human-perceivable information in the distortion process. It compares the source image and a sample of the source image after going through a distortion channel. A perfect quality image is used as a source. In the absence of any distortions, the source signal passes through the Human Visual System (HVS) channel of a human observer before entering the brain, which extracts cognitive information from it. For distorted images, we assume that the reference signal has passed through another "distortion channel" before entering the HVS [12]. VIF is derived by comparing the Source and the distorted image through Human Visual System (HSV) models. The comparison can be expressed as:

$$VIF = \frac{\sum_{j\in subbands} I(\vec{C}^{N,j};\vec{F}^{N,j}\mid S^{N,j})}{\sum_{j\in subbands} I(\vec{C}^{N,j};\vec{E}^{N,j}\mid S^{N,j})} \tag{9}$$

Where j is for the *j-th* subband level,

E and *F* denote the visual signal at the output of the HVS model from the reference and the test images in one sub band respectively,

C is the Random Field (RF) from a subband in the reference signal,
I is the set of spatial indices for the RF while *N* is covariance denoted by:

$$C_N = \sigma_n^2 I \tag{10}$$

2.3.4 Visual Signal to Noise Ratio

The Visual Signal to Noise Ratio (VSNR) quantifies the visual fidelity of distorted images [13]. It operates with both near-threshold and suprathreshold distortions to estimate visual fidelity of an image

$$VSNR = 20\log_{10}\left(\frac{C(I)}{\alpha d_{pc} + (1-\alpha)\frac{d_{gp}}{\sqrt{2}}}\right) \tag{11}$$

where, C(I) denotes the Root Mean Square (RMS) contrast of the original image I.

$$C(I) = \frac{\sigma_{L(I)}}{\mu_L(I)} \tag{12}$$

d_{pc} - perceived contrast of the distortion

d_{gp} - disruption of global precedence

$\alpha \in \lfloor 0,1 \rfloor$

2.3.5 Information Fidelity Criterion

Information Fidelity Criterion (IFC) is the mutual information between the source and the distorted images. Similar to the VIF, the source image is passed through a given statistical model for the source and the distortion (channel)[13]. The following equation is used to compute IFC.

$$IFC = \sum_{k \in subbands} I(C^{N_k,k}; D^{N_k,k} \mid s^{N_k,k}) \tag{13}$$

where,

$C^{N_k,k}$ denotes N_k coefficients from the RF,

C^k of the k-th subband and similarly for $D^{N_k,k}$ and $s^{N_k,k}$

2.4 Data Analysis

In this study, we obtained a 1080p video, which was split into four separate clips for analysis. Each of the clips was one minute in length and was extracted using PotPlayer at well spread out portions.

Notably, operations such as PSNR and SSIM can only be performed on images in MATLAB. Thus, video clips were further split into frames in order to compute the PSNR and SSIM. A 60-second video clip could contain as much as a thousand frames depending on the framerate. The video in this study had a framerate of 30fps resulting in 1800 frames every minute. Figure 2 illustrates the flowchart we developed to break the video clips into frames. The flowchart was implemented in MATLAB with relevant functions for PSNR and SSIM computations. Functions to compute VIF, VSNR and IFC were imported as external functions from https://github.com/sattarab/image-quality-tools.

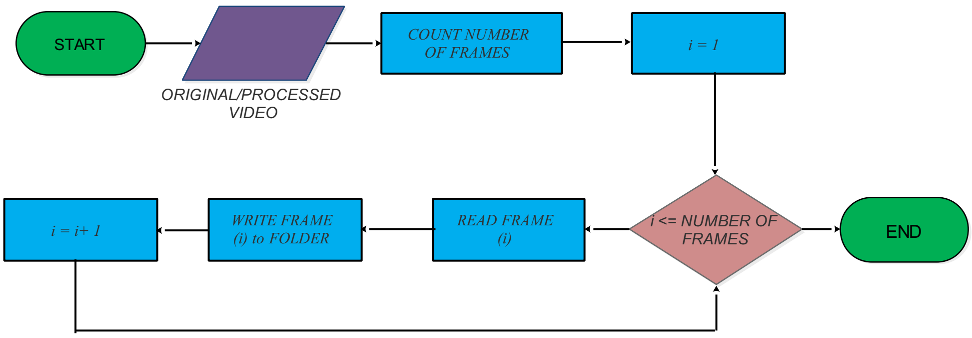

Figure 2: Flowchart to break down the video into respective frames

Figure 3 illustrates the procedure to compute the metric values for an entire video clip. As shown in the flowchart, the metric value is calculated for each frame and then stored in a sum variable. When every frame has been analysed, the total value is divided by the number of frames

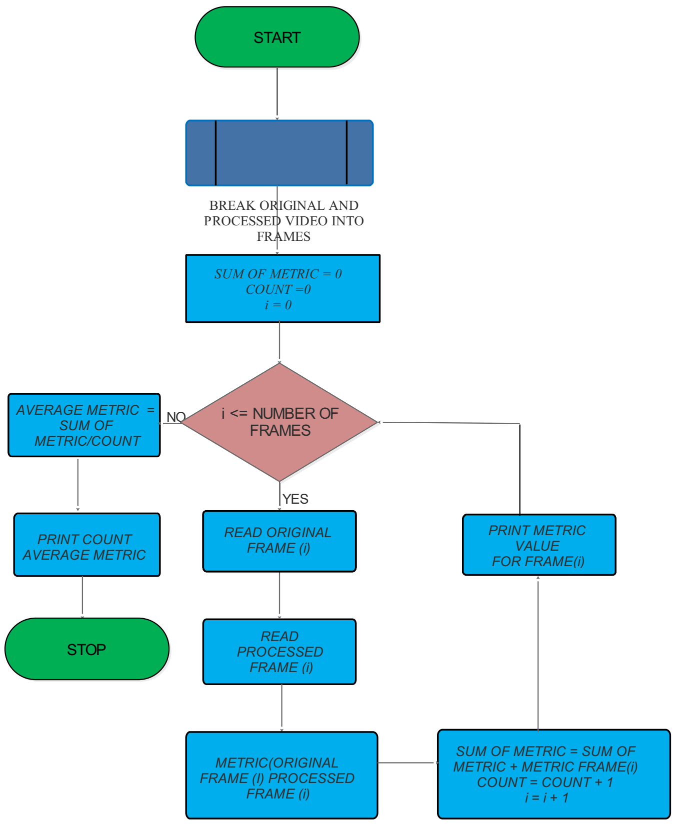

Figure 3: Flowchart to evaluate average metric value

3. Results and Discussion

Following the foregoing methodology, the average metric values for each video clip as well as

Clip 1		Average PSNR	Average SSIM	Average VSNR	Average IFC	Average VIF
	Data video (test 1)	27.76	0.92	21.04	0.67	0.23
	Data video (test 2)	30.08	0.96	24.50	1.01	0.31
	Youtube	36.83	0.96	33.15	1.26	0.37
	Facebook	32.25	0.94	23.47	1.87	0.44
Clip 2	Data video (test 1)	29.47	0.96	24.85	0.89	0.31
	Data video (test 2)	29.96	0.96	25.99	1.13	0.35
	Youtube	38.79	0.98	33.68	1.51	0.43
	Facebook	33.43	0.97	24.50	2.21	0.49
Clip 3	Data video (test 1)	26.83	0.91	18.83	0.63	0.20
	Data video (test 2)	28.29	0.93	22.20	0.91	0.28
	Youtube	37.79	0.98	31.81	1.35	0.39
	Facebook	33.99	0.97	24.88	2.10	0.48
Clip 4	Data video (test 1)	28.38	0.95	22.41	0.89	0.28
	Data video (test 2)	28.83	0.95	23.52	1.11	0.32
	Youtube	38.40	0.98	33.84	1.58	0.45
	Facebook	32.81	0.96	23.60	1.67	0.42
Overall Video	Data video (test 1)	28.11	0.93	21.78	0.77	0.26
	Data video (test 2)	29.29	0.95	24.05	1.04	0.32
	Youtube	37.95	0.98	33.12	1.43	0.41
	Facebook	33.12	0.96	24.11	1.96	0.46

the overall rating for the video are shown in Table 2.

Table 2: Overall metrics for assessed video

Line graphs were generated for each metric to aid visualisation. These line graphs illustrate all the metric values for each video clip across all frames. Figure 4 shows the line graph for PSNR, Figure 5 for SSIM, Figure 6 for VIF, Figure 7 for IFC, Figure 8 for VSNR while Figures 9 and 10 show the overall metrics for each clip.

Results from each clip (as illustrated in Table 2, the line graphs as well as the bar charts) show that over several portions of the video, YouTube provides the best average metric values (for PSNR (37.98), SSIM(0.98) and VSNR(33.12)) while Facebook also present the best average values (for VIF(1.96) and IFC(0.46)). It should however be noted that Facebook clips were evaluated at a resolution of 720p, which is the allowed resolution on the platform. The first Datavideo tests (Test1) consistently yielded a lower video quality than the second tests (Test2) due to different settings. This shows that changes in encoding settings would yield different results in video quality.

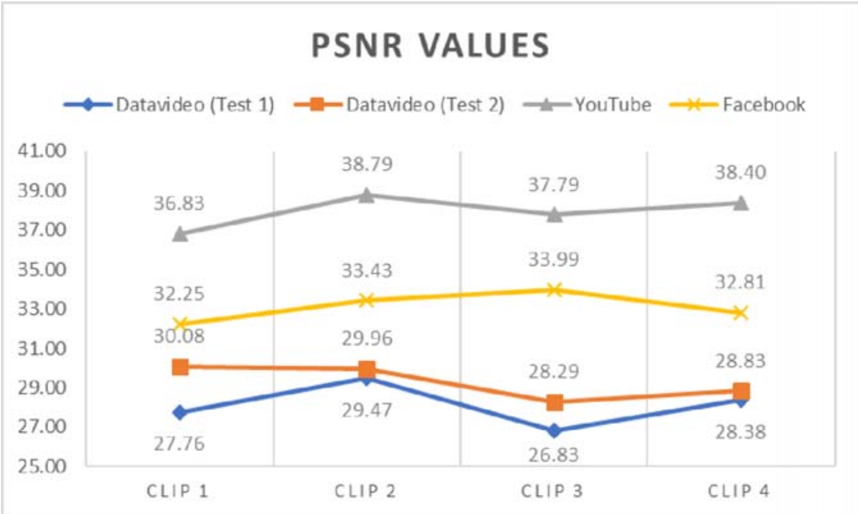

Figure 4: Line graph showing PSNR values

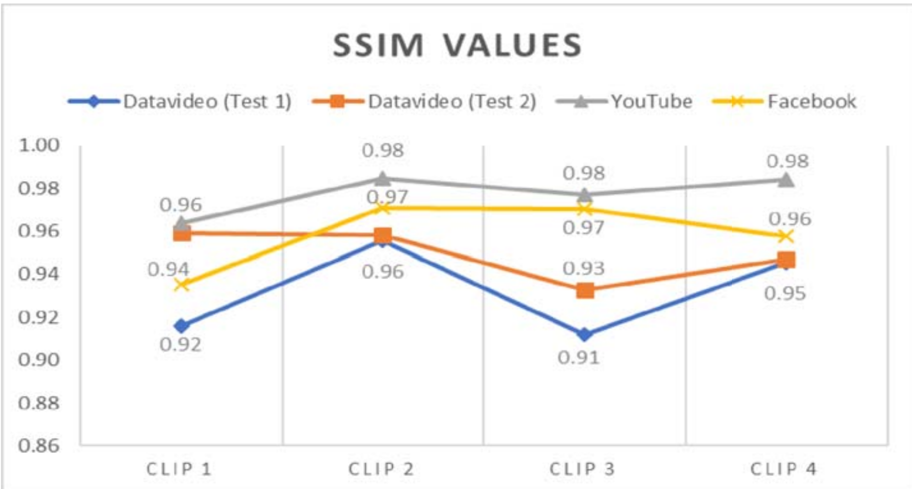

Figure 5: Line graph showing SSIM values

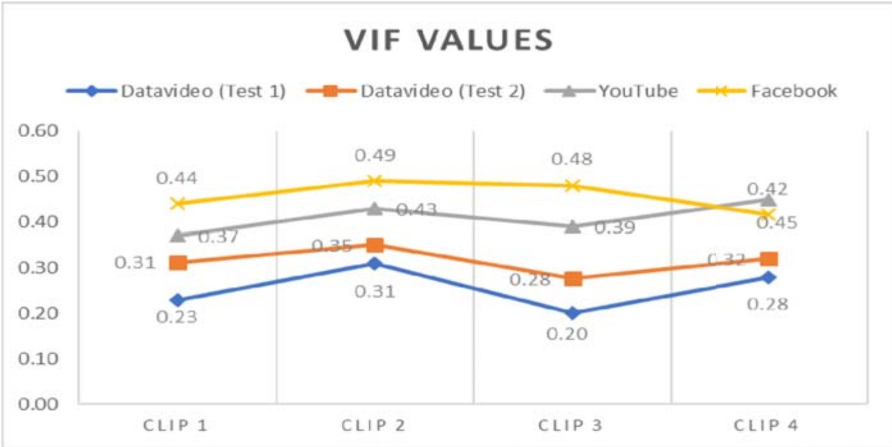

Figure 6: Line graph showing VIF values

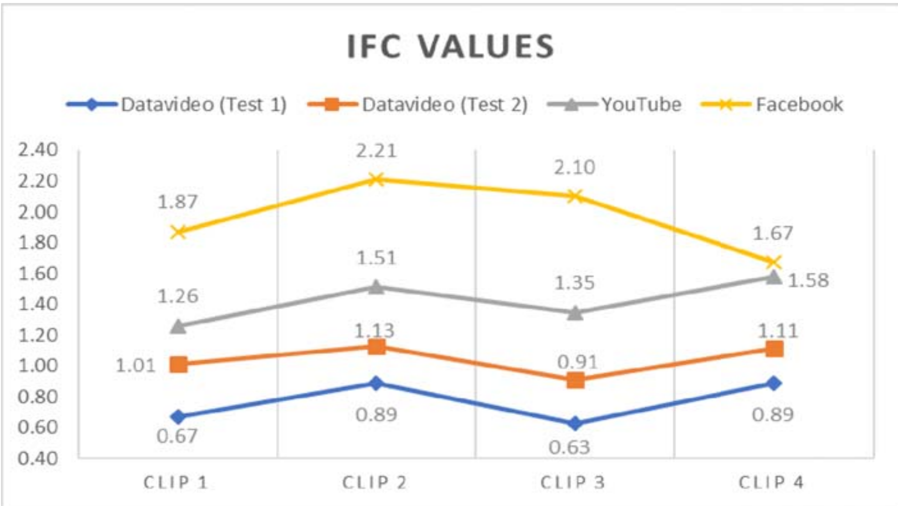

Figure 7: Line graph showing IFC values

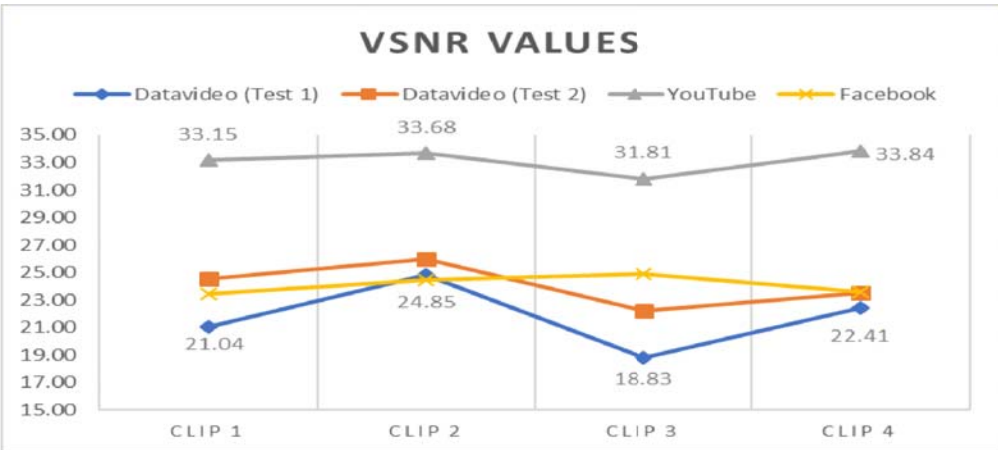

Figure 8: Line graph showing VSNR value

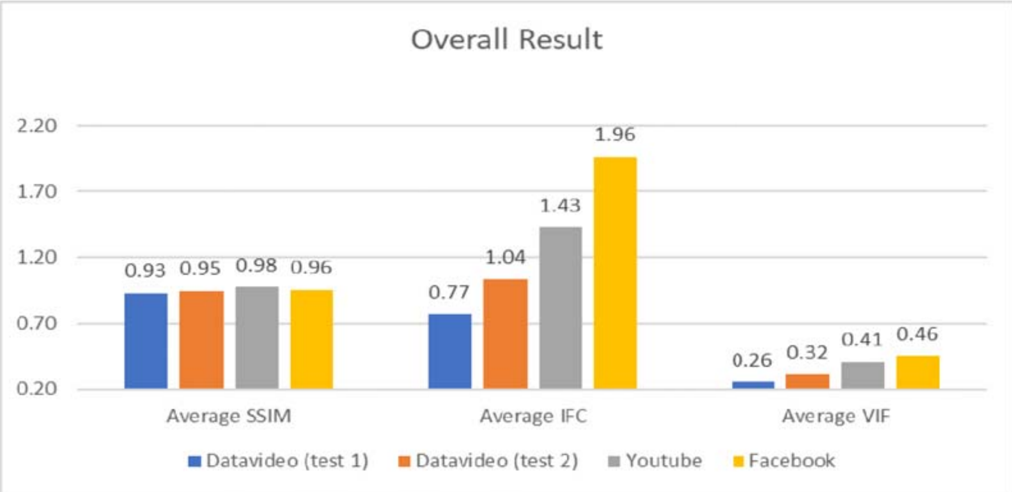

Figure 9: Bar chart showing overall SSIM, IFC and VIF values

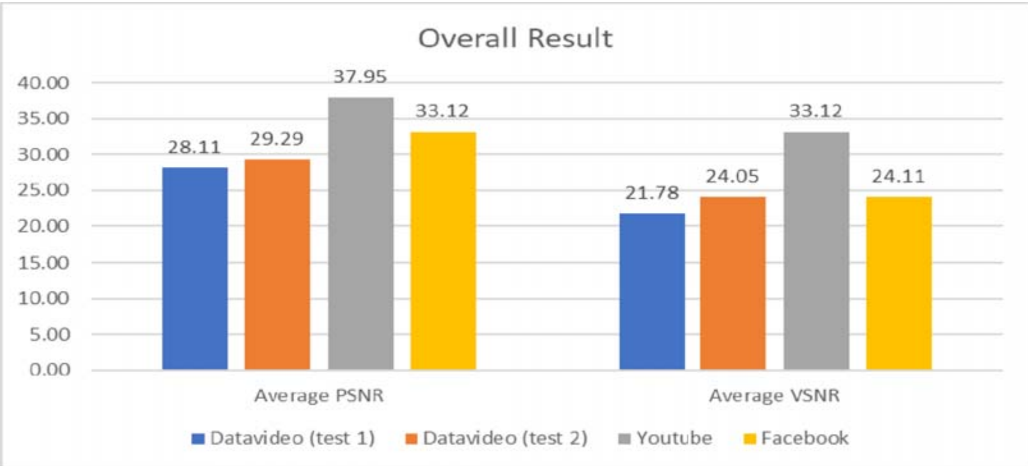

Figure 10: Bar chart showing overall PSNR and VSNR values

5. Conclusion

Video quality remains an important factor in all video related services especially for efficient multimedia applications in education. Through this study, we have been able to analyse the overall quality of videos when streamed using Datavideo as well as through YouTube and Facebook with objective video quality metrics. The results obtained shows that videos streamed across YouTube provide best quality over the other platform and device that were investigated in this study. Thus, YouTube can be posed to be suitable for multimedia applications in education

References

[1] V. R. Jonnalagadda, "Evaluation of Video Quality of Experience using EvalVid," no. July, 2012.

[2] C. Visual and N. Index, "Cisco Visual Networking Index : Global Mobile Data Traffic Forecast , 2017 – 2022," pp. 2017–2022, 2019.

[3] M. A. Usman, "Performance Evaluation of video Quality Assessment Methods Based on frames," no. April, 2014.

[4] R. I. Bt and B. T. S. Broadcasting, "Methodology for the subjective assessment of the quality of television pictures BT Series Broadcasting service," vol. 13, 2012.

[5] M. Shahid, A. Rossholm, B. Lövström, and H. Zepernick, "No-reference image and video quality assessment : a classification and review of recent approaches," pp. 1–32, 2014.

[6] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 660–668, 2008.

[7] A. M. Kaplan and M. Haenlein, "Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster," *Bus. Horiz.*, vol. 59, no. 4, pp. 441–450, 2016.

[8] B. Holmberg, *The evolution, principles and practices of distance education*, vol. 11. 2005.

[9] T. Wiegand, G. J. Sullivan, S. Member, G. Bjøntegaard, A. Luthra, and S. Member, "<Overview of the H.264 video coding standard.pdf>," vol. 13, no. 7, pp. 560–576, 2003.

[10] C. Gupta, *Data Communications and Computer Networks for Undergraduates*. 2006.

[11] J. You, U. Reiter, M. M. Hannuksela, M. Gabbouj, and A. Perkis, "Perceptual-based quality assessment for audiovisual services: A survey," *Signal Process. Image Commun.*, vol. 25, no. 7, pp. 482–501, 2010.

[12] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, 2006.

[13] D. M. Chandler and S. S. Hemami, "FR-TIP-VSNR : A Visual Signal-to-Noise Ratio for Natural Images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, 2007.