World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:17, No:1, 2023

Evaluation of Video Quality Metrics and Performance Comparison on Contents Taken from Most Commonly Used Devices

Pratik Dhabal Deo, Manoj P.

***Abstract*—**With the increasing number of social media users, the amount of video content available has also significantly increased. Currently, the number of smartphone users is at its peak, and many are increasingly using their smartphones as their main photography and recording devices. There have been a lot of developments in the field of video quality assessment in since the past years and more research on various other aspects of video and image are being done. Datasets that contain a huge number of videos from different high-end devices make it difficult to analyze the performance of the metrics on the content from most used devices even if they contain contents taken in poor lighting conditions using lower-end devices. These devices face a lot of distortions due to various factors since the spectrum of contents recorded on these devices is huge. In this paper, we have presented an analysis of the objective Video Quality Analysis (VQA) metrics on contents taken only from most used devices and their performance on them, focusing on full-reference metrics. To carry out this research, we created a custom dataset containing a total of 90 videos that have been taken from three most commonly used devices, and Android smartphone, an iOS smartphone and a Digital Single-Lens Reflex (DSLR) camera. On the videos taken on each of these devices, the six most common types of distortions that users face have been applied in addition to already existing H.264 compression based on four reference videos. These six applied distortions have three levels of degradation each. A total of the five most popular VQA metrics have been evaluated on this dataset and the highest values and the lowest values of each of the metrics on the distortions have been recorded. Finally, it is found that blur is the artifact on which most of the metrics did not perform well. Thus, in order to understand the results better the amount of blur in the data set has been calculated and an additional evaluation of the metrics was done using High Efficiency Video Coding (HEVC) codec, which is the next version of H.264 compression, on the camera that proved to be the sharpest among the devices. The results have shown that as the resolution increases, the performance of the metrics tends to become more accurate and the best performing metric among them is VQM with very few inconsistencies and inaccurate results when the compression applied is H.264, but when the compression is applied is HEVC, Structural Similarity (SSIM) metric and Video Multimethod Assessment Fusion (VMAF) have performed significantly better.

***Keywords*—**Distortion, metrics, recording, frame rate, video quality assessment.

I. Introduction

SOCIAL media consumption has been increasing constantly. World wide there are 3.8 billion social media users [9]. With the increase in social media consumption, photography and recording is also increasing. Almost 100 million hours of video content is consumed on Facebook every day, 500 hours of video are uploaded on YouTube every minute all over the world [10]. This makes it interesting to think, how well would the VQA metrics perform on these contents and which of them would be the best performing one. Video quality metrics are different types of algorithms that aim to predict how the viewers would perceive the video quality in real time. These metrics are used for several activities, such as comparing codecs and various other encoding configurations, assisting in production and live quality of experience. Therefore, if it is known that which metric would perform well in these contents, executing these tasks would be much more convenient and will provide better quality of experience.

There are several datasets available for VQA such as LIVE Video Quality Database (2010), in which the given conditions include MPEG-2 compression, H.264 compression, simulated transmission of H.264 compressed bit streams through error-prone IP wired and wireless networks [1], and Konstanz Natural Video Database (KoNViD-1k) (2017) which has 1200 videos with subjective data and attribute evaluation [2] and many more but they contain data taken from several high-end devices. This does not help our purpose since the aim is to evaluate the effects of common distortions on the contents taken from most widely used sensors. Also, with the enormous number of videos in particular datasets, it becomes difficult to analyze the content that is within the required spectrum. Hence, a dataset that has appropriate number of videos in it and has the required number of distortions from the most commonly used devices was needed to be constructed. Taking into consideration that smartphones are the most accessible camera, Android and iOS smartphones were used to create a part of the dataset. The other most accessible camera is a DSLR that many professionals or even enthusiasts use.

A clip of a few seconds was taken from each of these cameras and stored as the reference video for the analysis, since we are focusing on full reference metrics. Each of these clips have varying frame rates and different resolutions, which effects the details in the video.

Onto these original reference videos six different types of distortions were applied, after H.264 and H.265 compression was applied to them. The distortions which have been applied are the some of the most common types of distortions that video contents usually suffer from. The four reference videos are not

Pratik Dhabal Deo and Manoj P. are with HCL technologies LTD, India (e-mail: pratik1307xy@gmail.com, manoj.p@hcl.com).

steady and have considerable amount of shake which adds to the distortions in the videos. The videos are compared and the highest and lowest peaks of all the metrics are recorded. The recordings are done in manner of increasing levels of distortion and each metric is classified according to how they have performed on those distortions. Since it is known that with each level the distortion increases, the metrics should also provide results according to that. In case it does not provide matching results, it will be concluded that the metric has not been able to perform accurately on that particular artifact. It can also be possible that different metrics perform well on a particular distortion but poorly on the other distortion types. The devices used are an android phone (Redmi K20) which uses a SonyIMX582 sensor. The next device is an iOS device (iPhone 11) and then a Canon EOS200d mark ii DSLR.

II. Objective VQA

The best way to measure the quality of any video is by using subjective evaluation methods, since humans are the ultimate content consumers and know better about the quality of the content. The metric used in this evaluation is MOS (Mean Opinion Score), but this method is found to be expensive in terms of time and resources. Therefore, objective video quality metrics are widely used since they let the content creators and distribution organizations with means for making meaningful quality evaluations without convening viewer panels [3].

Objective video quality metrics can be classified according to the availability of the original image or video signal which is considered to be distortion-free or perfect quality and may be used as a reference to compare a distorted image or video signal against. They can be classified as Full-reference (FR), Reduced Reference (RR) and No-reference metrics.

FR metrics require the complete original video (reference video) for a frame wise comparison. This means that both of the videos (the original reference video and the video to test) should have the same properties and be spatially and temporally aligned. Any dissimilarities from the original video will amount to distortions. A few of the most used FR metrics are MSE (mean squared error), PSNR (Peak signal to noise ratio).

RR metrics do not require the complete video, but only a few particular features from the original video and then perform the evaluation on them.

NR metrics are those which do not need any reference video for the analysis of a video. The most common approach used in NR metrics is the estimation of artifacts and assessment of the information available in the bitstream of the video format [4].

A. Video Quality Metrics

The metrics selected for this experiment are VMAF, PSNR, SSIM, MSE and VQM since they are said to be ideal for objective quality assessment [11]. Most of the video quality metrics that are being used currently have been developed from the theory of image quality assessment (IQA) since videos are nothing but a continuous sequence of images. However, VQA involves distortions that are temporal, which do not apply on IQA. This causes the videos to be more distorted than images, yet spatial or image distortion still prevail in videos and affect the quality. The temporal and spatial effects combine to make the overall distortion more or less severe or visibly disturbing depending on when they appear in a video.

MSE: It is the average of the square of the errors. The larger the number the larger the error. There is no correct value for BMSE. Simply put, the lower the value the better and 0 means the model is perfect.

PSNR: It is a very commonly used video quality metric or a performance indicator. However, some studies have claimed that it has very poor correlation with subjective quality data, whilst others use it on the basis that is shows good correlation with subjective data [12]. A higher PSNR value indicates that the output quality is of good quality. For calculating the PSNR, MSE is used.

$$\text{MSE} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[f(i,j)-F(i,j)]^2}{M.N} \quad (1)$$

$$\text{PSNR} = 20.\log_{10}\left(\frac{255}{\sqrt{MSE}}\right) \quad (2)$$

Here f (i, j) is the original signal at pixel (i, j), F (i, j) is the reconstructed signal, and M x N is the picture size. The result is a single number in decibels, ranging from 30 to 40 for medium to high quality video. Even though many objective quality models have been developed in the past years, PSNR is still the most popularly used quality metric for evaluation of image and video contents.

SSIM: This is a widely used method to determine the perceived quality of digital television and cinematic pictures and other kinds of images and videos. It is used to measure the similarity between two images. SSIM is a perception-based model that considers image degradation as perceived change in structural information, while also incorporating important perceptual phenomena, including both luminance masking and contrast masking terms. The difference between SSIM and other methods like PSNR or MSE is that these methods estimate the absolute errors whereas structural information is the idea that the pixels that are spatially close in an image or a video have very strong inter-dependencies [5]. These dependencies contain important information about the structure of the objects in the frame. Luminance masking is a phenomenon whereby image distortions (in this context) tend to be less visible in bright regions, while contrast masking is a phenomenon whereby distortions become less visible where there is significant activity or "texture" in the image.

$$\text{SSIM} = \frac{(2\overline{xy} + C_1)\,(2\sigma_{xy} + C_2)}{[(\bar{x})^2 + (\bar{y})^2 + C_1]\,(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3)$$

Video Quality Metric: VQM is developed by ITS1 to provide an objective measurement for perceived video quality [13]. It measures the perceptual effects of video impairments including blurring, jerky/unnatural motion, global noise, block distortion and color distortion, and combines them into a single metric. The testing results show VQM has a high correlation with subjective VQA and has been adopted by ANSI as an objective video quality standard.

VMAF: This is a full reference-based metric that has implemented image metrics like Visual Information Fidelity (VIF), and Detail Loss Metric (DLM), put together with temporal differences between frames. The final score by VMAF is the output from a support vector machine (SVM) regressor [6]. VMAF score lies between 0-100 where the closer the value is to 100, the better the quality. VMAF framework also allows others to retain it for their individual use-cases with the inevitable outcome of losing the comparison with others if necessary.

III. Database

The evaluation of the metrics has been done on the video samples taken from SonyIMX582 sensor, iPhone 11 and a DSLR. This dataset contains a total of 90 video samples. Four reference videos have been recorded on these devices through which the distorted videos have been derived. The clips are subjected to six different distortions among which four are compression artifacts and two are due to bad sensor (camera) quality. Those distortions are: blur, fisheye, jitter, noise and brightness and contrast. The videos are also compressed using H.264 format and the device which was found to record the sharpest videos among all the devices was subjected to an additional H.265 compression, which was later used to create a separate dataset. Out of the four reference videos, three were taken in daylight and one was taken in low light in order to measure the performance of the metrics in low light as well.

H.264 Compression: It is a type of video compression standard available for high-definition videos. This compression is also referred to as MPEG-4 Part 10 and Advanced Video Coding (MPEG-4 AVC). H.264 is a block-oriented, compensation-based video compression standard, defining many tools, bitrates and resolutions. It also supports up to 8K ultra high definition. Any codec that is based on H264 compresses a digital video in a manner that it only requires half of the storage of MPEG-2. Using this compression, a codec can maintain the same video quality while using only half of the storage. H.264/AVC was created to pioneer a video standard that can deliver good video quality at lower bitrates than the existing standards without overly complicating the design in order to keep the implementation practical also comparatively inexpensive to implement. The H.264 is very flexible and can be applied onto many applications, networks and systems including those with varying bitrates and resolutions, broadcasts, storage etc. H.264 is adopted within many verticals and by many devices, from professional decoders to mobile devices and browsers [14]. The users are already accustomed to the high compression and the artifacts that tag along with it, yet there is limit to which the consumers can tolerate poor quality. This can have a significant negative impact in terms of revenue for the content providers.

H.265 compression: This is also known as HEVC (High Efficiency Video Coding) and was developed by JCT-VC [15]. This compression standard was created to double the compression efficiency of its predecessor AVC/H.264. HEVC performs better than its predecessor because it is able to define a bigger range of block sizes. For an instance HEVC can define blocks up to 64x64, but H.264 defines only up to 16x16.

The video compression artifacts can be divided into temporal and spatial artifacts. Artifacts are first categorized by whether they are time/sequence-based (temporal) or location-based (spatial) [7]. If the artifact is visible even when the video is paused, then it is a spatial artifact. If the artifacts are more visible while the video is playing, then it is likely to be a temporal artifact. The temporal artifacts can be jerk, flicker, noise, grain etc. A few examples of spatial artifacts are blur, blocking, fish-eye effect etc.

Blocking/Jitter (spatial): Blocking is known by several names, such as tiling, mosaicking, pixelating etc. It is basically the distortion which creates various tiny blocks like structures all over the video. It happens when a compressed video is streamed through a low bandwidth connection. During decompression, the output of certain blocks make the surrounding look similar to it and seem like large blocks. With the increase in the size of the display, the blocking usually becomes more visible (keeping the resolution same). However, if the resolution is increased, the blocking artifacts become much less visible, hence making the video look better.

Blurring (spatial): It happens when there is a loss of high spatial image frequency, usually in at sharp edges. It is generally known as “fuzziness” or “unsharpness”, since it makes the video look less sharp and appear out of focus.

Noise/Grain (temporal): Grainy videos are those in which you can see visible grain like disturbances all over the video. This is also referred to as noisy video. It can be caused by a lot of factors like bad production and low-quality settings while transcoding.

Fish-eye effect (spatial): Recordings from fish-eye lenses contain a certain kind of geometric distortion which results in the deformation of the video. This distortion makes the video (mostly the edges) look spherical. It is sometimes a needed effect, but it also happens sometimes due to bad lens corrections. This distortion can be corrected using prior information, such as calibration patterns and better lens design specifications.

IV. Evaluation and Results

A. Evaluation of Metrics on Sony IMX582 Sensor

A collection of frames from the dataset that has been created using this device is included below. There are 18 videos in six different categories of distortions. The distortions blur, contrast, brightness and fisheye are very visible even if the images are very small, but distortions like noise and jitter can only be visible if the video is in motion or if the video is watched in its original size. Since these are static frames reduced in size, jitter and noise are not very distinguishable. Each VQA metric is passed through the evaluation on these videos and the results have been recorded in the form of lowest value and highest value separated by ‘-’.

From results in Table I, it is obvious that, as the levels of video artifact increase, the value of MSE increases constantly showing a decrease in the video quality. The highest recordings of MSE are seen in brightness, contrast and fisheye, where in

brightness and contrast, the recordings of the lowest peak and the highest peak observed are 5844, 7729 in brightness and 5259, 8768 in contrast, respectively, which are very high numbers for MSE. MSE has performed well in all the metrics except for noise and blur where there have been some fluctuations in the highest and lowest peaks which seem to be decreasing with the increase in distortion levels.

Fig. 1 Dataset taken from the device with six different distortions (blur, contrast, brightness, fisheye, jitter, noise) and three degradation levels

TABLE I
ANALYSIS OF DATA USING MSE FOR SONY IMX582

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	336 - 1358	406-1275	481-1187
Brightness	1937-3367	3661-5478	5844-7729
Contrast	1377-4766	3121-6939	5259-8768
Fisheye	1948-2940	2027-3101	2167-3461
Jitter	245-1700	298-1656	379-1584
Noise	478-1698	434-1654	433-1968

TABLE II
ANALYSIS OF DATA USING PSNR FOR SONY IMX582

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	16 - 22	17 - 22.04	17.383 - 21.303
Brightness	12.487 - 15.282	10 - 12	9.2 - 10.46
Contrast	11.349 - 16.73	9.7 - 13.18	8.7 - 10.9
Fisheye	13.4 - 15.2	13.2 - 15.06	12.7 - 14.7
Jitter	16.13 - 22.34	15.93 - 23.385	15.8 - 24:23
Noise	15.832 - 21.3	15.9 - 21.7	15.18 - 21.7

As the numbers in Table II show, the metric does not fluctuate at different metrics and has a constant performance across all the artifacts, other than blur, where the value of PSNR increases with the increase in the level of distortion, which is inaccurate, and also in noise, jitter, where the highest peaks increase with the distortion levels.

The results show that there are a few fluctuations when it comes to handling blur and jitter since there is an inconsistent rise and drop in the lower peaks. Other than these two artifacts, the metric has performed well, with the increase in the level of distortion, the values deviate from 1 and move closer to 0.

TABLE III
ANALYSIS OF DATA USING SSIM FOR SONY IMX582

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	0 - 10	0 - 9.57	0 - 9.1
Brightness	9.9 - 87.3	10 - 81.7	10.3 - 74
Contrast	12.3 - 100	14.4 - 100	16.5 - 74.8
Fisheye	6.16 - 18.28	5.5 - 17.28	5.5 - 17.9
Jitter	4.2 - 15	0 - 14.6	0 - 14.4
Noise	9.8 - 54.8	10 - 41	7 - 32

TABLE IV
ANALYSIS OF DATA USING VMAF FOR SONY IMX582

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	0 - 10	0 - 9.57	0 - 9.1
Brightness	9.9 - 87.3	10 - 81.7	10.3 - 74
Contrast	12.3 - 100	14.4 - 100	16.5 - 74.8
Fisheye	6.16 - 18.28	5.5 - 17.28	5.5 - 17.9
Jitter	4.2 - 15	0 - 14.6	0 - 14.4
Noise	9.8 - 54.8	10 - 41	7 - 32

Table IV shows that VAMF has evaluated inaccurate results for all the artifacts other than blur which has a clear decrease in numbers. For contrast, there is a constant increase in the evaluated scores. This can be associated to the fact that at times, increasing the contrast can enhance the picture quality.

TABLE V
ANALYSIS OF DATA USING VQM FOR SONY IMX582

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	4.7 - 8.4	5.2 - 8.3	5.5 - 8.1
Brightness	5.9 - 12.5	7.8 - 13.9	9.5 - 15.07
Contrast	8.9 - 16.5	12.9 - 20	17.3 - 22.3
Fisheye	9.6 - 11.5	9.8 - 11.9	10 - 12
Jitter	4.1 - 9.5	4.6 - 9.3	5.1 - 9.2
Noise	5.7 - 10	5.6 - 12.1	5.9 - 14

As seen by numbers in Table V, the VQM provides accurate estimations for almost all the artifacts since as the level of the applied artifacts increases the values of VQM also increase. The closer the value is to 0, the smaller the amount of distortion. Blur and jitter have decreasing highest peaks, which is not correct.

Result

It can be concluded that VQM and SSIM have performed the best with almost all the results completely accurate other than a few fluctuations in the highest peaks or the lowest peaks. The resolution and the processing of the sensor defines a lot about an image/video quality, hence the metrics that did not perform well in this device might perform better on other devices. As for PSNR and VMAF, they were not able to handle many artifacts properly and produced many inaccurate results. And as for MSE, the values seem to be accurate according to the metric, i.e., the higher the value, the worse the quality is excluding the minor fluctuations in noise.

B. Evaluation of Metrics on iPhone 11

The iPhone 11 was used to take a similar video clip of a few seconds and the metrics were run on the video with video degradation of various types. Due to the increase in the resolution of the camera and a better image sensor in the camera module, it can be expected that a few metrics can perform better than the previous results.

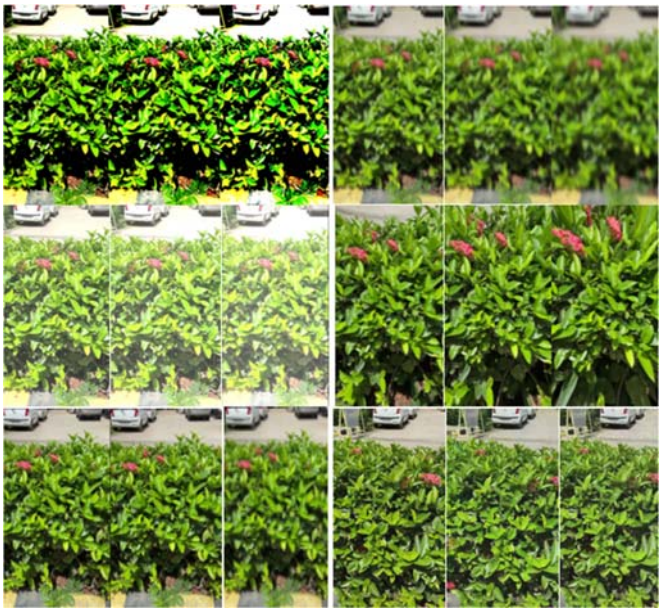

Fig. 2 Dataset taken from the device with six different distortions (contrast, blur, brightness, fisheye, jitter, noise and three degradation levels

TABLE VI
ANALYSIS OF DATA USING MSE FOR IPHONE 11

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	515 - 602	690 - 789	623 - 709
Brightness	1,841 - 4304	3,471 - 5,998	5,486 - 8,036
Contrast	1,252 - 5224	2,446 - 8,780	3,617 - 11,342
Fisheye	3,689 - 4,699	3,931 - 4,891	4,085 - 5,507
Jitter	351 - 450	445 - 545	663 - 773
Noise	78.5 - 375	129 - 722	211 - 2,195

As seen from Table VI, the MSE numbers keep increasing with the level of distortions. The evaluations from this device are significantly better than the evaluations from the previous device. The metric has produced accurate results through all of the artifacts since the numbers continue to rise with the level of distortions. Other than blur, where the numbers have fluctuated a little, MSE has performed better here.

TABLE VII
ANALYSIS OF DATA USING PSNR FOR IPHONE 11

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	20.3 - 21	19.1 - 19.7	19.6 - 20.1
Brightness	11.7 - 15.4	10.3 - 12.72	9.08 - 10.7
Contrast	10.9 - 17.15	8.69 - 14.2	7.58 - 12.54
Fisheye	11.41 - 12.46	11.2 - 12.18	10.7 - 12
Jitter	21.5 - 22.6	20.7 - 21	19.2 - 19.9
Noise	22.3 - 29.1	19.5 - 27	14.7 - 24.8

As according to Table VII PSNR has performed significantly better than the previous camera since this time jitter and noise have also been handled well, yet PSNR has evaluated wrong readings for blur since there are a few fluctuations in the highest and lowest peaks.

TABLE VIII
ANALYSIS OF DATA USING SSIM FOR IPHONE 11

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	0.61 - 0.66	0.6 - 0.65	0.66 - 0.69
Brightness	0.4 - 0.8	0.4 - 0.84	0.4 - 0.7
Contrast	0.2 - 0.6	0.2 - 0.54	0.19 - 0.45
Fisheye	0.38 - 0.44	0.36 - 0.44	0.3 - 0.4
Jitter	0.59 - 0.67	0.57 - 0.64	0.55 - 0.6
Noise	0.3-0.8	0.2-0.7	0.18 - 0.7

The scores evaluated by SSIM in the previous section had a few fluctuations for the artifacts blur and jitter, but as we can see here jitter has been handled by SSIM, however, blur and brightness have been mishandled as shown by the fluctuations in the readings.

TABLE IX
ANALYSIS OF DATA USING VMAF FOR IPHONE 11

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	0 - 14.3	0 - 16	100 - 100
Brightness	79.2 - 100	58 - 100	55.19 - 91.18
Contrast	78 - 100	28.6 - 100	28.4 - 93.8
Fisheye	3.74 - 30.4	3 - 29	2.6 - 29
Jitter	17.6 - 52.1	13.6 - 42	13.3 - 39.5
Noise	53.3 - 81.3	38.6 - 71	27.4 - 62.1

VMAF has performed a lot better in these data when compared to the previous section. Blur and brightness have been mishandled by VMAF and there a lot of 100 seen in the readings. This can be associated to the increase in the resolution of the camera as compared to the previous section.

TABLE X
ANALYSIS OF DATA USING VQM FOR IPHONE 11

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	5.7 - 7	6.3 - 7.5	6.8 - 8.2
Brightness	5. 8 - 12.9	7.6 - 14	9.41 - 15.04
Contrast	8.5 - 16	11.4 - 19.255	14.16 - 21.6
Fisheye	12 - 13.3	12.2 - 13.5	12.2 - 13.9
Jitter	5.05 - 6.6	5.6 - 6.9	6.3 - 7.8
Noise	2.6 - 5.5	3.4 - 7.7	4.13 - 10.2

From the readings of Table X, we can see that VQM has been able to handle all the artifacts properly with almost accuracy. As the distortion levels go high, the readings deviate from 0.

Result

In this section VQM is the best performing metrics since all the readings were accurate. MSE and PSNR were able to handle all the artifacts properly but were not able to produce accurate estimates for the blur artifact. SSIM handled artifact jitter better in this section than in the previous but could not handle blur as well as brightness properly. VMAF has also performed better when compared to the previous section but was not able to handle blur and brightness.

C. DSLR (Canon EOS 200d Mark ii)

This DSLR was used to record a clip of a few seconds and then similar distortions have been applied to them. The clips from the DSLR have occasional blurring out of objects since another object (the flowers) tried to be kept in focus.

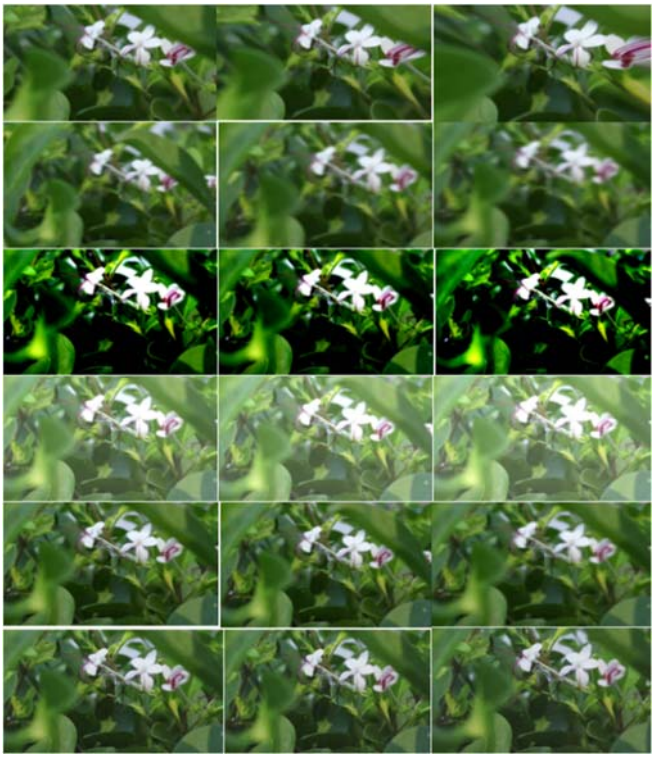

Fig. 3 Dataset taken from the device with six different distortions (fisheye, blur, contrast, brightness, jitter, noise) and three degradation levels

TABLE XI
ANALYSIS OF DATA USING MSE FOR DSLR

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	18 - 36	47 - 81	72 - 77
Brightness	1,826 - 1,836	2,615 - 2,632	3,537 - 3,569
Contrast	1,357 - 3,132	2,080 - 4,086	2,962 - 5,153
Fisheye	1,159 - 1,807	1,380 - 2,217	1,651 - 2,376
Jitter	20 - 37	31 - 51	59.34 - 90.1
Noise	31 - 1,087	63 - 1,711	78 - 935

As compared to the previous cameras the readings are lower in MSE, yet it was not able to handle blur and noise accurately since there are some fluctuations in the readings of the lowest and the highest readings.

TABLE XII
ANALYSIS OF DATA USING PSNR FOR DSLR

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	32.6 - 35.4	29 - 31.4	71 - 798
Brightness	15.4 - 15.5	13.9 - 13.9	12.6 - 12.6
Contrast	13.1 - 16.8	12 - 14	11 - 13.4
Fisheye	15.5 - 17.4	14.6 - 16.7	14.3 - 15.9
Jitter	32.3 - 34.9	30.9 - 33.2	18.5 - 30.3
Noise	17.7 - 33.1	17.4 - 30	18.4 - 29

PSNR has been able to perform accurately since as the videos get more degraded the PSNR value also decreases. Other than blur where there have been fluctuations in the readings of the metric, the metric has performed fairly well.

TABLE XIII
ANALYSIS OF DATA USING SSIM FOR DSLR

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	0.93 - 0.96	0.9 - 0.95	0.8 - 0.9
Brightness	0.89 - 0.91	0.86 - 0.88	0.88 - 0.86
Contrast	0.53 - 0.68	0.42 - 0.55	0.31 - 0.8942
Fisheye	0.78 - 0.84	0.77 - 0.83	0.77 - 0.83
Jitter	0.87 - 0.91	0.86 - 0.9	0.83 - 0.89
Noise	0.13 - 0.84	0.1 - 0.7	0.07 - 0.75

SSIM has been able to perform very well in all the artifacts since with degradation, the readings deviate from 1. There exists a minor fluctuation in noise with the highest peaks.

TABLE XIV
ANALYSIS OF DATA USING VMAF FOR DSLR

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	0 - 21	0 - 14	0 - 10
Brightness	94.18 - 100	91 - 100	89 - 100
Contrast	100 - 100	100 - 100	100 - 100
Fisheye	0 - 28	0 - 24	0 - 21
Jitter	3.7 - 43	0.68 - 38	0 - 32
Noise	53 - 79	50 - 75	42 - 68

VMAF has performed accurately for the samples in this device as the numbers keep decreasing as distortions keep increasing. It is also seen that the number of 100 present in the readings are higher than the previous sections.

TABLE XV
ANALYSIS OF DATA USING VQM FOR DSLR

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	2.3 - 3.6	2.4 - 3.8	2.8 - 6.0
Brightness	5.3 - 5.5	6.2 - 6.5	7.1 - 7.3
Contrast	7.6 - 12.8	9.5 - 14.4	11.2 - 16
Fisheye	7.1 - 7.8	7.3 - 8	7.5 - 8.5
Jitter	2.1 - 3.6	2.5 - 3.8	3.2 - 6.3
Noise	1.6 - 8.4	2.2 - 9.2	2.3 - 9.1

VQM has performed consistently through all the artifacts, since the readings increase as the distortion increases.

Result

VQM has been able to perform consistently throughout all the artifacts. SSIM performs well, other than minor fluctuations in noise. MSE also had a few fluctuations in noise and blur. PSNR also manhandles blur.

D. Sony Imx582 (Low-Light)

The same Android device was used to record clips in low light at 720p resolution. The distortions seem to be more severe on these clips when compared to the previous clips. Sample images of the complete dataset are not included since due to low-light conditions; it would be difficult to estimate the distortions unless the original picture frames with the original resolutions are provided.

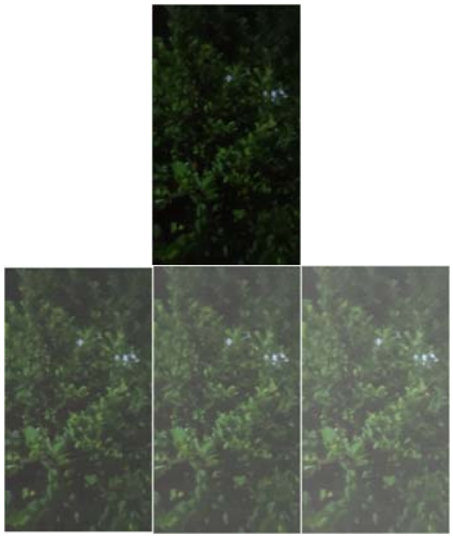

Fig. 4 Original frame with the sample frames demonstrating the increasing degradation as brightness level increases

TABLE XVI
ANALYSIS OF DATA USING MSE SONY IMX582 LOW LIGHT

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	3.5 - 18.9	5.8 - 26.7	8.6 - 33.4
Brightness	1226 - 1229	2406 - 2401	3978 - 3986
Contrast	534 - 620	604 - 781	641 - 876
Fisheye	87.3 - 147	96 - 203	89 - 304
Jitter	2.5 - 13.1	4 - 19.7	6.3 - 26
Noise	48 - 269	221 - 729	119 - 488

As seen from the readings of XVI, MSE has evaluated all the artifacts correctly, except for noise and fisheye where a few fluctuations are noticed for the highest and lowest recordings.

TABLE XVII
ANALYSIS OF DATA USING PSNR SONY IMX582 LOW LIGHT

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	35.3 - 42.6	33.8 - 40.4	32.8 - 38.7
Brightness	17.2 - 17.24	14.3 - 14.3	12.1 - 12.1
Contrast	20.1 - 20.8	19.2 - 20.1	18.7 - 20
Fisheye	26.4 - 28.7	25 - 28.3	23.2 - 28.5
Jitter	36.9 - 43.9	35.1 - 42	33.9 - 40
Noise	28.8 - 31.2	21.2 - 27.3	19.1 - 24.6

PSNR performs well in all of the artifacts other than a few fluctuations in fisheye since as the levels of distortions increase, the value decreases constantly, while in fisheye the values fluctuate by a minor value.

TABLE XVIII
ANALYSIS OF DATA USING SSIM SONY IMX582 LOW LIGHT

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	0.95 - 0.93	0.88 - 0.97	0.87 - 0.96
Brightness	0.7 - 0.79	0.6 - 0.7	0.6 - 0.6
Contrast	0.3 - 0.39	0.2 - 0.29	0.25 - 0.26
Fisheye	0.75 - 0.9	0.76 - 0.9	0.7 - 0.9
Jitter	0.92 - 0.97	0.89 - 0.96	0.87 - 0.95
Noise	0.25 - 0.83	21.2 - 27.3	19.14 - 24.6

SSIM has not been able to perform consistently throughout the artifacts. There are fluctuations in the artifacts blur, contrast, fisheye and SSIM has evaluated accurate results for only for the rest of the three artifacts.

TABLE XIX
ANALYSIS OF DATA USING VMAF SONY IMX582 LOW LIGHT

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	74 - 98.9	72 - 100	74 - 100
Brightness	93.8 - 100	93.9 - 100	93.1 - 100
Contrast	49 - 60	40 - 54	31 - 51
Fisheye	1.1 - 19.8	0 - 12	0 - 18
Jitter	1.1 - 19.8	0 - 12.9	0 - 18.1
Noise	28.2 - 68.3	13 - 53	6.3 - 35

VMAF has evaluated only the artifacts- contrast and noise accurately, the evaluations for the rest of the artifacts are inaccurate since there are several fluctuations in the highest and the lowest readings of the metrics.

TABLE XX
ANALYSIS OF DATA USING VQM SONY IMX582 LOW LIGHT

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	1.1 - 2.22	1.4 - 2.7	1.6 - 3.1
Brightness	6.4 - 6.6	8.5 - 8.7	10.5 - 10.7
Contrast	7.5 - 8	8.2 - 9	8.7 - 9.6
Fisheye	3.6 - 5.9	3.9 - 6.1	4.5 - 6.6
Jitter	0.8 - 1.7	1.1 - 2.4	1.3 - 2.9
Noise	2.4 - 7.3	3.7 - 9.6	4.5 - 6.6

VQM has performed accurately on all the artifacts since as the distortions increase, the values deviate from zero.

Results

In low light conditions, the most mishandled artifact was found to be fisheye, where four out of the five metrics have evaluated wrong readings. The best performing metrics are found to be PSNR and VQM.

V. BLUR CALCULATION AND EVALUATION ON HEVC

A. Blur Calculation

It is seen that the most difficult artifact to handle by the majority of the VQA metrics is blur since almost all the metrics have evaluated wrong results in the sample inputs that contained the artifact blur. In order to understand the results produced by the VQA metrics, the amount of blur on each of the videos was calculated. The amount of blur is calculated using the variance of the Laplacian method. The Laplacian method works by highlighting the areas of an image containing severe intensity changes, similar to the Sobel and Scharr operators. Just like these operators, the Laplacian method, which is very basic, is often used for detecting edges to identify blur since we know that the number of edges visible in a blurry image is less and decreases even further with the increase of blur. It is assumed that if there is high variance in the image, then there is a wide spread of responses, both edge-like and non-edge like, representative of a normal, in-focus image. But in the case the variance is low, there is a tiny spread of responses, indicating there are very few edges in the image. The final value of the variance is based on the threshold value. Below a

particular threshold value, the image will be classified as blurry, otherwise the image is not blurry.

TABLE XXI
AMOUNT OF BLUR PRESENT IN THE DATA SAMPLES

	IMX582(Low light)	IMX582	iPhone	DSLR
Original	115.2	2012	3678	72.98
Blur1	46.7	14.04	21.69	6.08
Blur2	35.25	10.08	16.38	6.08
Blur3	29.24	7.28	10.67	5.62

It is seen that the original videos have high values of variance, and as the blur is applied onto the frames of video, the variance decreases. The sensor of the iPhone has been calculated to have the highest variance, meaning that the iPhone's recording has the sharpest and most clear edges. Apart from the quality of the recording device, another reason behind these results can be that, the recording using this device was done in bright daylight conditions which helped in recording sharper videos. The DSLR has been calculated to have the lowest of all variances since it had a main object in focus; this made the background blurred, while keeping the object in focus.

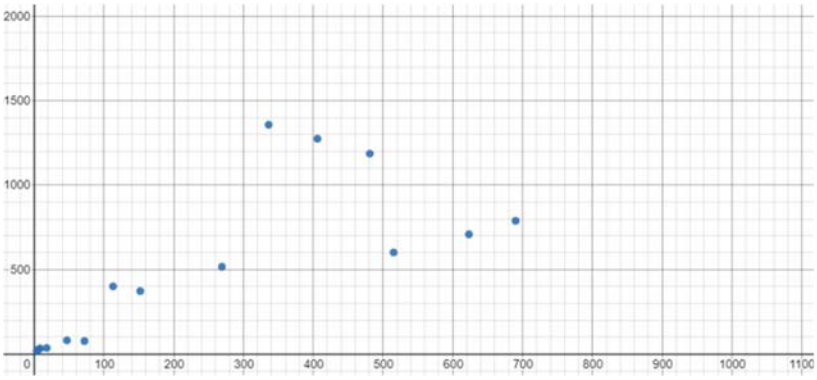

Fig. 5 Blur points under MSE of the complete dataset

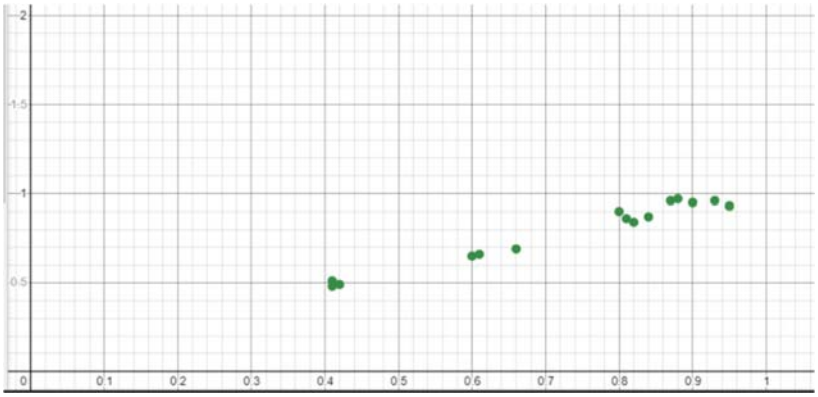

Fig. 6 Blur points under SSIM of the complete dataset

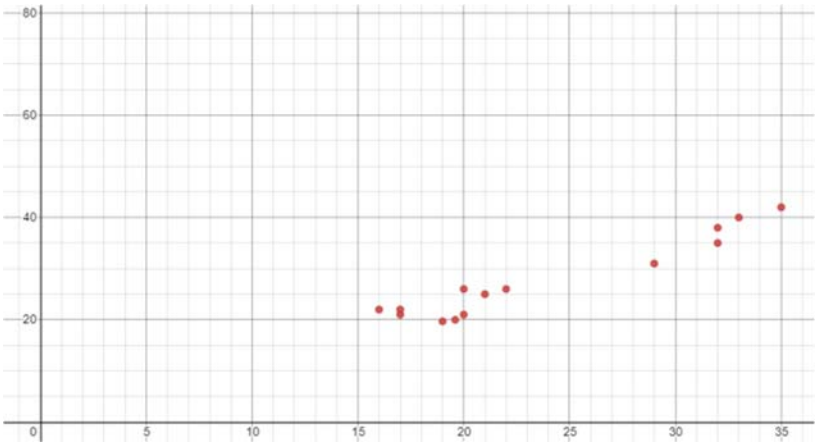

Fig. 7 Scatter plot of Blur under PSNR for complete dataset

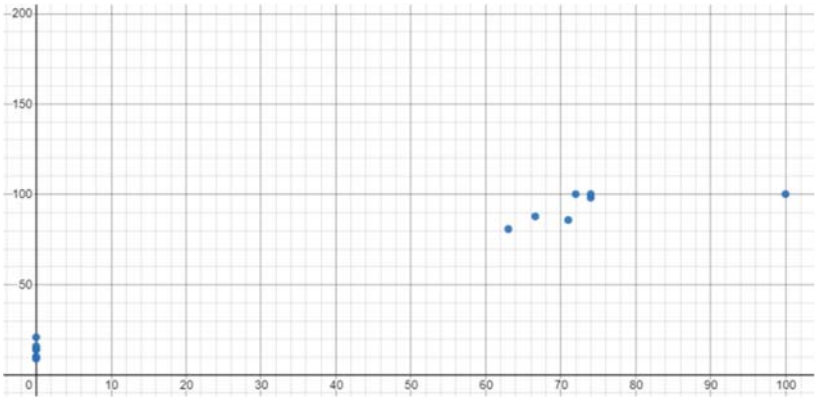

Fig. 8 Scatter plot of Blur under VMAF for complete dataset

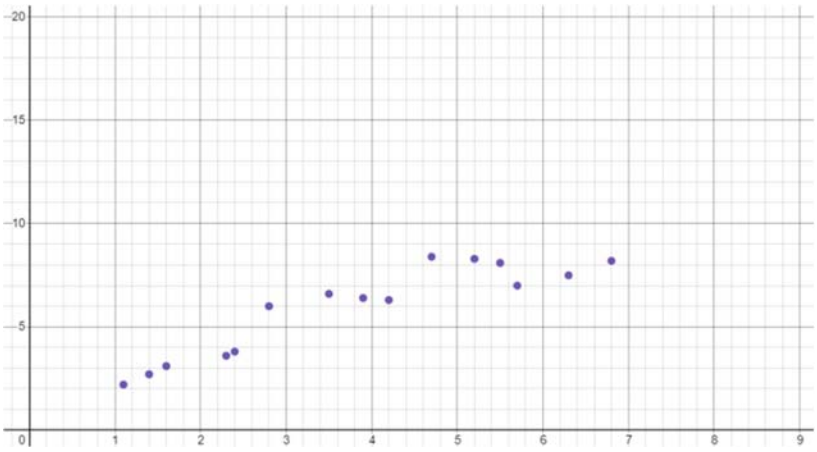

Fig. 9 Scatter plot of Blur under VQM for complete dataset

B. HEVC Evaluation

Since the iPhone 11 was found to have the sharpest and clearest of edges, we can now apply HEVC compression to the samples and know its effects. This is useful since HEVC is the successor of H.264 and it is important to know if the successor has any significant advantages over its predecessor when it comes to contents not taken under best suited conditions.

TABLE XXII
ANALYSIS OF DATA USING MSE FOR HEVC

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	113.5 - 401.09	152.6 - 373.3	269.4 - 517.2
Brightness	583.8 - 1,389.2	1118 - 1921	1,752 - 2569
Contrast	391 - 1738	780.7 - 2308.5	1167 - 2807
Fisheye	1202 - 1493	1283 - 1562	1319 - 1749
Jitter	113 - 667	143 - 672	212.5 - 605.4
Noise	38 - 756	82 - 753	152 - 780

TABLE XXIII
ANALYSIS OF DATA USING PSNR FOR HEVC

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	20.5 - 26	22.4 - 26	21 - 24
Brightness	16.7 - 20.4	15.2 - 17.6	14 - 15.6
Contrast	15.7 - 22.2	14.5 - 19.2	13.6 - 17.4
Fisheye	16.4 - 17.3	16.1 - 17	15.7 - 17
Jitter	20 - 7.6	20 - 26.5	20.3 - 25
Noise	19. 3 - 2.3	19.3 - 29	19.2 - 26.3

As seen from Table XXII, there are minor fluctuations in the artifacts blur, jitter and noise whereas in the H.264 evaluation, error was only obtained in the blur artifact.

PSNR has minor fluctuations when it evaluated blur and jitter. The H.264 evaluation also had fluctuations in blur but evaluated the other artifacts correctly.

TABLE XXIV
ANALYSIS OF DATA USING SSIM FOR HEVC

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	0.81 - 0.86	0.84 - 0.87	0.82 - 0.84
Brightness	0.77 - 0.96	0.76 - 0.95	0.76 - 0.93
Contrast	0.73 - 0.9	0.72 - 0.85	0.72 - 0.83
Fisheye	0.76 - 0.77	0.76 - 0.77	0.76 - 0.77
Jitter	0.8 - 0.87	0.8 - 0.86	0.8 - 0.83
Noise	0.81 - 0.86	0.84 - 0.87	0.82 - 0.84

TABLE XXV
ANALYSIS OF DATA USING VMAF FOR HEVC

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	63.6 - 81	66.6 - 88.2	71.2 - 86.7
Brightness	87.7 - 100	83.8 - 95.6	77.3 - 88.4
Contrast	4.7 - 30	0 - 12.6	0 - 4.7
Fisheye	0 - 25	0 - 23	0 - 21
Jitter	39 - 54.1	24.7 - 40.5	18.4 - 34.6
Noise	83.5-95.8	75.6-91	67.3-85.4

In the H.264 evaluation, SSIM had mishandled blur and

brightness both, but in HEVC the artifact brightness has been evaluated correctly and only blur contains fluctuations.

VMAF has performed better in this section than its H.264 counterpart since only blur contains fluctuations and the other artifacts have been evaluated well; whereas in H.264, both blur and brightness were mishandled.

VQM has performed consistently across all the data with H.264 compression, but in HEVC data, VQM has not evaluated blur properly.

TABLE XXVI
ANALYSIS OF DATA USING VQM FOR HEVC

Distortion Type	Distortion level 1	Distortion level 2	Distortion level 3
Blur	3.5 - 6.6	3.9 - 6.4	4.2 - 6.3
Brightness	3.1 - 9	4.1 - 9.3	5.1 - 9.6
Contrast	5 - 11	6.6 - 11.7	8.1 - 12.5
Fisheye	8 - 8.7	8.1 - 9	8.1 - 9
Jitter	3 - 7.2	3.2 - 7.1	4 - 6.9
Noise	1.7 - 7.5	2.3 - 7.5	3 - 7.5

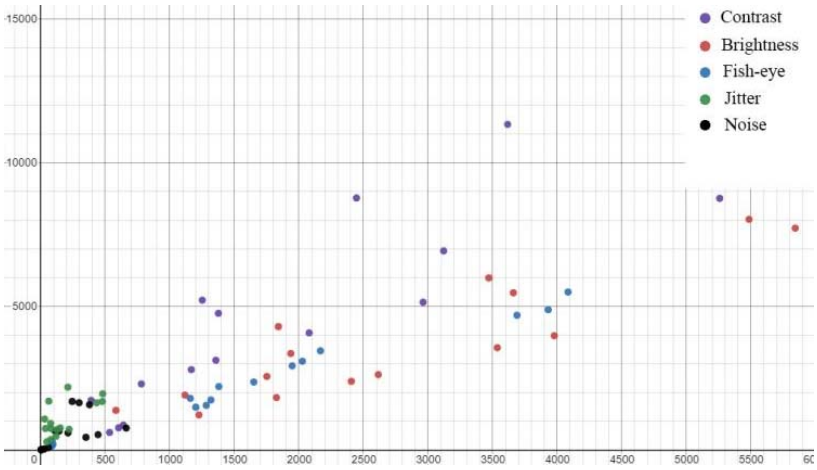

Fig. 10 Scatter plot of brightness, contrast, fisheye, jitter, noise under MSE for complete dataset

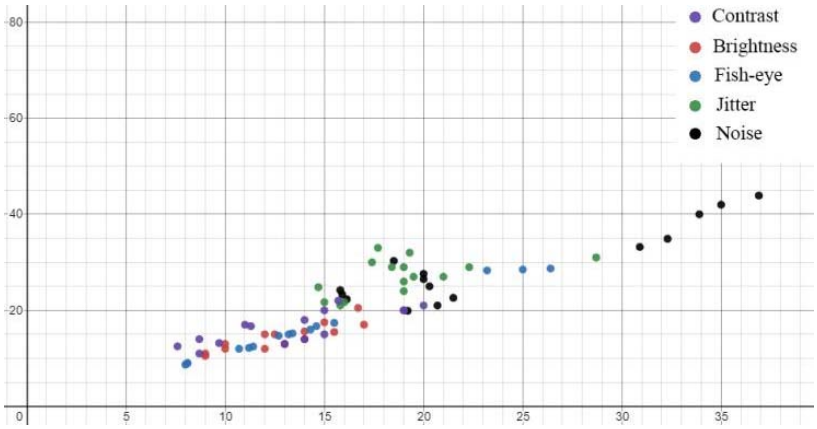

Fig. 11 Scatter plot of brightness, contrast, fisheye, jitter, noise under PSNR for complete dataset

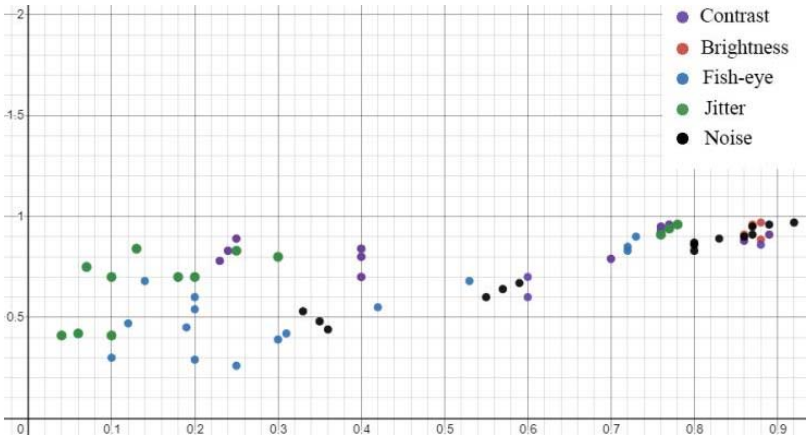

Fig. 12 Scatter plot of brightness, contrast, fisheye, jitter, noise under SSIM for complete dataset

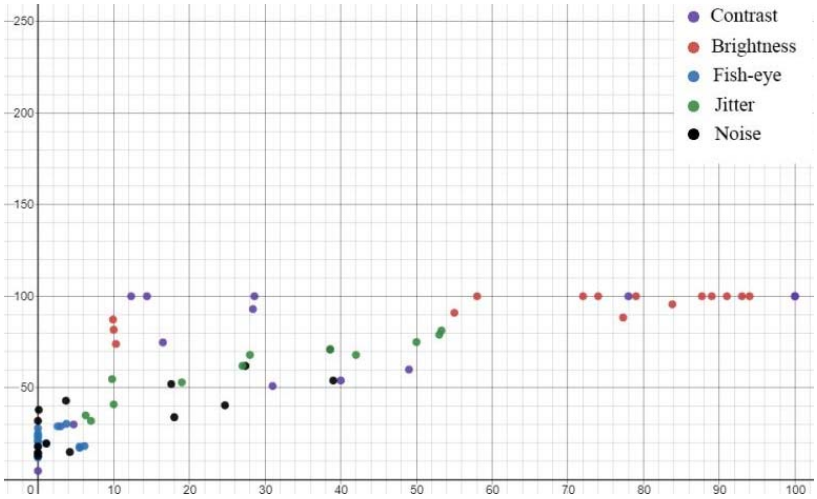

Fig. 13 Scatter plot of brightness, contrast, fisheye, jitter, noise under VMAF for complete dataset

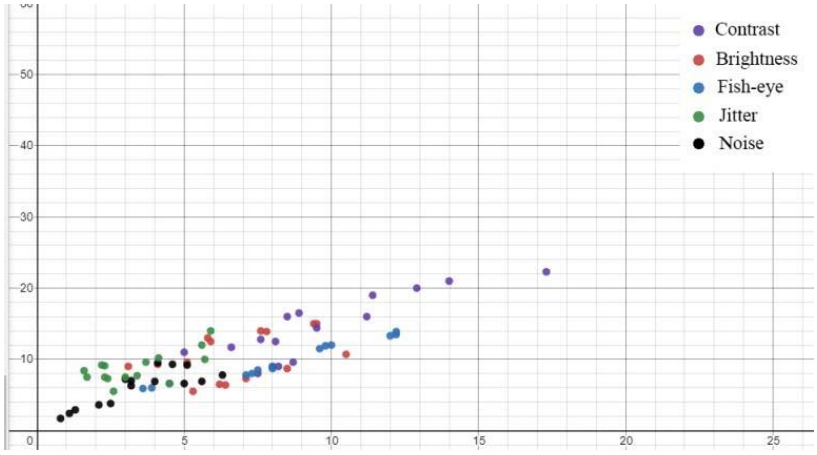

Fig. 14 Scatter plot of brightness, contrast, fisheye, jitter, noise under MSE for complete dataset

Results

The metrics SSIM and VMAF have performed comparatively better in this compression and other metrics have experienced minor fluctuations in their readings. VQM, which has been the best performing metric under H.264 compression, did not perform similarly under HEVC. Similar to the H.264, even here blur is the most mishandled artifact.

VI. Conclusion

Objective video/image quality metrics are used for many purposes such as to optimize algorithms and parameter settings of video processing systems, dynamically monitor and adjust the quality, to benchmark video processing systems and algorithms, and to compare two video systems solutions etc. Therefore, knowing which metric has the best performance under particular conditions helps a lot to ensure that the tasks are completed with utmost efficiency.

On testing the VQA metrics on contents taken from most commonly used cameras, it was found that the metric VQM outperforms all of the well-known metrics with very little or no fluctuations in the evaluations. MSE has performed well, other than in the case of blur and noise. MSE had inaccurate readings in all of the blur samples and two of the sample collections of noise. PSNR has also performed inaccurately for all the samples that contain blur. For the initial test conditions, PSNR failed in evaluating jitter and noise but has performed well in other artifacts. SSIM failed in the first two test conditions of blur and in the initial conditions of brightness, jitter and noise. VMAF did not perform well in the initial test conditions of blur and brightness but has performed significantly better than other metrics and performs better with the increase of resolution and quality of the devices. It is observed that the most difficult artifact that metrics find complex to handle is blur, where the majority of the metrics provided inaccurate readings. As the recording device improved, the number of metrics mishandling the blur artifact reduces, yet blur remains the most mishandled artifact. The second most inaccurately evaluated artifact is noise where three of the metrics recorded fluctuating readings. The next mishandled artifact is jitter where only in the initial test conditions did the metrics mishandle the artifact. The last mishandled artifact is brightness where just two metrics

mishandled the artifact in the initial test conditions. It is observed that the contrast and fish-eye artifacts have been handled very well by all of the metrics. It is also seen that VMAF performs strangely in terms of artifacts that deal with the dynamic range or the color corrections of the video, like contrast. The reason for this is that increasing or adjusting the contrast sometimes makes the contents looks better. VMAF is found to perform inaccurately when it comes to color correction [8]. Finally, the amount of blur in each recording was calculated by applying the variance of the Laplacian method, where it was found that the content recorded on the iPhone has the sharpest edges. After the application of HEVC compression on the same dataset, the most consistently performing metric, VQM, failed to assess all the distortions accurately. Hence, it can be concluded that if the contents to be evaluated are compressed using H.264, then using VQM will provide the best results; however, if the content is compressed using HEVC, SSIM and VMAF are better choices to provide the best results.

REFERENCES

[1] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, "Study of Subjective and Objective Quality Assessment of Video", IEEE Transactions on Image Processing, vol.19, no.6, pp.1427-1441, June 2010. W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[2] Vlad Hosu; Franz Hahn, Mohsen Jenadeleh; Hanhe Lin, Hui Men, Tamás Szirányi, Shujun Li "The Konstanz natural video database (KoNViD-1k)" IEEE, 3 July 2017

[3] Miguel O. Martínez-Rach,Pablo Piñol, Otoniel M. López, Manuel Perez Malumbres, José Oliver,and Carlos Tavares Calafate 2014

[4] Shahid, M., Rossholm, A., Lövström, B. et al. No-reference image and video quality assessment: a classification and review of recent approaches. J Image Video Proc 2014, 40 (2014).

[5] Sara, U., Akter, M. and Uddin, M. (2019) Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. Journal of Computer and Communications, 7, 8-18. doi: 10.4236/jcc.2019.73002.

[6] Mariela Fiorenzo, Claudio Righetti, María Cecilia Raggio, Fernando Ochoa & Gabriel Carro, Telecom Argentina S.A, 2019 SCTE.ISBE

[7] Read, Dwight. (2013). Artifact Classification: A Conceptual and Methodological Approach.

[8] Jnastasia Antsiferova, Dmitriy Vatolin, Dmitriy Kulikov, Sergey Zvezdakov "Hacking VMAF with Video Color and Contrast Distortion" july 2019

[9] We are social 2020, *We are social*, accessed 18 October, <https://wearesocial.com/uk/blog/2020/01/digital-2020-3-8-billion-people-use-social-media/>

[10] Vox 2016, *Vox*, accessed 20 October, <https://www.vox.com/2016/1/27/11589140/facebook-says-video-is-huge-100-million-hours-per-day-huge>

[11] Streaming media 2019, *streaming media*, accessed 20 October, <https://www.streamingmedia.com/Articles/Editorial/Featured-Articles/Buyers-Guide-to-Video-Quality-Metrics-130675.aspx?utm_source=related_articles&utm_medium=gutenberg&utm_campaign=editors_selection>