Video Compression Study
h.265 vs h.264

Jonathan E. Pryor

Abstract—**H.265 video compression (also known as High Efficiency Video Encoding (HEVC)) promises to provide double the video quality at half the bandwidth, or the same quality at half the bandwidth of h.264 video compression [1]. This study uses a Tektronix PQA500 to determine the video quality gains by using h.265 encoding. This study also compares two video encoders to see how different implementations of h.264 and h.265 impact video quality at various bandwidths.**

Index Terms—**HEVC, h.264, h.265, video, compression, encoder.**

I. INTRODUCTION

VIDEO encoding is used to reduce the bandwidth that raw video requires for transmission. It also reduces the file sizes of stored video. Compression methods such as h.264 and h.265 use prediction of moving objects between certain frames as a way to compress videos and reduce bandwidth. H.265 uses Adaptive Motion Vector Prediction for inter-frame prediction [1]. It should be noted that h.265 is a more computationally expensive compression method, which is likely the reason that it has not yet become as ubiquitous as h.264. Another type of compression is JPEG 2000. This is a wavelet based compression that uses the Discrete Cosine Transformation (DCT) [2]. This type of compression works on individual frames rather than predicting frames. This is a good option for high-bandwidth situations, but may not provide acceptable quality at low bandwidths. VP8 and VP9 are alternate encoding algorithms which are competitors with h.264 and h.265 respectively.

This study investigates only h.264 and h.265 compression using two different implementations of both h.264 and h.265. Two different hardware encoders were used in this study. For the purposes of this publication, they will be referred to as Encoder A and Encoder B. Each encoder is able to utilize both h.264 and h.265 algorithms. Encode A uses an ARM processor with dedicated hardware video encoders. Encoder B uses an FPGA to perform video encoding.

A. Types of Video Analysis

There are multiple ways to measure video quality. All methods compare compressed video to the reference uncompressed video. There are noise-based measurements which calculate the signal-to-noise ratio of the compressed video to the reference video, and there are perceptual based measurements, which take into account human perception to generate a rating. Perceptual based measurements are the preferred method of

Jonathan E. Pryor is employed by NASA Marshall Space Flight Center and attends the University of Alabama in Huntsville, Huntsville, Al, USA e-mail: jonathan.pryor@nasa.gov.

rating a compression algorithm because noise-based ratings can be misleading. Take the example provided by Tektronix.

Fig. 1. Signal-to-Noise Example provided in Tektronix Documentation [3]

The photo on the left looks more pleasing to the eye. However, this photo actually has a lower signal-to-noise ratio.

B. Picture Quality Analyzer

The Tektronix PQA500 is a system which performs video quality analysis [4]. By providing a reference and compressed video, it can generate a range of measurements. The measurements of interest are Picture Signal to Noise Ratio (PSNR) and Picture Quality Rating (PQR). The focus of this study will be the PQR ratings generated using h.264 and h.265 compression.

C. Picture Quality Rating

PQR values provide a good indication of perceived video quality. They were developed using a 19th century method of determining Just Noticeable Differences (JND) [3]. In the case of the PQA500, 1 JND = 1 PQR. To have a better understanding of what the PQR rating means to the viewer, the table below provides a summary of what the PQR ratings practically mean.

PQR/JND	Meaning
0	Reference and test images identical
<1	Viewers cannot distinguish between videos
1	Viewers can barely distinguish differences between videos
2-4	Viewers can distinguish between videos. Good to excellent quality
5-9	Viewers can easily distinguish between videos. Good to fair quality
>10	Obvious differences. Poor to bad quality

TABLE I
SUMMARY OF PQR INTERPRETATION ACCORDING TO TEKTRONIX
DOCUMENTATION [3]

II. Factors

There are 4 factors involved in this study. Each possible factor is shown below. (table 2) There are two algorithms being explored; h.264 and its successor, h.265. There are two encodes, as mentioned in the Section 1. There are 5 bandwidths being tested, and there are a total of 6 representative scenes that are of interest.

III. Methods

Every combination of factors was tested. Each encoder compressed every scene at every bitrate using both h.264 and h.265 compression. Each and every video was played through the PQA500 along with its reference video. Data is generated and saved in a Comma Separated Variable (CSV) file. For each video, a CSV file is generated for the PSNR values and PQR values. The way the data is generated, there is a PSNR and PQR value for every frame of the video. The figures below show what the output data looks like. These particular figures show both PSNR and PQR data for Encoder A using both its h.264 and h.265 encoding hardware. Figure 2 shows the results at 4 Mbps, and Figure 3 shows the results at 12 Mbps.
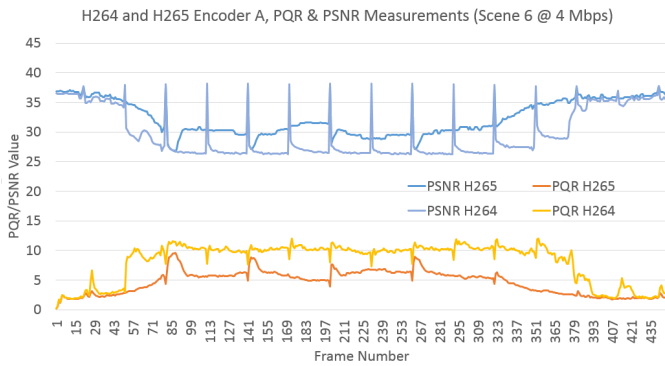

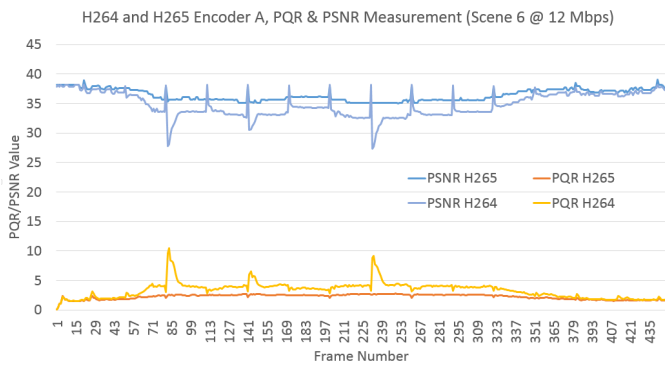Fig. 2. PQA500 Output, 4 Mbps

Fig. 3. PQA500 Output, 12 Mbps

By simply observing the PQR values of these figures, it is immediately apparent that the claim of h.265 being double the quality at the same bandwidth is a plausible and testable claim.

IV. 2^K Analysis

Because there are 4 factors in this study, it is possible to perform a 2^4 analysis. This type of analysis will reveal the most significant factors in this study and the significance of interactions between factors. Two possibilities for each factor was selected as shown in Table II.

Factor	Level -1	Level 1
Algorithm, A	H264	H265
Hardware, B	Encoder A	Encoder B
Bit Rate, C	1 Mbps	4 Mbps
Scene, D	Scene 1	Scene 4

TABLE II
FACTORS SELECTED FOR 2^4 ANALYSIS

	Encoder A				Encoder B			
	Scene 1		Scene 4		Scene 1		Scene 4	
Algorithm	1 Mbps	4 Mbps	1 Mbps	4 Mbps	1 Mbps	4 Mbps	1 Mbps	4 Mbps
H264	5.68	3.04	7.97	6.97	5.68	2.76	7.01	5.22
H265	3.34	1.93	6.58	5.28	3.04	1.97	5.75	4.69

TABLE III
PQR RESULTS FOR 2^4 ANALYSIS

The results of each combination of factors are shown in Table III. The procedure to arrive at this result was taken from Chapter 17 of [5]. The PQR values in this table represent the average PQR values of each scene. This average is generated in the same CSV results file from the PQA500.
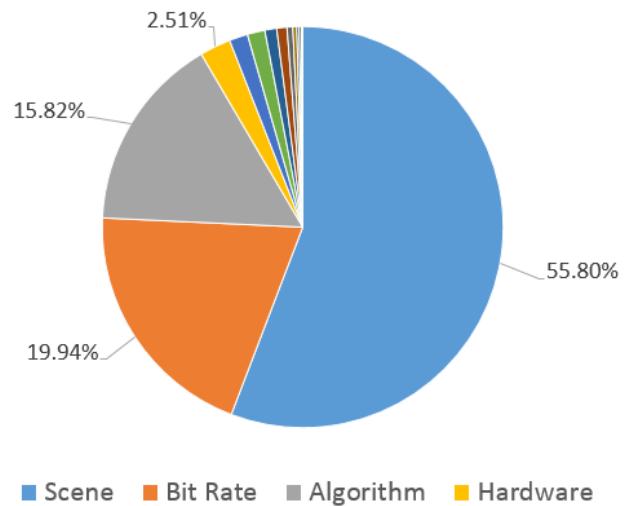

Fig. 4. Significance Results for 2^4 analysis

The results of this analysis are summarized in Figure 4. It is interesting that the scene has the largest overall impact on the PQR results. The scenes selection is followed by bitrate, algorithm, and lastly, the hardware selection.

This means that the hardware used to encode the video has very little impact on the PQR results. Choosing Encoder A or

Encoder B is the least important selection. Choosing a higher bitrate and selecting an h.265 encoder has a much bigger impact. It is likely that the scene may be pre-determined, but it is important to choose a representative scene when taking PQR measurements because it has such a profound impact on the results.

V. SCENE RESULTS

All of the scenes show different results. The figures in this section contain all of the results obtained in this study. The Y-axis shows the average PQR values, while the X-axis represents bitrates (1 Mbps, 2 Mbps, 4 Mbps, 8 Mbps, and 12 Mbps). Encoder A and Encoder B are represented for each of their algorithms (h.264 and h.265). The graphs also include horizontal bars to represent the quality levels outlined in Table I. This makes it easier to determine what quality of video is obtained for each encoder using each algorithm.
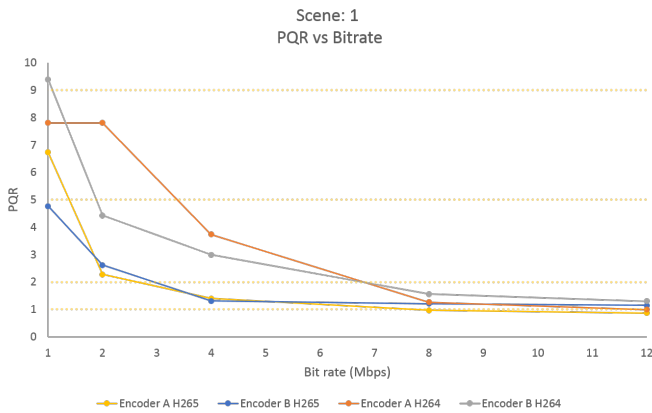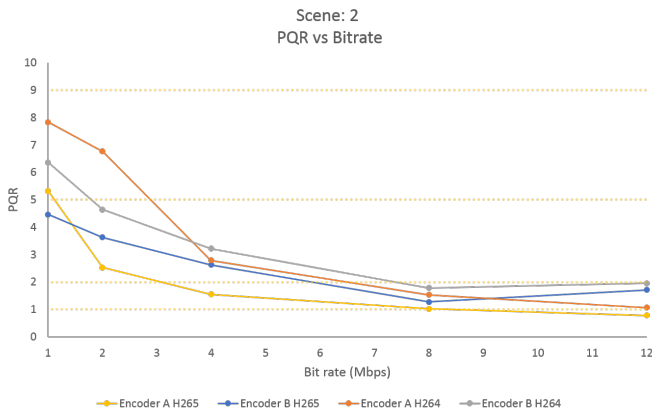
Fig. 5. Scene 1 Results
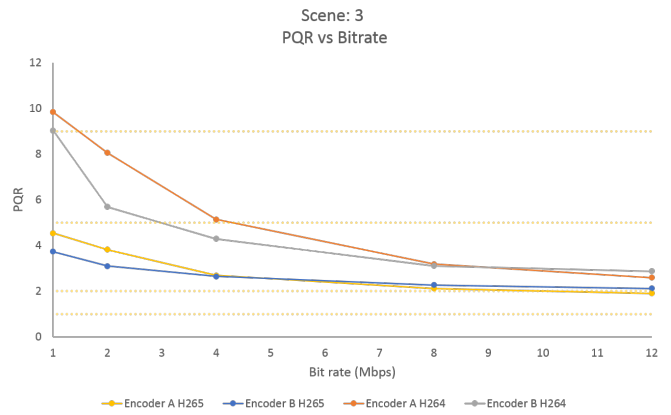
Fig. 6. Scene 2 Results

Fig. 7. Scene 3 Results
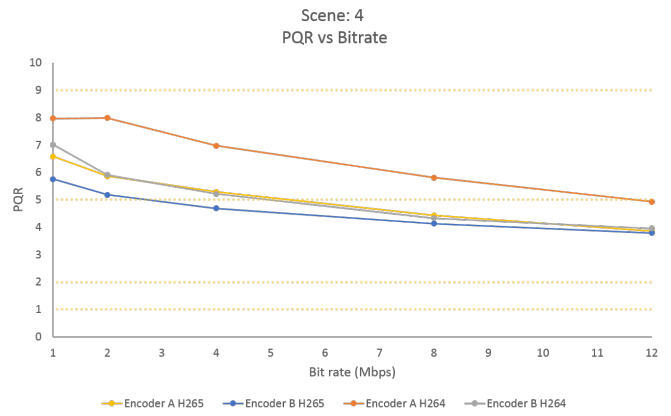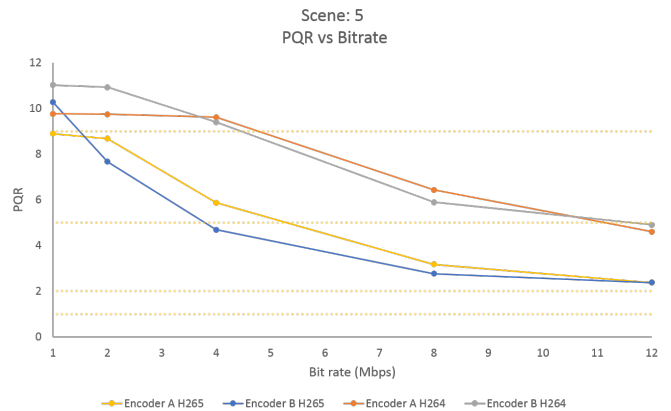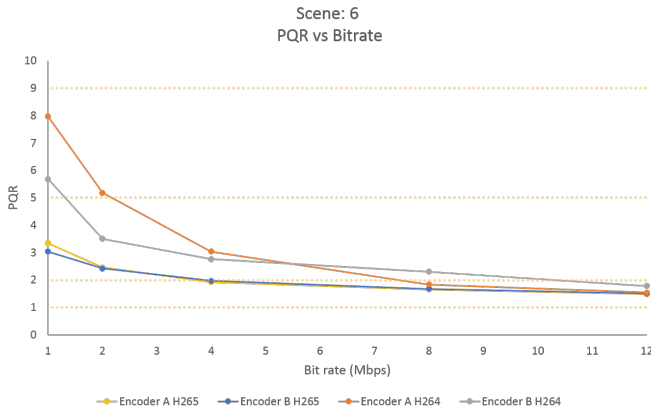
Fig. 8. Scene 4 Results

Fig. 9. Scene 5 Results

Fig. 10. Scene 6 Results

Mbps	Scene	Encoder B	Encoder A	Difference
12	Video 1	1.147955	0.868987	0.278968
	Video 2	1.712122	0.777554	0.934568
	Video 3	2.113891	1.896575	0.217316
	Video 4	3.79051	3.851428	-0.060918
	Video 5	2.379502	2.370172	0.00933
8	Video 1	1.208552	0.965849	0.242703
	Video 2	1.275069	1.023524	0.251545
	Video 3	2.268451	2.120817	0.147634
	Video 4	4.128992	4.429365	-0.300373
	Video 5	2.766014	3.167514	-0.4015
4	Video 1	1.308326	1.398816	-0.09049
	Video 2	2.621488	1.546753	1.074735
	Video 3	2.6482	2.686852	-0.038652
	Video 4	4.688632	5.283137	-0.594505
	Video 5	4.688603	5.875813	-1.18721
2	Video 1	2.616988	2.272637	0.344351
	Video 2	3.628826	2.528985	1.099841
	Video 3	3.091973	3.810917	-0.718944
	Video 4	5.179402	5.858548	-0.679146
	Video 5	7.670425	8.676266	-1.005841
1	Video 1	4.769659	6.73339	-1.963731
	Video 2	4.465652	5.326486	-0.860834
	Video 3	3.729418	4.53466	-0.805242
	Video 4	5.752618	6.580311	-0.827693
	Video 5	e10.272194	8.892453	1.379741

Fig. 11. All PQR resutls for Encoder A and Encoder B using h.265 compression (scenes 1 - 5)

In every scene, the h.265 encoders outperform the h.264 encoders. Depending on the scene, this difference can be drastic or subtle. Between Encoder A and B, it can be observed that the difference is minimal. Some scenes favor Encoder A, while other scenes favor Encoder B. This is also true for bitrates. In almost every scene, some bitrates favor Encoder A, while other bitrates favor Encoder B.

Figure 11 shows the differences taken between each encoder. Using these differences, the results in Figure 12 were generated. Because there are less than 30 samples, the t distribution was used to arrive at these results.

The confidence intervals were also calculated for two questions: 1) At what confidence interval is h.265 better than h.264 for Encoder A. 2) At what confidence interval is Encoder B better than Encoder A. By gathering all PQR averages for every scene, and weighting all scenes and bitrates equally, we can say with 99.9999% confidence that for Encoder A, h.265 is better than h.264. For the same weighting as our first questions, we can say that Encoder B is better than Encoder A with only 62.2% confidence. These confidence intervals were calculated using the methods discussed in Chapter 13 of [5]. Specifically section 13.4.1, which compares paired observations. As an example, we will look at the calculation for comparing Encoder A and Encoder B (only using five scenes due to time constraints and only using their h.265 encoders).

Standard Deviation	0.78984404
Alpha	0.378
Confidence Level	**62.2**
Sample Size	25
Mean	-0.14217388
Confidence Interval	0.141888175
Confidence Low	-0.284062055
Confidence High	-0.000285705

Fig. 12. Confidence level calculated from results in Figure 11

This would not be enough to make a decision between Encoder A and Encoder B without knowing how to properly weight the scenes and bitrates. It would be too early in this study to make a proper selection.

VI. Conclusion

It is conclusive that h.265 outperforms h.264 in all situations. However, if a selection between Encoder A or Encoder B was necessary, factors other than video quality would need to be assessed. Other factors that would come into play are cost, weight, ruggedness, speed of encoding, ease of use, and other metrics.

Acknowledgment

References

[1] "HEVC / H.265 Explained."
[2] "JPEG 2000: How does it work?."
[3] "Understanding PQR, DMOS, and PSNR Measurements," tech. rep.
[4] "Picture Quality Analysis System | Tektronix."
[5] R. Jain, *The Art of Computer Systems Performance Analysis*. John Wiley & Sons, Inc., 1991.