

COMPRESSED VIDEO QUALITY

Iain Richardson, Vcodex Ltd., UK

1. Introduction

The concept of video compression goes hand in hand with the switch from analogue to digital video technology that has taken place over the last 25 years. Video delivered to televisions, computers, and smartphones typically arrives in a compressed form. The bandwidth and file size savings that compression provides are a significant benefit for consumer and business applications, making it possible to send and receive high-definition video over limited capacity networks. However, for digital archive applications, compression can be problematic, especially when it introduces loss or distortion into a video signal.

‘Born digital’ often means ‘born compressed’ and it is increasingly likely that newly-created digital video material will have gone through at least some level of lossy compression. For this reason, it is important to understand the effect of video compression on visual quality. In this paper, I will introduce the concept of video compression and its relationship to video image quality. I will consider the factors that influence visual quality, including technical factors such as codecs and coding parameters, as well as the complex and only partly-understood factors that govern our perception of moving images. I will introduce methods of measuring and quantifying video quality and show how it is possible to compare the quality and performance of video processing systems, despite the limitations of quality measurement.

2. Compressed video

Video is compressed using a video encoder and decompressed using a video decoder. The encoder/decoder pair is known as a ‘codec’.

Figure 1 shows a video encoding and decoding scenario. A video source is encoded to produce a compressed bitstream or file that may be transmitted or stored. The compressed file is decoded to produce a video output that can be displayed. The compressed file typically occupies less space than the original video source, making it easier to store or transmit. The compressed file contains elements to aid in storage, transmission, and playback, such as resynchronisation markers, which help a decoder handle transmission or storage errors.

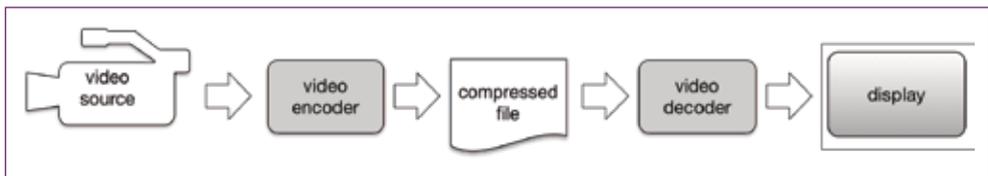


Figure 1. Video encoding and decoding.

There are many image and video codec formats. These include ISO/IEC and ITU-T international standards such as MPEG-2, H.264/AVC, and H.265/HEVC; open source formats such as VP8 and VP9; and proprietary codec formats (Richardson, 2017). A video coding standard specifies a compressed bitstream format and a method of decoding a compressed bitstream.

Video codecs may be lossless or lossy. A lossless codec creates a compressed bitstream that can be decoded to re-create an identical copy of the original video sequence. Lossless video codecs typically provide modest levels of compression, such that the compressed bitstream is around 2x–4x smaller than the original video clip. A lossy codec discards information dur-

ing compression so that the decoded video clip is not identical to the original clip. Typically, a lossy video codec can achieve much higher compression than a lossless codec, 50x or more, at the expense of degradation in video quality. I will discuss the nature of lossy compression in Section 4.

2.1 Compressed video example

A source video clip *Tractor* consists of approximately 28 seconds of video. Each video frame is 1920x1080 pixels (1080p format) and there are 25 frames per second. The original video file occupies 2.15 GBs of disk space.

The source file is encoded using the HEVC/H.265 video codec, using four different compression settings. The following table shows the file size for each setting. Lossless compression yields a compression ratio of 2.4:1, i.e., the compressed size is 2.4x smaller than the original. Three levels of lossy compression, low, medium and high, produce compression ratios of between 63:1 and an impressive 1343:1.

File	File size	Compression ratio
Original, <i>Tractor</i>	2.15 GBs	1:1
Lossless compression (HEVC)	854 MBs	2.5:1
Low compression (HEVC)	24 MBs	63:1
Medium compression (HEVC)	6.5 MBs	330:1
High compression (HEVC)	1.6 MBs	1343:1

Figure 2 illustrates the effect of each compression setting on a single frame of the video clip. At the top left is a close-up of a frame from the original clip. The slight blurring on the wheels and background is due to the movement of the tractor relative to the camera. At the top right is a close-up of the same frame from the low-compression version. It is difficult to see any difference, even though this uses lossy compression. When the full sequence is played back at 25 frames per second, it is even more difficult to see any difference from the original clip. This low-compression version could be described as *visually lossless*, i.e. visually indistinguishable from the original, even though the decoded video data is not identical to the source video.

The medium compression version, lower-left, has some visible degradation, for example around the lettering on the tractor's body. The high compression version, lower-right, is obviously distorted, though it is still recognizably a tractor!



Figure 2. Compression and video quality.

3. Compressed Video Quality and Digital Preservation

The effects of lossy compression can range from very obvious to nearly imperceptible. Why is it important to consider these effects in a digital preservation context?

Since the 1990s, more and more moving image content has been born digital, i.e., created and stored in a digital form. Much of this content is also born compressed. For example, video material recorded on a present-day device such as a camcorder or SLR is increasingly likely to be captured in a compressed form, as Figure 3 illustrates. The camera device has a video encoder built in and video is encoded before recording onto an SD card or other storage. Consumer cameras and camcorders capture video in lossy formats such as H.264/AVC. At the time of writing (2017), only high-end motion picture cameras are capable of recording lossless RAW video. Entry-level professional video cameras, high-end consumer or prosumer cameras, camcorders, SLR cameras, and smartphones capture only lossy video in formats such as MPEG-2 Video, H.264/AVC, and H.265/HEVC.

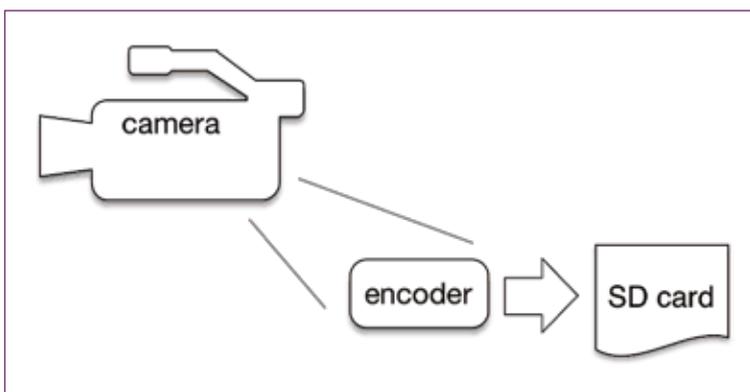


Figure 3. Camcorder capturing in compressed format.

This means that newly-created video material received by an archive is increasingly likely to originate in a compressed, lossy format. Any processing step can introduce further loss, including but not limited to:

- Editing
- Post-production
- Resizing / resampling
- Colour space conversion
- Transcoding for distribution
- Transcoding for streaming

For example, each of the stages shown in Figure 4, which may be entirely outside the control of an archive, may introduce quality loss in the video signal. From a digital preservation perspective, it is important to understand the mechanisms that affect video quality and how to measure quality. In the next section, I will look at issues that specifically affect a user's perception of compressed video quality.

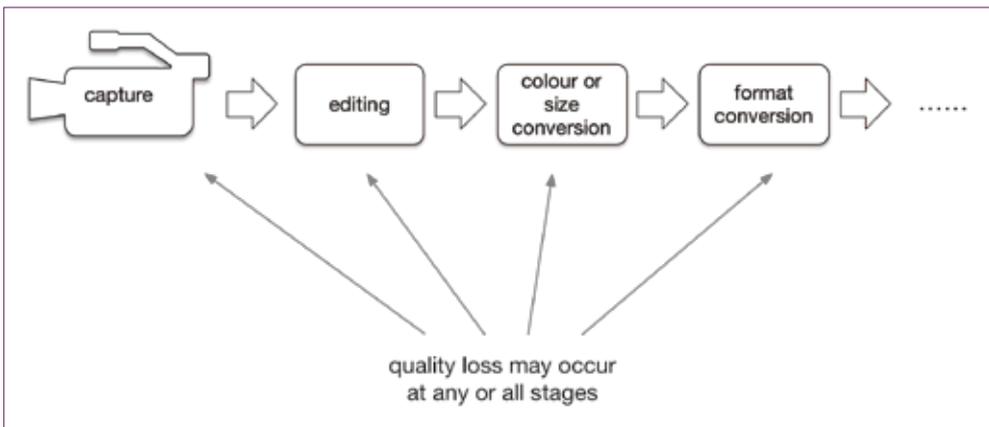


Figure 4. Video processing and quality loss.

4. What Affects Video Quality?

Many different factors can affect the quality perceived by a viewer of a video clip. I will group these factors into system factors, content factors, and human factors.

4.1 System factors

System factors are technical elements that influence video quality. These include encoder settings, codec format, codec implementation, and transmission issues.

In a video codec such as H.264 or HEVC, a number of encoder settings affect the amount of compression. In particular, the quantization parameter (QP) controls the amount of compression and the amount of quality loss. Quantization discards less significant information in a video scene. In general, a higher QP gives increased compression but also increased distortion in the video clip.

Each video coding standard uses a different set of algorithms to compress and decompress video. These differences have an impact on video quality, especially as compression increases.

For example, differences in transform block sizes, prediction types, and deblocking filter designs influence the way in which video quality degrades as compression increases. Two video encoders may be compatible with the same video coding standard but may be implemented using very different design choices. This in turn can significantly affect the visual quality at different compression ratios.

Errors during storage or transmission can introduce significant distortion when the video clip is decoded. Bit errors and glitches in a compressed video stream are often expanded or magnified when the clip is decoding, so that a single error can affect many frames of video. Transmission delays can lead to freezing / buffering which has a detrimental effect on perceived video quality.

4.2 Content factors

Content factors are inherent properties of the video clip that affect the way a human observer may notice, or may not notice, video quality losses. For example, we can see only a certain level of detail in an image. Details smaller than this threshold are not visible. We can see a greater level of detail when the image is static or moving slowly. When objects in a video scene are moving faster, relative to the position of our gaze, our sensitivity to detail goes down.

Our perception of visual scenes is influenced by masking. With spatial masking, we are less likely to notice small changes in a detailed or busy area of a scene. For example, an error or distortion effect will be very obvious in the top section of Figure 5 and least obvious in the semi-random texture of the beach in the bottom section. Temporal masking means that we don't tend to notice details such as errors when there is a sudden, significant change such as a cut between scenes.



Figure 5. Errors and distortions will be most obvious in the top section of this scene.

What this all means is that we are more likely to notice errors and distortions in less detailed, slower-moving parts of a video scene and less likely to notice errors in highly detailed or rapidly changing parts of a scene. This in turn means that the type of video content may affect our perception of video quality. For example, we may rate the quality of a fast-moving or detailed scene *higher* than a slow-moving or less-detailed scene, even though the same amount of image degradation is present.

4.3 Human factors

Human factors are aspects of our environment, experience, and attention that affect how we perceive video quality (Callet and Niebur, 2013). When we are looking at a video image, the environment influences our perception. For example, light levels in the room, viewing distance, the presence or absence of backlighting, and even the level of physical comfort can influence an observer's opinion of the image.

While a very large amount of visual information reaches the retina (Figure 6), the amount of information we can attend to at any one time is limited to less than 10kbits per second, according to Anderson, Van Essen and Olshausen (2005). Because of this, we fill in much of the visual scene based on our experience and expectation of how the visual world works. Hence, the human observer can only really attend to a small subset of a displayed video image. Furthermore, the amount of data that is retained in medium- and long-term memory is even smaller. This means, for example, that our perception of video quality is dominated by the most recent 5–10 seconds of video (Moss, et al., 2016).



Figure 6. Visual attention and information content, adapted from Anderson, Van Essen and Olshausen (2005).

Our opinion of video quality is influenced by factors such as the presence or absence of human figures, perhaps because we learn to pay particular attention to other humans, and by whether or not we have a task to carry out. For example, we perceive visual quality differently if we are carrying out a specific task when watching a video clip, compared with passively watching a video (Zhong, et al., 2004).

A human observer's attention tends to be attracted to unusual or unexpected items in the visual field. There is some evidence that the distortions due to high compression can disproportionately affect our perception, because they stand out and attract attention (Redi, et al., 2011).

Our understanding of the way a human observer acquires and processes visual information is still only partially complete. The complexities of the active process of human vision (Findlay and Gilchrist, 2003) make it exceptionally difficult to measure a subjective concept such as video quality.

5. Measuring and Comparing Compressed Video Quality

The range of factors discussed above make it difficult to measure video quality accurately and consistently. Our perception of a visual scene is fundamentally subjective and no two individuals will have exactly the same perception of a video clip. At most, we can make an approximate evaluation of quality, either through experimental assessment by human observers— **subjective quality assessment**—or through the use of measurements and computational methods— **objective quality measurement**.

5.1 Subjective assessment

Subjective quality assessment involves determining video quality based on the opinions of human observers. It is typically carried out by asking multiple observers each to rate the quality of multiple video clips and then analysing the results.

Many different subjective quality assessment methods are described in the literature (Pinson and Wolf, 2003) and in several industry standards such as ITU-T BT.500 (ITU-T, 2012). For example, the Double Stimulus Continuous Quality Scale (DSCQS) method presents the viewer with (A) an unimpaired reference clip and (B) the same clip after it has passed through a system such as a video codec (Figure 7). The viewer is asked to give each clip a score by marking on a line with five intervals ranging from Excellent to Bad. Each viewer is asked to rate a number of pairs of original and modified clips. The order of clips is randomized so that the viewer does not know which is the original and which is the modified version. The results from multiple viewers are combined to give a Mean Opinion Score (MOS) which indicates the subjective visual quality of each sequence under test (Streijl, Winkler and Hands, 2016).

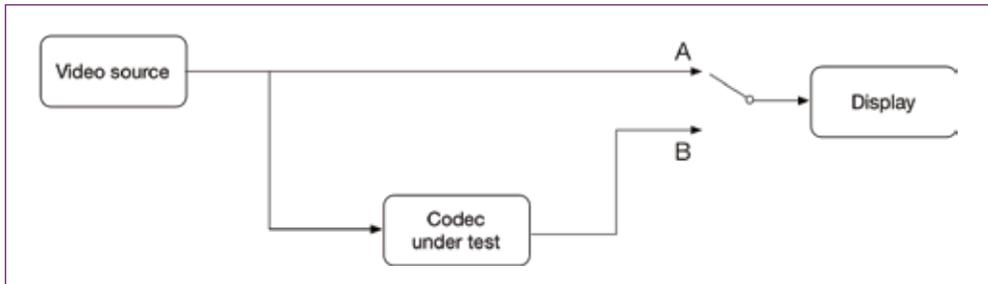


Figure 7. Double Stimulus quality testing.

The results of a subjective quality test can vary depending on the viewer and on the sequence under test. It is therefore recommended to repeat each test with multiple clips and multiple assessors. Carrying out subjective quality tests with many assessors is time consuming and expensive. Because of this, objective metrics, i.e., quality measurements that can be calculated automatically, are often used instead of, or as well as, subjective tests.

5.2 Objective measurement

Numerical measures of image distortion include Mean Squared Error (MSE) and Peak Signal to Noise Ratio (PSNR). PSNR is widely used for estimating distortion in compressed images and video. It is computationally simple to calculate and produces a single number. A higher PSNR should equate to better image quality. However, simple numerical measures such as PSNR do not fully reflect the way human observers notice, or do not notice, quality degradation. For example, Figure 8 shows two close-ups from video images. Both *Tractor* (left) and *Snow Scene* (right) have approximately the same PSNR, i.e., the same numerical level of distortion compared with the original video clips. However, it is much easier to notice the quality loss in *Tractor* because the large, homogeneous and recognisable image features are clearly distorted. On the other hand, the distortion in *Snow Scene* is masked by the detailed, semi-random textures of the tree branches.



Figure 8. Tractor (left) and Snow Scene (right), approximately the same numerical level of distortion (Peak Signal to Noise Ratio).

Many attempts have been made to develop image and video quality metrics that more accurately reflect the results of subjective quality assessment. Several objective quality metrics are presented in ITU-T Recommendation J.247 (ITU-T, 2008). The Structural Similarity Index (SSIM) is reported to perform well compared with other objective metrics (Wang, et al., 2004). Each of these metrics attempts to model the visual effects introduced by distortions such as blurring and blockiness, and to provide a numerical estimate of subjective video quality. Objective metrics have varying success in predicting subjective quality.

5.3 Comparing performance

Despite the challenges and limitations of subjective and objective quality measurement, it is at least possible to **compare** the performance of different systems. For example, if we consider a single video clip coded in multiple ways and at a range of compressed bitrates, then most of the well-known subjective or objective quality measures will give an indication of relative performance.

If we measure the quality of the clip at a series of compressed bitrates, we can obtain a **rate-distortion curve** such as the curves shown in Figure 9. This example shows four rate-distortion curves. The upper curves are generated by coding Video Clip A using two different lossy video codecs, Codec A and Codec B. Each point on the curve records the average bitrate and quality for a single coding pass, i.e., for a complete encoding and decoding process using a single set of parameters. By repeating the coding process and changing the coding parameters, we can obtain a complete curve. For example, increasing QP will produce a video clip with a lower bitrate and a lower quality. Repeating this with multiple QPs produces a curve such as the examples in Figure 9. If we repeat this process with two different video codecs (Codec A and Codec B), we can compare the performance of these codecs across a range of compression bitrates.

For Video Clip A, Codec B performs consistently better than Codec A, i.e. the quality of the decoded video clip is better at every bitrate. If we repeat this for a second video clip, Video B, we can expect to get a different set of curves. In the example of Figure 9, the two codecs perform very similarly to each other when compressing Video B. Codec A is slightly better at low bitrates, i.e. the decoded quality is slightly higher, but Codec B is better at high bitrates.

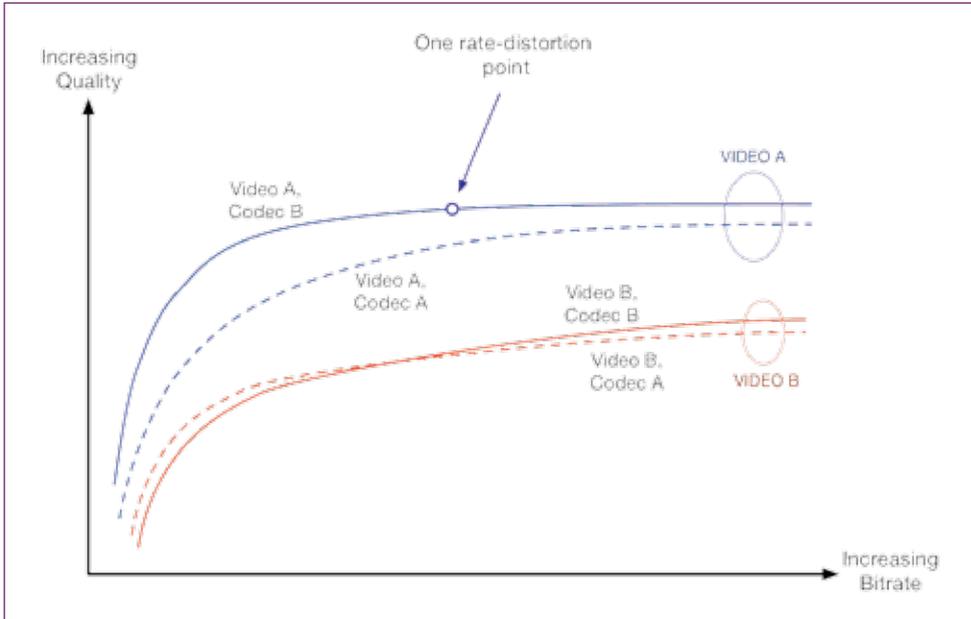


Figure 9. Rate-distortion curves : two clips and two codecs.

Let us consider a real example. Figure 10 shows subjective test results for two HD video clips, *Campfire Party* and *BQ Terrace*, compressed using four different lossy video codecs (source: Baroncini, et al., 2017). Each clip is coded at a range of compressed bitrates. HM is the HEVC/H.265 reference video codec and JEM, PA, and PB are three experimental video codecs under test. The X-axis indicates the compressed bitrate and the Y-axis indicates the Mean Opinion Score (MOS). Each point on the graph represents the average subjective quality score of multiple observers.

For *Campfire Party*, we can see that the experimental codecs out-perform the HEVC codec at all the tested bitrates. At every bitrate (x-axis), the experimental codecs give a higher MOS (y-axis). For the second clip, *BQ Terrace*, the experimental codecs out-perform the HEVC codec at lower bitrates. At the highest bitrate (around 1700 kbits/second), there is little difference between the four codecs, probably because none of the decoded clips shows obvious distortion. From these charts, we can see that the new codecs perform better than the older HM codec, at least for the video clips and bitrates tested.

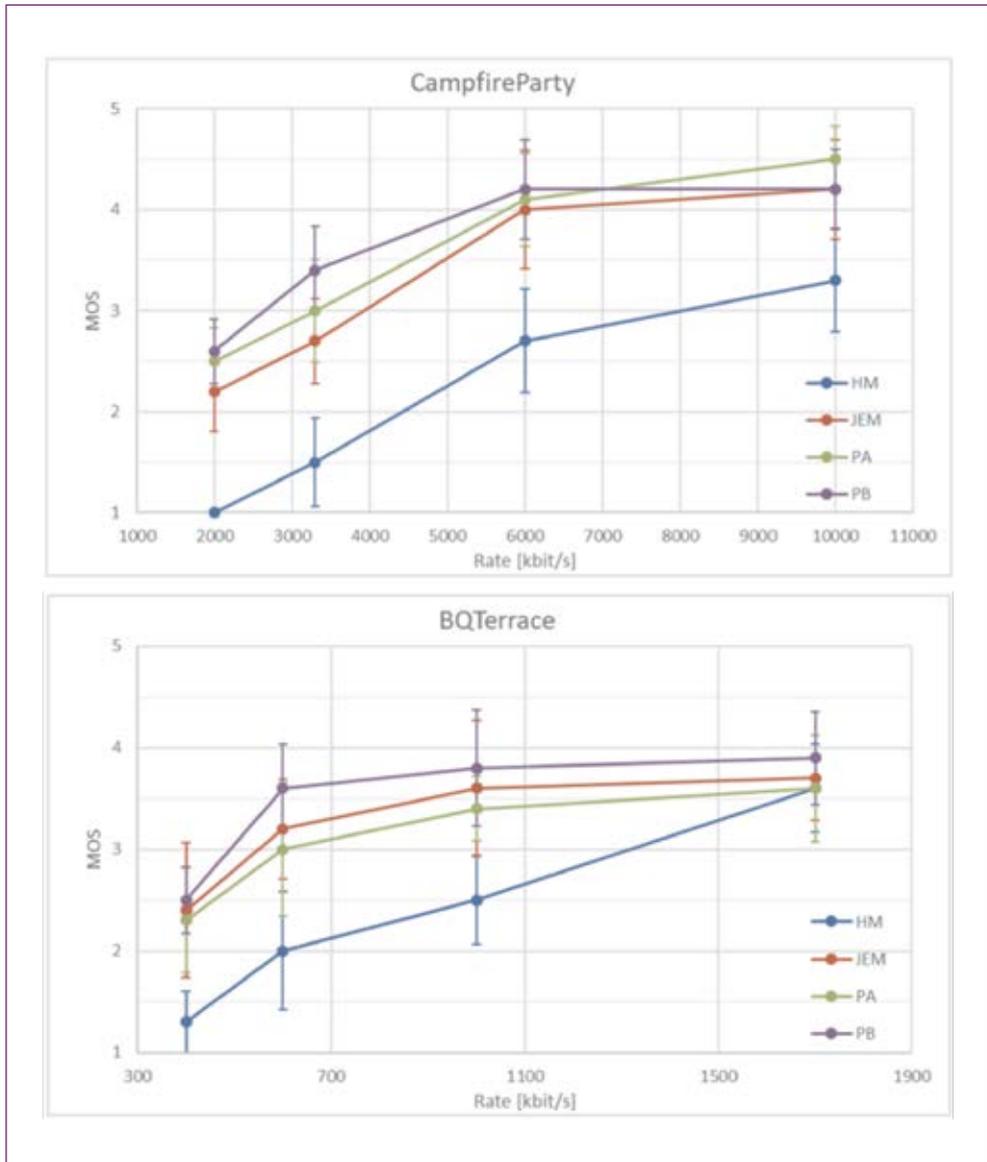


Figure 10. Subjective evaluation of two video clips, each coded using four codecs at a range of bitrates, from Baroncini, et al., 2017.

6. Conclusions

The increasing trend towards born-compressed video, i.e., video that is encoded at source using lossy compression, makes it necessary to understand lossy video compression and its associated quality issues.

Compressed video quality is a complex topic with some unresolved questions. For example, a definitive and universally agreed measure of compressed video quality is still some way off. The range of subjective and objective measurements and testing procedures can be confusing, especially when trying to compare different systems tested using different methods.

When attempting to evaluate the quality and performance of video processing systems or stages, the following practical considerations may be helpful:

1. Test (or ask for test results) using a range of video content and at a range of compressed bitrates.
2. Test using content that is representative of your own application area. For example, content with graphics or other computer-generated material may perform very differently from natural video scenes.
3. Where practical, consider testing using multiple methods, for example two or more of the following:
 - a. Subjective testing with a single observer.
 - b. Subjective testing with multiple observers, following a protocol such as DSCQS.
 - c. Objective testing, if possible using a more sophisticated alternative to PSNR.
4. When comparing systems, where possible, compare results using the same video sources, the same range of compressed bitrates and the same quality metrics.

Despite the challenges and uncertainties associated with video quality measurement, lossy video compression is likely to be with us for some time to come.

7. References

- Anderson C., Van Essen D. and Olshausen B., 2005. "Directed visual attention and the dynamic control of information flow." *Encyclopedia of Visual Attention*. Academic Press/Elsevier. Available at: <https://redwood.berkeley.edu/bruno/papers/attention-chapter.pdf> [Accessed 14 January 2018].
- Baroncini, V., Hanhart, P., Wien, M., Boyce, J., Segall, A., and Suzuki T., 2017. "Results of the Joint Call for Evidence on Video Compression with Capability beyond HEVC". *ITU-T and ISO/IEC Joint Video Exploration Team Document JVET-G1004*. Available at: http://phenix.it-sudparis.eu/jvet/doc_end_user/documents/7_Torino/wg11/JVET-G0004-v2.zip [Accessed 14 January 2018].
- Callet, P. L. and Niebur E., 2013. "Visual attention and applications in multimedia technologies." *Proc IEEE*, Sept 2013. Available at: <http://ieeexplore.ieee.org/abstract/document/6547645/> [Accessed 14 January 2018].
- Findlay, J. and Gilchrist, I., 2003. *Active Vision: The psychology of looking and seeing*. Oxford University Press.

- ITU-T, 2008. *Recommendation J.247, Objective perceptual multimedia video quality measurement in the presence of a full reference*. Available at: <https://www.itu.int/rec/T-REC-J.247/en> [Accessed 14 January 2018].
- ITU-T, 2012. *Recommendation BT.500, Methodology for the subjective assessment of the quality of television pictures*. Available at: <https://www.itu.int/rec/R-REC-BT.500> [Accessed 14 January 2018].
- Moss F., Wang K., Zhang F., Baddeley R. and Bull, D., 2016. "On the optimal presentation duration for subjective video quality assessment." *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 26, No. 11. Available at: <http://ieeexplore.ieee.org/abstract/document/7172512/> [Accessed 14 January 2018].
- Pinson, M. and Wolf S., 2003. "Comparing subjective video quality testing methodologies." *Proc. Visual Communication and Image Processing*. Available at: <http://dx.doi.org/10.1117/12.509908> [Accessed 14 January 2018].
- Redi, J., Liu, H., Zunino, R. and Heynderickx, I., 2011. "Interactions of visual attention and quality perception." *Proc. SPIE 7865, Human Vision and Electronic Imaging XVI*. Available at: <http://dx.doi.org/10.1117/12.876712> [Accessed 14 January 2018].
- Richardson, I., 2017. "Video compression codecs: a survival guide." *IASA Journal* no. 47.
- Streijl, R., Winkler S. and Hands, D., 2016. "Mean Opinion Score (MOS) revisited: methods and applications, limitations and alternatives." *Multimedia Systems Journal*, vol. 22, no. 2. Available at: <https://link.springer.com/article/10.1007/s00530-014-0446-1> [Accessed 14 January 2018].
- Wang, Z., Bovik, A., Sheikh, H. and Simoncelli, E., 2004. "Image quality assessment: from error visibility to structural similarity." *IEEE Transactions on Image Processing*. Available at: <http://ieeexplore.ieee.org/abstract/document/1284395/> [Accessed 14 January 2018].
- Zhong, Y., Richardson, I., Sahraie A. and McGeorge, P., 2004. "Influence of task and scene content on subjective video quality." *Image Analysis and Recognition*. Available at: https://doi.org/10.1007/978-3-540-30125-7_37 [Accessed 14 January 2018].