Technical Disclosure Commons

Defensive Publications Series

March 2020

Using a rate-distortion curve to determine an ideal maximum video bit rate

Michael J. Horowitz

Erik Språng

Using a rate-distortion curve to determine an ideal maximum video bit rate

ABSTRACT

The target bit rate for a video that is to be transported across a network to a receiver is typically determined by the maximum throughput of the network. Such an approach is incognizant of the statistical properties of the video source and can result in inefficient use of bandwidth and/or compute resources. This disclosure describes techniques, based on the rate-distortion curve of the video source and encoder, to determine an ideal maximum bit rate for a video encoder, e.g., a bit rate that produces excellent video quality without undue load on network or compute resources. The techniques automatically adapt to and account for statistical parameters of the video source such as spatial resolution, frame rate, source complexity, etc. The techniques are applicable to a wide variety of video encoder types.

KEYWORDS

- Video encoding
- Video transmission
- Bit rate
- Rate-distortion curve
- Video source complexity
- Video source resolution
- Video frame rate
- Distortion measure
- Quantization step size

BACKGROUND

In video communications applications such as video conferencing, a video encoder compresses source video to a specified target bit rate in order to transport it over a network to a receiver. Often, the target bit rate is determined by the maximum throughput or bandwidth of the network; however, as is often the case in enterprise video conferencing, the available bandwidth for coded video is higher than that necessary to provide a high-quality video experience. In such cases, video conferencing applications typically use a pre-specified maximum to limit bit rate. However, even in cases where the available bandwidth is high, it is advantageous to use bandwidth efficiently.

Generally, increasing the target bit rate of an encoder results in higher quality video. At low bit rates, relatively small increases in target bit rate lead to relatively large increases in quality. As the target bit rate increases, it takes larger and larger rate increases to achieve the same increase in quality. Eventually, at sufficiently high bit rates, the video quality is excellent and increasing the rate to achieve a small, perhaps imperceptible, improvement in quality is an inefficient use of bandwidth as well as CPU resources (due to the fact that encoding video at higher bit rates is more CPU intensive).

Ideally, a maximum bit rate value can allow an encoder to produce excellent video quality at a rate that is not so high as to be bandwidth inefficient. Such a maximum bit rate, which produces excellent video quality without loading the network, is herein referred to as the *ideal maximum bit rate*. Several factors affect the value of the ideal maximum bit rate, e.g., the encoder type and implementation, the spatial resolution, the frame rate, the complexity of the source video, etc. For a given encoder, a more complex video (e.g., a video with lots of motion and high-contrast detail) typically requires a higher bit rate to achieve excellent quality compared with one of lower complexity.

In practice, it is common to set a maximum bit rate that is a function of the spatial resolution and frame rate of the source. The value for the maximum bit rate is usually selected by empirical means with the goal of providing the encoder with ample rate to produce excellent quality video over a wide variety of sources. This approach can lead to inefficient bandwidth use due to the fact that the maximum bit rate might be too high for lower complexity video. Also, in extreme cases, the maximum bit rate can set at a value that is not high enough to maintain quality. As such, not accounting for video complexity when determining a maximum bit rate leads to inefficient use of bandwidth and compute resources.

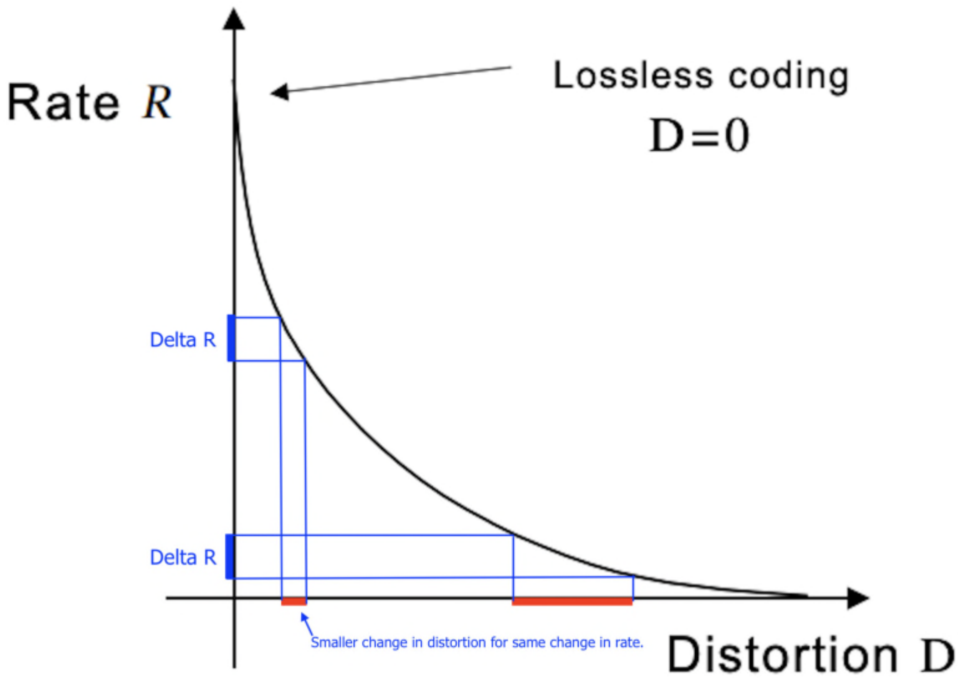

Fig. 1: An example rate-distortion curve

The aforementioned relationship between video quality, e.g., distortion, and bit rate is captured by a rate-distortion (RD) curve, illustrated in Fig. 1. The higher the target bit rate, the higher the quality or equivalently, the lower the distortion, measured, e.g., using mean-square error, peak signal-to-noise ratio, structural similarity index (SSIM), or other measures of distortion. The rate-distortion curve for a particular encoder and video source shows the distortion level that the encoder will introduce to the coded video signal at a specified target bit rate. Rate-distortion curves typically have a non-linear, convex shape. As illustrated in Fig. 1, the shape of the RD curve is such that a small increase in bit rate when operating at a relatively low bit rate results in a large decrease in distortion, whereas the same, small increase in rate when operating at a relatively high bit rate results in a much smaller decrease in distortion.

Video encoders such as H.264, H.265, VP8, VP9, and AV1 work by modeling (predicting) source video using information from previously encoded frames or previously encoded information in the same frame. These encoders then transform and quantize the difference between the source and prediction. The quantization process effectively controls the point on the rate-distortion (RD) curve where the encoder operates.

Roughly speaking, the encoder controls the quantization step size. The larger the quantization step size, the more error, e.g., distortion, introduced by the encoding process, and the lower the resulting bit rate of the coded video. In VP8, VP9, and the other video encoding formats listed above, a quantization parameter (QP) is used directly or indirectly to control the quantization step size. The quantization step size varies directly with QP value, e.g., the larger the QP value, the larger the corresponding quantization step size.

In video encoders starting with H.264, the relationship between QP (and QIndex) and quantization step size follows a power law. For example, $qss = a \times 2^{(QP-b)}$ where qss is the quantization step-size, a is a positive constant, and b is a constant whole number. In practice, it is typical for a video encoder to be configured to achieve a target bit rate. The encoder's rate control procedure then adjusts the QP value (and consequently the quantization step size) in order to achieve the target rate. The quantization parameter is generally readily available to external applications.

DESCRIPTION

This disclosure describes techniques to determine an ideal maximum bit rate for a video encoder, e.g., a bit rate that produces excellent video quality without undue load on network or compute resources. The techniques automatically adapt to and account for such video-source parameters as spatial resolution, frame rate, source complexity, etc. The techniques are applicable to a wide variety of video encoder types.

The techniques leverage the aforementioned correlation between the slope of the RD curve and the quality of the encoded video. As seen from Fig. 1, when the absolute value of the slope is small (i.e., the slope is more horizontally oriented), video quality and the encoded bit rate, tend to be low: the RD curve has lower bit rates and higher distortion where its slope is small. Conversely, video quality tends to be higher when the absolute value of the slope is large (i.e., the slope is more vertically oriented). Rather than explicitly computing distortion (e.g., mean square error, SSIM), which can be computationally burdensome, the techniques compute distortion as the square of the measured quantization step size. The use of the square of the quantization step size as a distortion measure is consistent with the fact that square-error distortion for an optimally designed uniform scalar quantizer (e.g., the type of quantizer used in VP8, VP9, H.264, H.265, and AV1) is proportional to the square of the quantization step size.

Using the square of the quantization step size as the distortion measure, an estimate of the slope of the RD curve between two bit rates, $R(i)$ and $R(i-1)$, is obtained as follows:

RD_slope = $(R(i-1) - R(i)) / (qss(i-1)^2 - qss(i)^2)$, where $qss(x)$ is the quantization step size that the encoder uses to encode the source at bit rate, $R(x)$. In practice, RD_slope can assume small, negative values that can present numerical difficulties, especially at low bit rates. Therefore, for convenience, S, a variant of RD_slope, defined as $S = -\frac{1}{RD\_slope} = -\frac{qss(i-1)^2 - qss(i)^2}{R(i-1) - R(i)}$, is used to estimate the ideal maximum bit rate. The computation of S is of relatively low complexity: One divide and two additions.

Example: *Target bit-rate versus RD_slope (and S) for a VP9 encoder of a 720p, 30 fps video source*

Target Bit Rate (kbps)	QIndex	Quantization Step Size	Quantization Step Size Squared	RD_slope	Value of S
4000	43	50	2500		
3500	45	52	2704	-2.45	0.4
3000	47	54	2916	-2.36	0.4
2500	52	59	3481	-0.88	1.1
2000	60	67	4489	-0.50	2.0
1500	67	74	5476	-0.51	2.0
1000	78	85	7225	-0.29	3.5
500	97	106	11236	-0.12	8.0
250	131	185	34225	-0.01	92.0

Table 1: Bit rate as a function of *RD_slope* and *S*

A VP9 encoder encodes a 720p, 30 fps video captured using a video loopback application at a set of specified target bit rates. The QIndex (derived from the QP value) is measured and the corresponding quantization step size found using the ac_qlookup look-up table in the open-source libvpx. The measured QIndex values are averaged using exponential moving average as follows: QIndex_avg = (250 × QIndex_avg + 5 × QIndex) >> 8, where the value of QIndex is for the current frame, and the symbol >> represents a right bit shift. Table 1 indicates the target bit rate as a function of RD_slope (and *S*). At low bit rates, both the small, negative values of RD_slope and the more numerically stable values of *S* are observed.

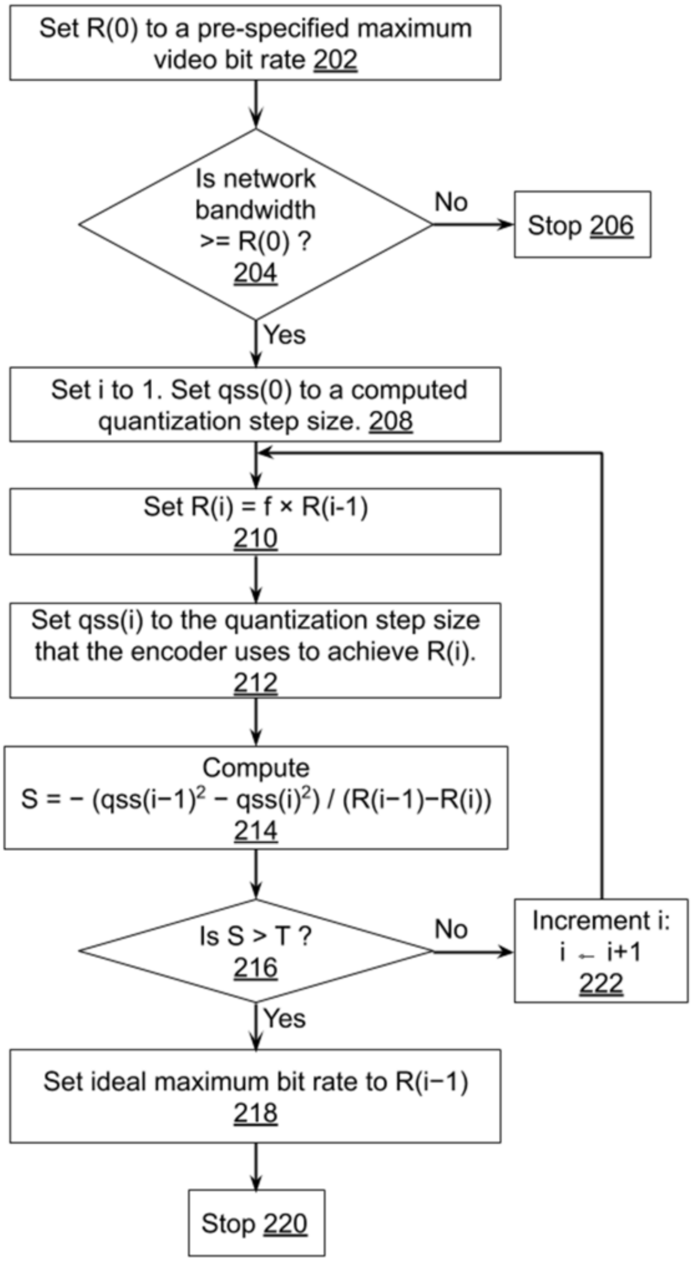

Fig. 2: Using a rate-distortion curve to determine an ideal maximum video bit rate

Fig. 2 illustrates using a rate-distortion curve to determine an ideal maximum video bit rate, per techniques of this disclosure. A video encoder with a pre-specified maximum bit rate begins encoding source video. The initial bit rate *R*(0) is set to the pre-specified maximum bit rate (202). The initial bit rate is high enough to result in excellent video quality for nearly any source, but is also likely to unduly load network and compute resources. If the available network bandwidth is less than *R*(0) (204), then the procedure to find the ideal maximum bit rate terminates (206). If the available network bandwidth is greater than *R*(0), then *qss*(0) is set to the quantization step size that the encoder uses to achieve bit rate R(0), and the index *i* is set to 1 (208). As in the above VP9 example, the values used to compute quantization step size may vary, and it can be advantageous to average such values.

At the present iteration, the bit rate *R*(*i*) is set to a fraction *f* of the previous bit rate (210): $R(i) = f \times R(i-1)$, where a reasonable value for *f* might be in the range 0.5 to 0.95. Encoding continues. The quantization step size for the current iteration *qss*(*i*) is set equal to the quantization step size that the encoder uses to achieve *R*(*i*) (212). *S* is computed as (214)

$$S = -\frac{qss(i-1)^2 - qss(i)^2}{R(i-1) - R(i)} \quad .$$

S is compared to a predetermined threshold *T* (216). The value of *T* can be determined empirically by examining video quality at various bit rates for a given encoder over several source videos. If $S > T$ (218), the ideal maximum bit rate is determined as $R(i-1)$, and the procedure stops (220). In this case, the encoder is operating at a point on the RD curve where further bit rate reduction may result in visible quality degradation. If $S \leq T$, the encoder is likely operating at a point on the RD curve where encoder quality is excessively high, e.g., network and compute resources are unduly being loaded. The index *i* is incremented (222), the procedure repeats from 210.

In the above procedure, the QP value of the encoder can be monitored on a frame-by-frame basis (or at other suitable intervals). If the target bit rate of the encoder is at the ideal maximum bit rate and the QP value rises rapidly resulting in a sudden decrease in video quality (as might happen if the video scene were to suddenly become more complex and remain that way), the maximum bit rate of the encoder can be increased to the pre-specified maximum bit rate or to a value between the current value of the ideal maximum bit rate and the pre-specified maximum bit rate, and the procedure to find the ideal maximum bit rate run again. The exact value to which the maximum bit rate of the encoder is increased depends on how much the QP value has increased. In consort with the increase in the maximum encoder bit rate, the threshold T can also be decreased, e.g., by 10%.

CONCLUSION

This disclosure describes techniques, based on the rate-distortion curve of the video source and encoder, to determine an ideal maximum bit rate for a video encoder, e.g., a bit rate that produces excellent video quality without undue load on network or compute resources. The techniques automatically adapt to and account for statistical parameters of the video source such as spatial resolution, frame rate, source complexity, etc. The techniques are applicable to a wide variety of video encoder types.