The CABS$_{score}$
A definite video codec metric

Alberto Vigata

email: `alberto.vigata@gmail.com`

July 23, 2014

Abstract

Measuring video codec performance has traditionally been a simple matter. One selects an encoder and runs a sequence set through it producing the resulting compressed bitstreams. Subsequent comparisons of the size of the resulting bitstream against some chosen metric (SSIM [1], PSNR, ...) yields the framework for comparison of performance between different setups. While this is a valid proposition some problems are inherent with it. The lack of standardization between benchmarking setups makes their results not comparable and therefore relative. This paper introduces the $CABS_{score}$ (Codec Average Bitrate Savings) which provides a definite, clearly stated framework to the benchmarking environment.

I. INTRODUCTION

A. Motivation

With numerous video codec schemes being released frequently, the need to understand their performance characteristics is more prevalent than ever. Codec performance benchmarking is done directly by those who are releasing the product or by a third party. While results are usually shared publicly, a sense of opaqueness can remain unless all the materials used during the benchmarking process are released to the public. Even when the later is available, no standardized way seems to be available to compare with other setups. $CABS_{score}$ tries to address this issue by providing a definite framework in which anyone can produce benchmarking scores and share them unambiguously with the world.

B. Informal description

$CABS_{score}$ is an intuitive and simple score, from which any reader can get a sense of the underlying performance of the codec at hand. To start with, a reference codec configuration is selected towards which all other measurements will be referred to. From this particular configuration the score calculates the codec average bitrate savings through a determined range of bitrates. The result of the score is the percentage of savings from the reference configuration.

II. $CABS_{score}$ COMPONENTS

The intent of this specification is to establish a framework in which anyone can define their own scores, and being able to share it with others in a seamless manner. In this context, the specification essentially establishes the parameters surrounding the testing setup with enough specificity to remove ambiguity, and with enough flexibility to allow versatility in benchmarking scenarios. In detail, a particular $CABS_{score}$ is defined with all the following required parameters:

1) A Nickname

The nickname of the score gets added to the subscript when referring to it. This acts as a shorthand for it. A nickname is only meaningful in the context of the other defining parameters of the score, like: the sequence set, bitrage range ... Example nicknames could be '4K' for $CABS_{4K}$, 'UHD' for $CABS_{UHD}$

2) A Sequence Set

The specific set of uncompressed sequences that are used as input to the encoders. Such sequences must have the same resolution $\mathcal{S}$ but they can vary in the frame count.

3) A Bitrate Range

The range of bitrates $\{br_{low} \rightarrow br_{high}\}$ where the score is going to be performed and where the RD curves are going to be evaluated

4) A Reference Codec Configuration

The reference RD curve that all the other curves are compared to. This can be defined through a specific configuration for an encoder or providing the encoded bitstreams of the encoded configuration

5) A Frame based metric used for comparison

The specific metric used for measuring the distortion introduced by the codec due to the encoding process. For example SSIM or PSNR. Pseudo-code or equivalent algorithmic description must be provided to instruct how to derive the metric numbers on a frame by frame basis.

A. NOTES

All this parameters must be provided along side the score for it to be valid. However, the general goal is that a specific nickname will be generally known to be equivalent to all the other parameters as short hand. For example $CABS_{4KvsHD1}$ could be used as the nickname for one specific set of parameters available in some public repository, such specific $CABS_{score}$ would measure performance of codecs on 4K video compared to an HD era video codec.

1) Materials distribution

The particular way in which the components of a $CABS_{score}$ are distributed are beyond the scope of this specification. However, here are a few suggestions:

rev 0.98 draft

- The sequence set (see II-2) can be provided as a set of files along with their SHA-1 hashes[2] to verify their signature
- The reference binaries used to produce the reference bitstream (see II-4) can be provided precompiled for a specific platform, along side with exact configuration files that produce the reference bitstreams. Alternatively, just the reference bitstreams can be provided as they can be decoded with the video decoder under study, and resulting reference RD curves can be derived from it given the availability of II-2
- Specific implementations of frame-based metrics like SSIM or PSNR must be provided. It is suggested that alongside reference code that produces the reference metric results, the resulting metric values for the sequence set in II-2 are provided. For example, if metric is SSIM based, the values that this particular SSIM implementation metric produces on II-2 can be provided for validation purposes.

III. CALCULATION

The following steps must be followed to produce a $CABS_{score}$ that is complaint with this spec:

A. RD Curve sampling

Due to the intrinsic non-linear nature of the rate distortion curve for a specific codec configuration, the RD curve must be empirically sampled. This specification requires that the RD curve is sampled with a density of at least 10 points through the bitrate range of coverage. For example, if the bitrate range is $\{500_{kbps} \rightarrow 2500_{kbps}\}$ a encoded bitstream must be obtained per configuration with resulting bitrate every 200_{kbps} for the range.

B. RD Curve interpolation

In order to extrapolate values outside the sampled RD curve in the previous step, linear interpolation between the sampled points must be used. The resulting function $ref(x)$ corresponds to the mathematical extrapolated representation of the RD curve. Figure 1 shows a sampled RD curve for the x264 100-2500kbps for the PSNR metric. While other types of interpolation maybe possible, linear interpolation is chosen for its simplicity.

C. Determine Metric range from the bitrate reference range for a specific sequence

The Average Bitrate Savings is informally defined as the average distance in bitrate between the reference RD curve $ref(x)$ and the testing curve $g(x)$. In Figure 2 the green area represents the $CABS$ area of interest with a range of $\{200_{kbps} \rightarrow 800_{kbps}\}$

The average distance is calculated in the y axis (the metric axis) between two points: a lower M_{low} and a higher M_{high}. Both points are formally derived from the reference curve as follows:

$$M_{low} = ref(br_{low}) \tag{1}$$
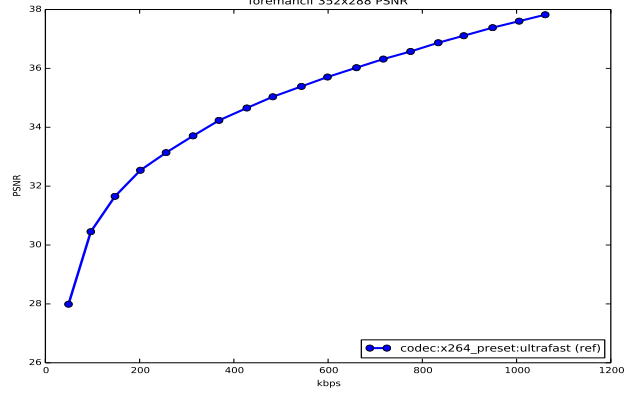$$M_{high} = ref(br_{high}) \tag{2}$$

Figure 1. x264 RD curve for foreman CIF, medium preset, from 100kbps to 2000kpbs, PSNR
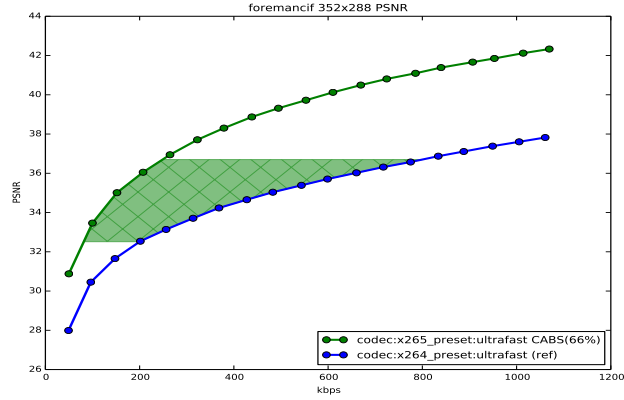

Figure 2. x265 in green $g(x)$ vs x264 in blue $ref(x)$. Area in green is the area of coverage for ABS calculation with range $\{200_{kbps} \rightarrow 800_{kbps}\}$. foreman CIF, ultrafast preset,, PSNR, $CABS$ score averages 66% bitrate savings

D. Calculate bitrate RD Curve differentiation for a specific sequence

In order to find the average distance between the reference RD curve $ref(x)$ and the measuring curve $g(x)$ we measure such distance on a N number of points on the extrapolated curves and then we produce the average. N must be greater or equal than 10000. More formally:

$$step = \frac{M_{high} - M_{low}}{N} \tag{3}$$

$$y_k = step \cdot k + M_{low} \tag{4}$$

$$AverageBitrateSavings_{score_{seqi}} = \frac{1}{N} \sum_{k=0}^{N} \frac{ref^{-1}(y_k) - g^{-1}(y_k)}{ref^{-1}(y_k)} \tag{5}$$

E. Average over sequences

To obtain the final score, we average the ABS_{score} over multiple sequences to obtain the final $CABS_{score}$

$$CABS_{score} = \sum AverageBitrateSavings_{score_{seqi}} \tag{6}$$

rev 0.98 draft

REFERENCES

[1] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh and Eero P. Simoncelli *The SSIM Index for Image Quality Assessment*
https://ece.uwaterloo.ca/ z70wang/research/ssim/

[2] FEDERAL INFORMATION PROCESSING STANDARDS PUBLICATION *Secure Hash Standard*
http://csrc.nist.gov/publications/fips/fips180-4/fips-180-4.pdf