An algorithm for a quality-optimized bit rate ladder generation for video streaming services using a neural network

Andreas Kah[a], Maurice Klein[a], Christoph Burgmair[b], Markus Rasokat[b], Wolfgang Ruppel[a], and Matthias Narroschke[a]

[a]RheinMain University of Applied Sciences, Wiesbaden, Germany
[b]Joyn GmbH, Munich, Germany

ABSTRACT

For video streaming services, a bit rate ladder is generated by encoding each video signal at various bit rates and associated spatial resolutions. For a bit rate ladder that maximizes the subjective quality at a minimum bit rate, it was found that the VMAF of the highest provided quality should not exceed 95, which is on average associated with the same subjective quality as the original signal. Second, all VMAF differences between adjacent renditions should ideally be not greater than 2 as this guarantees indistinguishable subjective quality on average. The generation of a bit rate ladder fulfilling these constraints faces the difficulties that (i) today's encoders cannot be instructed to achieve a certain VMAF and (ii) a certain VMAF can be achieved by various combinations of bit rate and spatial resolution. These difficulties result in a content-dependent multidimensional solution space for generating the quality-based bit rate ladder at a minimum bit rate. In this paper, an algorithm is presented which can generate such a bit rate ladder. The algorithm determines the VMAF of nine initial encodings of the signal. Using a specifically designed and trained neural network, the VMAF of 5805 combinations of bit rate and spatial resolution is predicted from the nine initial ones. Based on these predictions, a bit rate ladder is extracted and further refined until all VMAF constraints are fulfilled. Experiments show that the algorithm requires 3.6 encodings per provided VMAF on average. A VMAF of 95.07 is achieved on average for the highest provided quality and a VMAF difference between adjacent renditions of 1.92.

Keywords: ABR, bit rate ladder, video streaming, VMAF, subjective quality, neural networks

1. INTRODUCTION

The use of over-the-top media services applying adaptive bit rate (ABR) video streaming is continuously growing. Hereby, digital video signals are encoded at various bit rates and spatial resolutions, resulting in associated qualities. An encoded video signal with a certain bit rate and spatial resolution is denoted as a rendition. Each rendition is associated with a quality. The set of all renditions is denoted as a bit rate ladder. Content Delivery Networks (CDN) are used to provide the encoded video signals to the end-user devices. Depending on the capabilities of the end-user device and the individual transmission rate of the internet connection, a rendition is typically selected with a maximum bit rate that is equal to or lower than the individual transmission rate. For display purposes, each end user device scales the decoded video signal to its display resolution.

A basic generation method is the "one-size-fits-all" fixed bit rate ladder, as described by Aaron et al.[1] or in Apple's specification for creating HTTP Live Streaming (HLS).[2] Here, a fixed predefined set of combinations of bit rates and spatial resolutions is used for each video content, regardless of the content complexity. Using fixed predefined combinations for all video content results in a lower quality for high complexity video content than for low complexity video content. As the quality varies at a given bit rate, a high user experience cannot be guaranteed by such a bit rate ladder.

Nowadays, there are content-dependent bit rate ladder designs that consider the content complexity, in particular per-title,[1] and shot-based[3] encodings. Per-title encoding refers to a bit rate ladder optimization for the entire video content. Shot-based encoding, also known as per-scene encoding, refers to a more granular optimization of bit rate ladders for each shot of each video content. A further aspect that is typically considered in the generation process is the storage costs. To limit them, the number of renditions in a bit rate ladder is typically limited to around 5 to 15.[1,2,4] Reznik et al.[4] propose a variable number of renditions for the bit rate ladder depending on the complexity of the video content.

There are further content-dependent approaches, such as Katsenou et al.,[5] which introduced an approach for estimating a bit rate ladder based on the objective metric Video Multi-Method Assessment Fusion (VMAF).[6] They consider a fixed bit rate range for the ladder and that the next higher bit rate is twice the previous bit rate. They also monitor the slope of the bit rate VMAF curve; when the derivative falls below a certain threshold, no further higher quality is considered. This varies the length of the bit rate ladder. MiPSO[7] uses a scene-based optimization method that optimizes the bit rate ladder for the maximum possible quality or the minimum possible bit rate. Each scene is encoded with a fixed set of combinations of bit rates and spatial resolutions. The combinations that are on the constructed convex hull are selected as renditions. FAUST[8] predicts an optimized bit rate ladder for each scene using an artificial neural network. Hereby, rendition qualities are selected in a specified peak signal-to-noise ratio (PSNR) range with equidistant PSNR values.

One major drawback of these content-dependent bit rate ladder designs is that the renditions are still primarily selected based on the bit rate, as in the case of Katsenou et al.[5] or MiPSO.[7] Whereas FAUST[8] selects the renditions based on the PSNR metric, which has been shown to have a lower correlation with subjective quality than VMAF.[9] In addition, there are several commercial solutions, such as from Bitmovin,[10] CAMBRIA,[11] MUX,[12] and Brightcove,[13] for which the bit rate ladder generation has not been published in detail.

The key aspect of the user experience is the subjective quality. To consider this aspect, Kah et al.[14] developed theoretical foundations for a quality-based bit rate ladder design. The renditions resulting from this design maximize subjective quality under the constraint of minimal bit rate. To judge the subjective quality of the encoded video signal, Kah et al.[14] use VMAF. The VMAF score ranges from 0 to 100, where 0 corresponds to low and 100 to high subjective quality. The VMAF score of a video signal is computed as the average of the VMAF score of each frame. Kah et al.[14] put the focus on the determination of the theoretical VMAF constraints. The authors determined the highest VMAF and the lowest VMAF to be provided with the bit rate ladder, as well as the maximum VMAF difference between each two bit rate-wise adjacent renditions.

The generation of a bit rate ladder fulfilling these constraints faces the major difficulties as explained in the following. Today's encoders cannot be instructed to achieve a certain VMAF score. However, they can be instructed to reach a certain bit rate via the rate control. Figure 1 shows a block diagram of the generation of a rendition. First, the video signal to be coded is scaled to the desired spatial resolution S_{coded}. Second, the scaled video signal is encoded by the video encoder using a bit rate target R_{RC}, which is passed to the rate control of the video encoder. Usually, this rate control bit rate is not exactly reached by the encoder and the resulting bit rate R_{coded} differs from the desired bit rate R_{RC}. Furthermore, the VMAF can only be determined after decoding the encoded video signal and after scaling it back to the original spatial resolution. Another major difficulty is the fact that a video signal with a certain VMAF can be generated by various combinations of bit rate and spatial resolution. These difficulties result in a multidimensional solution space for generating the quality-based bit rate ladder at minimum storage costs. Additionally, this solution space varies with the complexity of the video content.

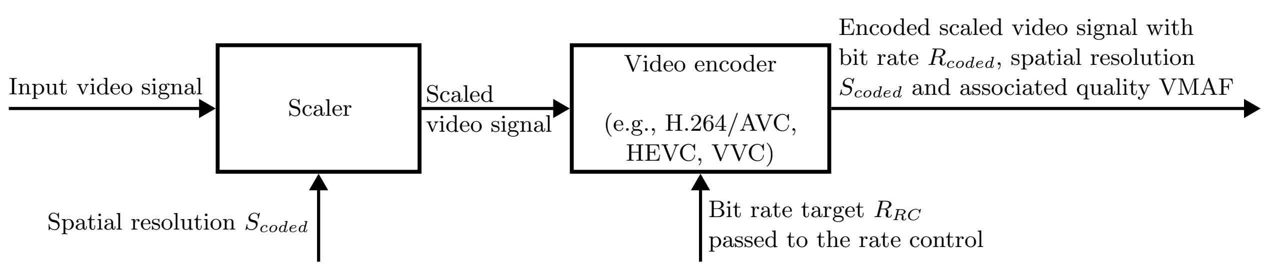

Figure 1. Block diagram of the generation of a encoded scaled video signal.

In this paper, an algorithm is presented for a bit rate ladder generation meeting all theoretical VMAF constraints given by Kah et al.[14] with a limited effort. The key component of this algorithm is a predictor, which can predict the VMAF of a large number of combinations of bit rates and spatial resolutions. Using these predictions the renditions are effectively selected.

The remaining paper is structured as follows. Section 2 summarizes the quality-based bit rate ladder design of Kah et al.,[14] which forms the theoretical background of this work. All details of the presented algorithm are described in Section 3. The evaluation results are presented in Section 4 and Section 5 concludes the paper.

2. A QUALITY-BASED BIT RATE LADDER DESIGN

The quality-based bit rate ladder design of Kah et al.[14] consists of K renditions, each of a certain VMAF score $VMAF_1, \ldots, VMAF_k, \ldots, VMAF_K$, see Figure 2. The bit rates of the renditions are $R_1, \ldots, R_k, \ldots, R_K$. The K VMAF scores are derived from the three quality parameters, which were determined by subjective tests using a 4K TV, as described in the following. The first parameter is the maximum provided quality $VMAF_K$. For a maximum user experience, it should ideally be set to the lowest possible VMAF, for which a video signal is subjectively indistinguishable from the original video signal. This minimizes the bit rate, and thus storage and network costs, while still ensuring optimal subjective quality. Based on Kah et al.,[14] this VMAF score is determined to be 95. The second parameter is the minimum provided quality $VMAF_1$. As stated by Kah et al.[14] it is advised to be set to the lowest VMAF score for which video is still acceptable for watching by the users depending on the situation of permanent viewing or temporary impairment, e.g. of approx. 30 seconds. Such a strategy minimizes encoding and storage costs by avoiding renditions not being watched by users due to unacceptable subjective quality. To achieve an acceptance rate larger than 90 % for free video streaming services or larger than 70 % for paid video streaming services, this VMAF score is set to 79, according to Kah et al.[14]

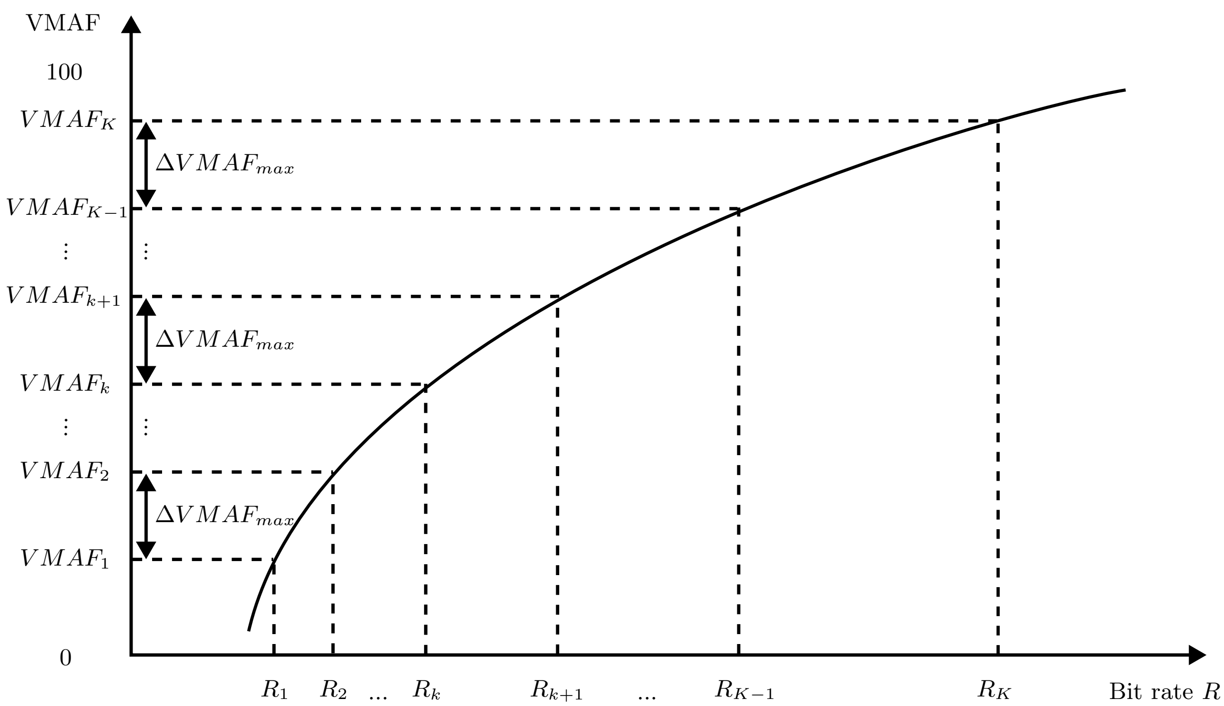

Figure 2. Bit rates $R_1, \ldots, R_K$ and associated VMAF scores $VMAF_1, \ldots, VMAF_K$ of a bit rate ladder limiting the quality difference to $\Delta VMAF_{max}$.

The third parameter of the design is the maximum quality difference between two neighboring renditions k and $k+1$, which can be expressed as $\Delta VMAF = VMAF_{k+1} - VMAF_k$. It should ideally be set small enough such that the subjective quality of the video signal is the same for each pair of neighboring renditions k and $k+1$ in average. This way, any potential quality difference due to not fully exploiting the available transmission rate T of the user's internet connection can be avoided. In addition, switching between neighboring renditions remains subjectively unnoticeable. Thus, the temporal consistency of the video playback is maximized. Based on Kah et al.,[14] this difference is determined to be $\Delta VMAF_{max} = 2$. This results in a quality-based bit rate ladder, which ideally provides each video signal in 9 qualities associated with the VMAF scores 95, 93, ..., 81, and 79. Figure 2 illustrates the bit rate ladder, which limits the quality difference between two neighboring renditions to $\Delta VMAF_{max}$ for all transmission rates T with $R_1 \leq T \leq R_K$.

3. ALGORITHM FOR GENERATING A QUALITY-OPTIMIZED BIT RATE LADDER

The goal of the algorithm is to generate a quality-optimized bit rate ladder that exactly meets the defined VMAF scores $95, 93, \ldots, 81, 79$ and requires a minimum number of encodings for this purpose. Since VMAF is a continuous quantity, the target VMAF scores cannot be achieved precisely. Therefore, a VMAF tolerance interval is introduced, which can vary in size and has always a positive value. The VMAF tolerance interval is added to each target VMAF score. In the following, it will be referred to as ϵ. Depending on the size of the tolerance interval, the VMAF scores are met with varying precision. The parameter ϵ affects the encoding effort and the precision of the achieved VMAF scores: A larger ϵ results in greater deviations from the target VMAF scores. The tolerance interval affects the defined VMAF scores as follows. For the highest quality $VMAF_K$ of the bit rate ladder it holds that

$$95 \leq VMAF_K \leq 95 + \epsilon \tag{1}$$

and for the subsequent lower qualities $VMAF_k$ of the bit rate ladder applies that

$$VMAF_{k+1} - \Delta VMAF_{max} \leq VMAF_k \leq VMAF_{k+1} - \Delta VMAF_{max} + \epsilon \tag{2}$$

where $\Delta VMAF_{max}$ is set to 2: $VMAF_{k+1} - 2 \leq VMAF_k \leq VMAF_{k+1} - 2 + \epsilon$. This ensures that the described VMAF constraints are met. To ensure that $VMAF_1 = 79$ exists during the generation process, a lower boundary $VMAF_{min}$ is set as follows

$$VMAF_{min} \leq VMAF_1 \tag{3}$$

As introduced in Section 2, the quality-optimized bit rate ladder that fulfills the described quality constraints consists of a set of $K = 9$ renditions. Each rendition has a certain $VMAF_k$. The VMAF results from the rate control bit rate $R_{RC,k}$ and spatial resolution $S_{coded,k}$ used for encoding to $VMAF_k(R_{RC,k}, S_{coded,k})$. For simplicity, the rate control bit rate $R_{RC,k}$ is turned into R_k and will be referred to as bit rate in the following. The spatial resolution $S_{coded,k}$ is turned into S_k. The following assumptions are made for the generation of the bit rate ladder: $R_k < R_{k+1}$, $S_k \leq S_{k+1}$, and $VMAF_k < VMAF_{k+1}$.

Figure 3 shows a simplified block diagram of the proposed algorithm, which is described in more detail in the following subsections. The algorithm is applicable to both per-shot and per-title encoding. In the generation process of the quality-optimized bit rate ladder, initial VMAF scores $VMAF$ from performed encodings, additional VMAF scores $\widetilde{VMAF}$ interpolated from the initial VMAF scores and predicted VMAF scores $\widehat{VMAF}$ are used.

3.1 Generation of 3 x 3 initial VMAF scores

As a first step in generating the initial 3×3 VMAF scores $VMAF$, the maximum spatial resolution S_{max} is set to the original resolution of the input video signal and the minimum spatial resolution S_1 is set to 512×288 luminance samples. Subsequently, using these two spatial resolutions, the maximum bit rate B_{max} and the minimum bit rate B_1 are determined. The bit rate B_{max} combined with the maximum spatial resolution S_{max} should result in a quality greater than or equal to $95 + \epsilon$. This is realized by the rate control mode Constant Rate Factor (CRF) of FFmpeg.[15] The CRF value is adjusted until the required quality is achieved. The bit rate B_1 is determined in the same manner. However, the minimum spatial resolution S_1 is used and the quality

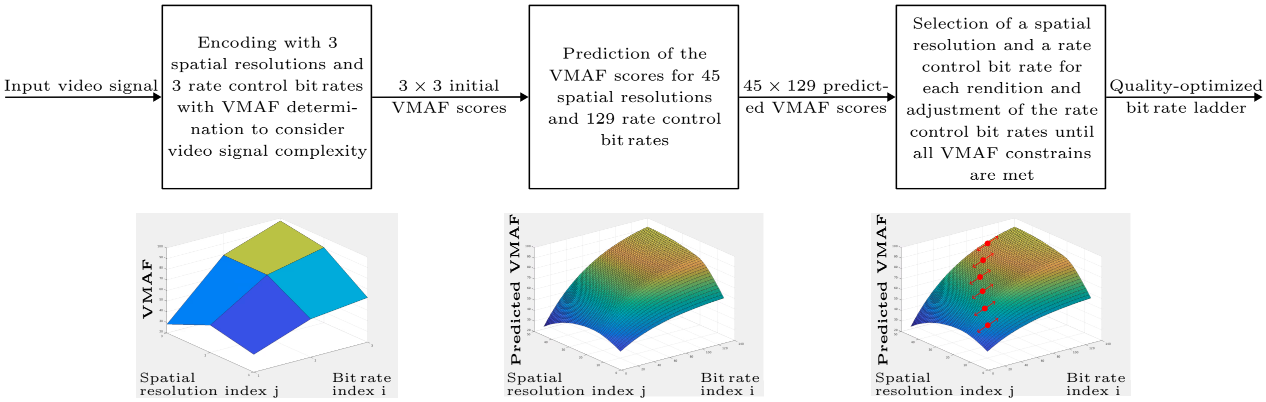

Figure 3. Simplified block diagram of the proposed algorithm.

$VMAF_{min}$ must be achieved. Based on the determined bit rates and spatial resolutions, the remaining spatial resolution and the remaining bit rate are set as follows. A further spatial resolution S_2 is defined, where the following applies: $S_1 < S_2 < S_{max}$ and is set to 1024 x 576 luminance samples. A further bit rate B_2 is defined with $B_1 < B_2 < B_{max}$ according to the following formula

$$\log_2(B_n) = \log_2(B_{n-1}) + \frac{\log_2(B_{max}) - \log_2(B_1)}{N-1} \tag{4}$$

with $n = 2$ and N being the number of bit rates, in this case $N = 3$. After determining the remaining bit rate and remaining spatial resolution, the 7 remaining VMAF scores are determined by encoding using the rate control mode 2-pass average bit rate of FFmpeg. At each encoding, the associated VMAF score $VMAF(B_n, S_m)$, $\forall n = 1, 2, 3$ and $\forall m = 1, 2, 3$ is calculated by decoding and scaling to the original resolution of the input video signal using bicubic filtering of the encoded video signal.

3.2 Generation of 45 x 129 predicted VMAF scores

The initial 3×3 VMAF scores $VMAF(B_n, S_m)$ are used to generate 45×129 predicted VMAF scores $\widehat{VMAF}(B_i, S_j)$, where $i = 1, \ldots, 129$ bit rates and $j = 1, \ldots, 45$ spatial resolutions. This results in 5805 predicted VMAF scores $\widehat{VMAF}(B_i, S_j)$. The bit rates B_i include the initial bit rates B_n plus an additional 126 bit rates calculated based on Equation 4. This results in a total of 129 bit rates. To obtain the predicted VMAF scores $\widehat{VMAF}(B_i, S_j)$, first interpolated VMAF scores $\widetilde{VMAF}(B_i, S_j)$ are generated. The first interpolated VMAF scores $\widetilde{VMAF}(B_i, S_j)$ are generated using the initial VMAF scores $VMAF(B_n, S_m)$ of the three resolutions S_m. For this purpose, a two-term power series model $VMAF = a \cdot B^b + c$ is used to generate these additional 129 VMAF scores for each of the three resolutions S_m. The unknown parameters a, b, and c are estimated with a non-linear least squares method. In a second step, VMAF scores are interpolated for the 42 additional spatial resolutions for each of the 129 bit rates. A piece-wise cubic Hermite interpolation[16] is used. This results in a total of 5805 interpolated VMAF scores $\widetilde{VMAF}(B_i, S_j)$ in which the initial VMAF scores $VMAF(B_n, S_m)$ remain.

To reduce interpolation errors, a neuronal network as shown in Figure 4 is applied to the 5805 interpolated VMAF scores $\widetilde{VMAF}(B_i, S_j)$. For that purpose, a convolutional neural network (CNN) is implemented. To extract different features from the interpolated VMAF scores $\widetilde{VMAF}(B_i, S_j)$, various filter sizes are used in convolutional blocks. The extracted features are then compiled to form an output using fully connected layers. To train the neural network, a large set of 2000 video signals is used, which includes video signals from the BVI-DVC database,[17] Tencent Video Dataset (TVD),[18] certain shots from Blender sequences,[19] and proprietary sequences. To mitigate prediction errors of the VMAF scores $\widehat{VMAF}(B_i, S_j)$, local low-pass filtering is applied to the VMAF vs bit rate curve separately for each resolution, and the resulting curves are clipped to the range of VMAF scores from 0 to 100.

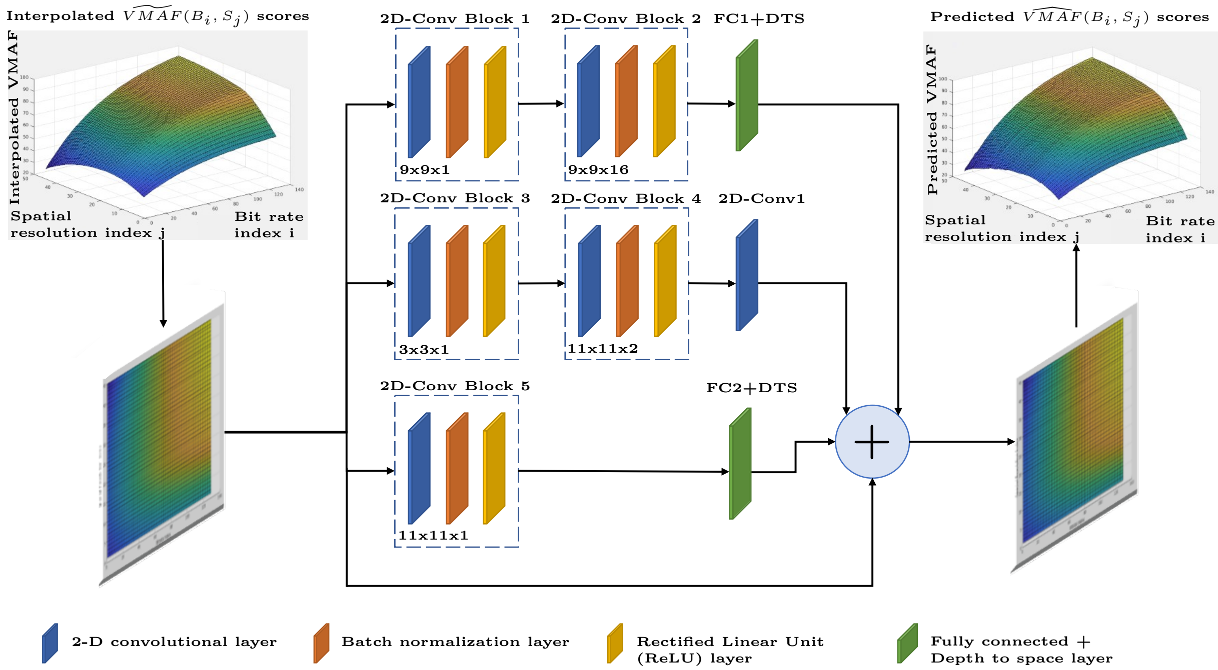

Figure 4. Neural network architecture applied to the interpolated VMAF scores.

3.3 Encoding of renditions from the predicted VMAF scores and quality verification

Starting with the rendition of the highest quality $VMAF_K(R_K, S_K)$, the target VMAF range is shown in Equation 1. All spatial resolutions leading to a VMAF in this range are considered. For each of these spatial resolutions a linear interpolation is performed between the VMAF score $\widehat{VMAF}(B_i, S_j)$ and the VMAF score $\widehat{VMAF}(B_{i+1}, S_j)$ which comprises $VMAF_K = 95 + \frac{\epsilon}{2}$. The VMAF score $VMAF_K$ is set in this way since the associated rate control bit rate is not exactly reached by the encoder and the resulting bit rate can be higher or lower than the desired bit rate. To minimize the bit rate, the spatial resolution that results in the lowest bit rate at the interpolated $VMAF_K$ is selected. This combination of the selected spatial resolution and bit rate is used for encoding of the rendition with the highest quality. $VMAF_K$ of the encoded rendition is calculated and used to verify that the VMAF score is within the range shown in Equation 1. If necessary, an adjustment of the bit rate for the rate control mode 2-pass average bit rate is made at the same spatial resolution until the quality constraint, see Equation 1, is met.

The subsequent renditions are selected based on the VMAF score of the previous renditions and must be in the range shown in Equation 2. For these renditions, it should be noted that the selected spatial resolution for the interpolation is less than or equal to the resolution of the previous rendition. For these renditions, the same iterative procedure of resolution selection, interpolation, and quality verification is performed as for $VMAF_K$.

4. EVALUATION RESULTS

The performance of the algorithm is assessed using a set of 10 HDTV video sequences from the MPEG dataset,[20] which is used in the international video coding standardization. These video signals, comprising for instance ArenaOfValor, BasketballDrive, BQTerrace, Cactus, MarketPlace, and RitualDance, are not included in the training of the neural network. Various values for ϵ are used to generate the quality-optimized bit rate ladder and the impact of ϵ is evaluated. For this evaluation, the average total bit rate of the bit rate ladder and the average number of encodings to generate a rendition are measured for all considered ϵ. To determine the average total bit rate of the bit rate ladder, the bit rates of all 9 renditions resulting form the algorithm are totaled up and then averaged over all 10 video sequences. For the measurement of the average number of encodings to

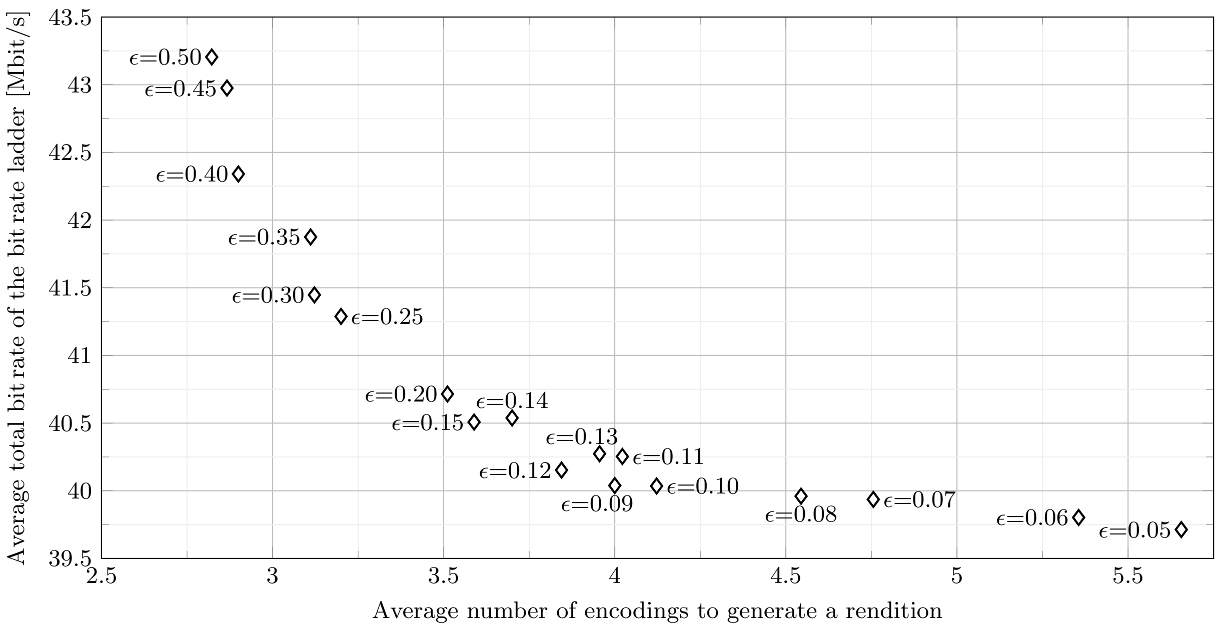

Figure 5. Measured average total bit rate of the bit rate ladder and the average number of encodings to generate a rendition for different values for ϵ.

generate a rendition, the number of performed encodings per video sequence is added up and divided by $K = 9$ renditions. The result is afterwards averaged over the 10 video sequences.

As described in Section 3, the VMAF tolerance interval ϵ can be used to achieve the quality constraints with adjustable precision. Figure 5 shows the results for different values of ϵ. With a large ϵ of 0.5, the algorithm requires an average of 2.8 encodings per rendition. However, ϵ of 0.5 is associated with a higher average total bit rate of 43.2 Mbit/s because the quality constraints are less precisely fulfilled. An average VMAF of 95.28 is achieved for the rendition of the highest quality. The average VMAF difference between two bit rate-wise adjacent renditions is 1.73. A small value for ϵ of 0.05 leads to an average of 5.7 encodings per rendition and an average total bit rate of 39.7 Mbit/s. An average VMAF of 95.02 is achieved for the rendition of the highest quality and the average VMAF difference between two bit rate-wise adjacent renditions is 1.98. When choosing a practical value of $\epsilon = 0.15$, the algorithm requires an average of 3.6 encodings per rendition and results in an average total bit rate of 40.5 Mbit/s. An average VMAF of 95.07 is achieved for the rendition of the highest quality and the average VMAF difference between two bit rate-wise adjacent renditions is 1.92.

5. CONCLUSION

In this paper, an algorithm for a quality-optimized bit rate ladder generation for video streaming services using a neural network is presented. The bit rate ladder is formed by a set of renditions, each of which is defined as an encoded video signal with a certain bit rate and spatial resolution. A bit rate ladder, which maximizes the subjective quality at a minimum bit rate and fulfills the following two constraints was found. First, the VMAF of the highest provided quality should ideally not exceed 95 as all VMAF scores larger than or equal to 95 are on average associated with the same subjective quality as the original input signal. Second, all VMAF differences between two bit rate-wise adjacent renditions should ideally be not greater than 2 as this guarantees indistinguishable subjective quality on average.

The generation of a bit rate ladder fulfilling these constraints faces the major difficulty that today's encoders cannot be directly instructed to achieve a certain VMAF and the VMAF can only be determined after decoding the encoded video signal. Another major difficulty is the fact that a video signal with a certain VMAF can be generated by various combinations of bit rate and spatial resolution. These difficulties result in a multidimensional solution space for generating the quality-based bit rate ladder at a minimum bit rate. Additionally, this solution space varies with the video content.

In this paper, an algorithm is presented which can generate such a bit rate ladder. First, the video signal is encoded for 9 combinations of 3 bit rates and 3 spatial resolutions and the corresponding VMAF scores are calculated. Using a specifically designed and trained neural network, the VMAF of 5805 combinations of 129 bit rates and 45 spatial resolutions are predicted from the 9 ones. Based on these 5805 combinations, a bit rate ladder is extracted, which is further refined afterwards by bit rate adjustments until all VMAF constraints are fulfilled. Experiments show that the algorithm can generate the desired bit rate ladder with an average as low as 3.6 encodings per provided rendition. A VMAF score of 95.07 is achieved on average for the highest provided quality and a VMAF difference of 1.92.

ACKNOWLEDGMENTS

The authors thank the German Federal Ministry of Education and Research for funding this work via the program FHprofUnt 2018, contract number 13FH152PX8.

REFERENCES

[1] Aaron, A., Li, Z., Manohara, M., De Cock, J., and Ronca, D., "Per-title encode optimization." Netflix, 14 December 2015 https://netflixtechblog.com/per-title-encode-optimization-7e99442b62a2. (Accessed: 17 June 2021).

[2] Apple Inc., "HTTP live streaming (HLS) authoring specification for apple devices | apple developer documentation." 20 July 2020 https://developer.apple.com/documentation/http_live_streaming/http_live_streaming_hls_authoring_specification_for_apple_devices. (Accessed: 20 July 2020).

[3] Katsavounidis, I., "Dynamic optimizer — a perceptual video encoding optimization framework." Netflix, 5 March 2018 https://netflixtechblog.com/dynamic-optimizer-a-perceptual-video-encoding-optimization-framework-e19f1e3a277f. (Accessed: 26 May 2020).

[4] Reznik, Y. A., Lillevold, K. O., Jagannath, A., Greer, J., and Corley, J., "Optimal design of encoding profiles for ABR streaming," in [*Proceedings of the 23rd Packet Video Workshop*], *PV '18*, 43–47, Association for Computing Machinery (June 2018).

[5] Katsenou, A. V., Zhang, F., Swanson, K., Afonso, M., Sole, J., and Bull, D. R., "VMAF-based bitrate ladder estimation for adaptive streaming," in [*2021 Picture Coding Symposium (PCS)*], 1–5 (June 2021). ISSN: 2472-7822.

[6] Li, Z., Aaron, A., Katsavounidis, I., Moorthy, A., and Manohara, M., "Toward a practical perceptual video quality metric." Netflix, 6 June 2016 https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652. (Accessed: 27 March 2020).

[7] Malladi, V. P. K., Timmerer, C., and Hellwagner, H., "Mipso: Multi-period per-scene optimization for http adaptive streaming," in [*2020 IEEE International Conference on Multimedia and Expo (ICME)*], 1–6 (July 2020). ISSN: 1945-788X.

[8] Zabrovskiy, A., Agrawal, P., Timmerer, C., and Prodan, R., "FAUST: Fast per-scene encoding using entropy-based scene detection and machine learning," in [*2021 30th Conference of Open Innovations Association FRUCT*], 292–302 (October 2021).

[9] Lee, C., Woo, S., Baek, S., Han, J., Chae, J., and Rim, J., "Comparison of objective quality models for adaptive bit-streaming services," in [*2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)*], 1–4 (August 2017).

[10] Zwantschko, G., "Encoding software: What is per-title encoding?." Bitmovin, 15 November 2017 https://bitmovin.com/per-title-encoding/. (Accessed: 27 March 2020).

[11] Capella Systems, LLC, "Cambria FTC." 14 July 2022 https://capellasystems.net/products/transcoding/cambria-ftc/. (Accessed: 14 July 2022).

[12] Dahl, J., "Instant per-title encoding | mux blog." Mux, 17 April 2018 https://mux.com/blog/instant-per-title-encoding. (Accessed: 26 May 2020).

[13] Brightcove Inc., "Video encoding platform - reduce costs & increase quality." 26 July 2022 https://www.brightcove.com/en/products/zencoder/. (Accessed: 26 July 2022).

[14] Kah, A., Friedrich, C., Rusert, T., Burgmair, C., Ruppel, W., and Narroschke, M., "Fundamental relationships between subjective quality, user acceptance, and the VMAF metric for a quality-based bit-rate ladder design for over-the-top video streaming services," in [*Applications of Digital Image Processing XLIV*], *Proc. SPIE* **11842** (August 2021).

[15] FFmpeg, "A complete, cross-platform solution to record, convert and stream audio and video." 17 April 2018 https://ffmpeg.org/. (Accessed: 2 June 2021).

[16] Fritsch, F. N. and Carlson, R. E., "Monotone piecewise cubic interpolation," in [*SIAM Journal on Numerical Analysis*], **17**, 238–246 (April 1980). ISSN: 1095-7170.

[17] Ma, D., Zhang, F., and Bull, D., "BVI-DVC: A training database for deep video compression," in [*IEEE Transactions on Multimedia*], (September 2021). ISSN: 1941-0077.

[18] Xu, X., Liu, S., and Li, Z., "A video dataset for learning-based visual data compression and analysis," in [*2021 International Conference on Visual Communications and Image Processing (VCIP)*], 1–4 (December 2021). ISSN: 2642-9357.

[19] Xiph.Org, "Test media." 19 July 2022 https://media.xiph.org/. (Accessed: 19 July 2022).

[20] Karczewicz, M. and Ye, Y., "JVET-Y2017: Common test conditions and evaluation procedures for enhanced compression tool testing." JVET, 25 February 2022 https://jvet-experts.org/doc_end_user/current_document.php?id=11473. (Accessed: 1 August 2022).