Accepted in IEEE International Conference on Consumer Electronics (ICCE 2019), 2019

Power-Efficient Video Streaming on Mobile Devices Using Optimal Spatial Scaling

Christian Herglotz, André Kaup (*Fellow IEEE*)
Multimedia Communications &
Signal Processing,
Friedrich-Alexander University
Erlangen-Nürnberg (FAU)
Erlangen, Germany

Stéphane Coulombe
Dept. of Software Engineering and IT
École de technologie supérieure (ÉTS)
Montréal, Canada

Ahmad Vakili
Summit Tech Multimedia
Montréal, Canada
www.summit-tech.ca

Abstract—This paper derives optimal spatial scaling and rate control parameters for power-efficient wireless video streaming on portable devices. A video streaming application is studied, which receives a high-resolution and high-quality video stream from a remote server and displays the content to the end-user. We show that the resolution of the input video can be adjusted such that the quality-power trade-off is optimized. Making use of a power model from the literature and subjective quality evaluation using a perceptual metric, we derive optimal combinations of the scaling factor and the rate-control parameter for encoding. For HD sequences, up to 10% of power can be saved at negligible quality losses and up to 15% of power can be saved at tolerable distortions. To show general validity, the method was tested for Wi-Fi and a mobile network as well as for two different smartphones.

Index Terms—video coding, resolution, low-power, smartphone, video streaming, VMAF

I. INTRODUCTION

During the past years, video streaming applications have conquered the mass markets such that nowadays, watching videos online can be performed with many portable devices, such as a smartphone or a tablet PC. Wireless networks using standards like Wi-Fi, 4G, or the upcoming 5G allow the streaming of high-quality video content at high resolutions. However, a major problem in video streaming is the power consumption of the portable device, because the streaming application can drain the battery quickly. As a consequence, operating times during streaming tend to be short and the consumer's quality of experience (QoE) is strongly impaired.

To tackle this problem, we take a close look at the power consumption of typical smartphones during video streaming. From the literature, we know that the bitrate and the resolution have a major influence on the power consumption in such a way that a lower bitrate and a lower resolution decrease the streaming power significantly [1]. On the other hand, decreasing both values also impairs the visual quality of the displayed video. Therefore, in this work, we perform a thorough analysis of the power consumption and the visual quality at different resolutions as well as bitrates and propose parameter combinations for optimal power-quality trade-offs. It is worth mentioning that in this context, we assume that the bitrate of the video stream is significantly smaller than the
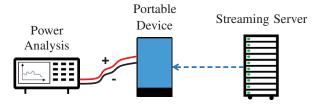

Fig. 1. Power consumption analysis for a portable device during wireless video streaming. The stream is provided by a remote streaming server.

capacity of the transmission channel such that the bitrate does not have to be considered.

To this end, we consider the power consumption of a smartphone that runs a video streaming application (cf. Fig. 1). The streaming video is coded with a common compression format like H.264/AVC [2] or HEVC [3] and streaming is performed via a wireless network. In this work, we restrict our considerations to H.264/AVC because, in 2018, this codec still accounted for more than three-thirds of all coded video data worldwide [4]. The power is considered to be the overall power of the device, which can be measured using the battery connectors. As a consequence, we consider the overall power consumption including network receiver power, decoding, rendering, and display. The application receives the stream, decodes the stream using an internal hardware decoder, and performs upsampling to obtain a full-screen output.

For the evaluation of the visual quality of the output video, we use an objective and a perceptual quality metric, which is trained on subjective data. For exhaustive power considerations, next to true power measurements, we exploit a video decoding energy model from the literature [1], which was successfully employed to model virtual reality streaming applications [5].

The method of reducing the resolution for optimizing the compression efficiency in video coding has been studied thoroughly in the literature. Wang et al. [6] performed an analysis of the distortion induced by downsampling and found that it is barely correlated to coding artifacts caused by quantization. As a consequence, they proposed a method to adaptively downsample image content to increase the compression efficiency
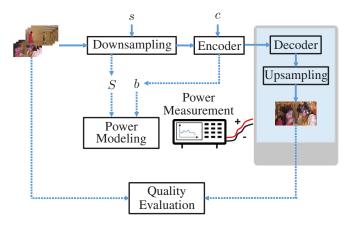
Fig. 2. Framework for the assessment of the mobile device's power consumption and the visual quality of the output video.

at low visual qualities. In a similar direction, Afonso et al. [7] proposed an adaptive frame-based downsampling scheme and obtained bitrate reductions of approximately 4% at the same objective visual quality. Further work showed that using neural networks for image enhancement, one can increase the quality of images distorted by downsampling [8], [9]. For practical applications, Dragić et al. showed that power can be saved by adapting the video resolution to the device's display [10].

In terms of the power consumption, Li et al. constructed a power model for H.264/AVC streaming and proposed to optimize the power-rate costs [11]. However, resolution downscaling was not considered. In a similar direction, a method for optimizing the software decoding energy using decoding-energy-rate-distortion optimization was proposed in [12]. In contrast, in this work we tackle hardware decoding and propose to exploit both resolution downsampling and quantization control for optimal visual qualities at a high power efficiency.

This paper is organized as follows. First, Section II summarizes the simulation setup used for power and quality assessment. Afterwards, Section III discusses the simulation results in detail and provides recommendations for optimal coding decisions. Finally, Section IV concludes this paper.

II. SIMULATION FRAMEWORK

A block diagram of our evaluation framework is depicted in Fig. 2. At first, we downsample the input sequence using bicubic interpolation. In order to do so, we use FFmpeg's internal spatial scaling algorithm, in which the target pixel width w_{target} and pixel height h_{target} can be freely chosen [13]. To determine the target resolution, we use the downscaling factor s which can be used to calculate the target resolution by

$$w_{\text{target}} = \text{round}\left(\frac{w_{\text{in}} \cdot s}{4}\right) \cdot 4, \tag{1}$$

$$h_{\text{target}} = \text{round}\left(\frac{h_{\text{in}} \cdot s}{4}\right) \cdot 4, \tag{2}$$

where w_{in} and h_{in} are the width and the height of the original input sequence, respectively. The rounding operation is

performed to obtain an integer number of pixels. The division by four, as well as the multiplication by four, ensures that the output resolution is covered by the video codec standards, which only allow multiples of four as width and height. As a consequence, s^2 represents the fraction of pixels in relation to the number of pixels in the original sequence. For example, if $s^2 = 0.5$, the total number of pixels after downscaling is half as high as in the original resolution.

Afterwards, the output sequence from downsampling is encoded with the standard encoder included in FFmpeg, i.e., x264 [14]. For the sake of simplicity, we do not change presets and only use the constant rate factor (crf) c for bitrate control, which is generally proportional to the quantization parameter (QP). In contrast to the QP, the crf is supposed to keep the perceptual quality (not the objective quality) of the compressed sequence at a constant level. This is done by encoding frames including a lot of motion with a higher QP and frames with small motion at a smaller QP. Furthermore, using the output resolution $S = w_{\text{target}} \cdot h_{\text{target}}$ and the output bitrate b from encoding, we are able to obtain power estimates from the power model that will be introduced in Subsection II-A.

In addition to power modeling for exhaustive simulations, we performed several power measurements to support our findings (right of Fig. 2). To this end, we measure the power consumption of a smartphone using a power meter attached to its battery connectors. A video streaming application is executed which receives the video stream from a remote server or reads it from the local memory, decodes the stream using an on-chip hardware decoder, upsamples the video to the display's resolution S_{display}, and shows the video on the screen. The measurement setup is the same as in [5].

For quality evaluation of the output video (bottom of Fig. 2), we simulate decoding and upsampling in the portable device using FFmpeg's decoder and the same spatial scaling algorithm as used for downsampling. For the interpolation filter, we again choose bicubic interpolation because it can easily be implemented using on-chip graphics processing units (GPUs) and it has low complexity demands such that the corresponding power consumption on a portable device would be small. We upsample the decoded video to HD resolution and compare with the original sequence that was also upsampled to HD, where we keep the aspect ratio constant (that means, either the width or the height corresponds to HD). The quality metrics used in this work are explained in Subsection II-B.

A. Power Modeling

To measure the impact of the resolution and other factors on the energy or power consumption during video streaming and coding, detailed investigations were performed and published in [1], [5], [15]. In particular, it was found that the power can be accurately modeled using a few parameters, which describe high-level properties of the bit stream. In this work, we take the power model for video decoding from [5] that reads

$$\hat{P} = p_0 + p_f \cdot f + p_S \cdot S + p_b \cdot b, \tag{3}$$

where the variable f is the frame rate, S is the resolution in pixels per frame, and b the bitrate of the video stream. The parameters p_f, p_S, and p_b are model parameters that describe a linear relationship between the variables and the overall power, and p_0 is a constant offset. Results reported in [5] and [1], in which the model was first proposed for energy estimations, indicate that this model is capable of estimating the overall power with a mean error of less than 5%.

In this paper, we exploit the fact that by changing the pixel dimensions of the frames and the quantization parameter during encoding, we can directly influence the overall power with the help of the two variables resolution S and bitrate b (cf. Fig. 2). Since the corresponding parameters p_S and p_b have positive values, we can decrease the overall power consumption. For resolution and bitrate control, we use the downsampling factor s and the crf c.

B. Quality Evaluation

In order to evaluate the visual quality of the output sequence, we exploit two different metrics. The first metric is the PSNR which is an objective metric used to assess the video quality of compressed images. The PSNR is a full reference metric calculated by

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{M^2}{\text{MSE}} \right), \qquad (4)$$

where M is the peak value a pixel can obtain ($M = 255$ for 8-bit sequences) and MSE the mean-square-error between the original sequence and the reconstructed sequence. The PSNR is calculated for each color component Y, U, and V and then averaged to the YUV-PSNR

$$\text{PSNR}_{\text{YUV}} = \frac{1}{8} \left(6 \cdot \text{PSNR}_Y + \text{PSNR}_U + \text{PSNR}_V \right), \quad (5)$$

in which the luma component Y has the highest weight [16]. The overall PSNR of a video is then obtained by calculating the mean YUV-PSNR over all frames.

A major drawback of the PSNR is that it does not always reflect the perceptual quality of a video. To tackle this issue, a lot of research was performed to obtain metrics reproducing the perceptual quality [17]–[19]. In this work, we consider the video multi-method assessment fusion (VMAF) [20] metric, which is a machine learning-based approach. The concept of VMAF is to use a support vector machine (SVM) that exploits the information from different quality metrics and maps them to subjective quality scores. The SVM is trained using a large database from true subjective video quality assessments and it can be shown that VMAF provides high correlations with subjective scores. The subjective scores were obtained using the popular mean-opinion-score (MOS) and the VMAF score maps them to the range $[0, 100]$ with 100 being the best score. For the target of mobile streaming in this paper, we use the VMAF-model version 0.6.1, which includes data that is explicitly trained for smartphones. It is available using the command line flag `--phone-model` [20].

TABLE I
MAIN SPECIFICATIONS OF THE TWO TESTED DUTS. THE DISPLAYS USE THE IN-PLANE SWITCHING (IPS) TECHNOLOGY FOR THIN-FILM-TRANSISTOR (TFT) LCDS.

	Module	Properties
Device A	CPU	Four 32 bit cores (max 2.5 GHz)
	GPU	Up to 578 MHz, 147.9 GFLOPS
	Memory	2 GB of LPDDR 3 (up to 1866 MHz)
	Multimedia	4K video decoding at 60 fps (H.264 or HEVC)
	Display	5 inches, 1080×1920 pixels, TFT/IPS
Device B	CPU	6-core
		4 low-power cores (max 1.4 GHz)
		2 high-performance cores (max 1.8 GHz)
	GPU	Up to 600 MHz, 153.6 GFLOPS
	Memory	3 GB of LPDDR 3 (up to 1866 MHz)
	Multimedia	4K video decoding at 60 fps (H.264 or HEVC)
	Display	5.5 inches, 1440×2560 pixels, TFT/IPS

III. EVALUATION

The evaluation is split into three parts. In the first part, the general setup for power modeling is presented. In the second part, the results from the quality-power assessment are discussed in detail. In the third part, recommendations for practical applications are derived.

A. Power Modeling Setup

For the evaluation of the quality-power performance, we encode 20 different sequences from the HEVC common test conditions [21]. The sequences are given in the resolutions 416×240, 832×480, 1280×720, and 1920×1080. The latter high-definition (HD) resolution is the largest resolution we consider because it is a common resolution of modern smartphone displays. The sequences mainly contain natural content including camera movement (sports, public places, indoor), three sequences were recorded with a static camera similar to videoconferencing, and three sequences show screen content.

The power evaluation is performed for two modern smartphones with the properties listed in Table I. For online streaming, a Wi-Fi connection was established and the video was streamed using a real-time messaging protocol (RTMP). In addition, 3G streaming was tested for device A. Hence, three different cases are evaluated, namely A-Wi-Fi, A-3G, and B-Wi-Fi.

The power model for the three tested cases (3) was trained using the sequences introduced above encoded with FFmpeg at original resolutions and crf values of 18, 23, 28, and 33, which corresponds to a 'subjectively sane' range [13]. To have more information on resolutions, four sequences were encoded at quarter common intermediate format (QCIF) resolution. All in all, the set for training and validation includes 86 single measurements. The training for optimal parameter values

	ε	p_0	p_f	p_S	p_b
A-Wi-Fi	3.80%	1.45	$7.83 \cdot 10^{-3}$	$7.06 \cdot 10^{-8}$	$1.08 \cdot 10^{-8}$
A-3G	3.41%	1.70	$7.75 \cdot 10^{-3}$	$7.40 \cdot 10^{-8}$	$1.02 \cdot 10^{-8}$
B-Wi-Fi	4.58%	1.20	$1.30 \cdot 10^{-2}$	$5.19 \cdot 10^{-8}$	$6.22 \cdot 10^{-9}$

TABLE III
SQUARED SCALING FACTORS s^2 AND CRF VALUES c FOR POWER-QUALITY
EVALUATION.

s^2	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.25	0.2
c	18	20	22	24	26	28	30	32	34	36

was performed using least-squares curve fitting. The mean estimation error over all M measurements is calculated by

$$\varepsilon = \frac{1}{M} \sum_{m=1}^{M} \left| \frac{\hat{P}_m - P_m}{P_m} \right|, \qquad (6)$$

where m is the measurement index, $\hat{P}_m$ the estimated power of the m-th measurement from (3), and P_m the measured power of the m-th measurement. To obtain the error, cross-validation was performed by separating the whole set into subsets, where each subset is based on a single input sequence. The resulting mean errors and the trained parameter values are given in Table II. As the mean error is below 5% for all tested cases, the model is highly accurate in estimating the true power of the streaming process.

B. Results

For an exhaustive analysis, we evaluate the visual quality and the streaming power for all permutations of the downscaling factors s^2 and crf values c shown in Table III. A subset of the resulting power-distortion curves depending on the coding resolution is plotted in Fig. 3 (BQTerrace sequence, case A-Wi-Fi). Each curve corresponds to a constant scaling factor, and the markers on the curves represent crf values, where the small crf values are located on the right.

The curves indicate that to obtain a maximum quality at low power, different combinations can be chosen. If high qualities are desired, it is helpful to choose the maximum resolution $s^2 = 1$ and only change the crf (blue curve). If power shall be saved, it makes sense to reduce the resolution significantly ($s^2 = 0.2$) and compress with medium crfs (green curve). The pink curve represents the Pareto front which links all combinations $\{s^2, c\}$ that are Pareto-efficient on a bilinearly interpolated curve.

In addition to PSNR considerations, we evaluate the distortion-power curve in terms of the VMAF metric to obtain perceptual results. The curves are plotted in Fig. 4. Again, we can see that the Pareto curve covers multiple values of the scaling factor and the crf. Interestingly, the VMAF scores
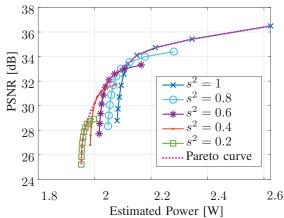

Fig. 3. Power-distortion curves for the BQTerrace sequence of test case A-Wi-Fi. Each curve represents a constant scaling factor as shown in the legend. Each marker on the curve represents a crf value, where the smallest value is located on the right. The Pareto curve links all Pareto efficient markers including points from the not displayed curves $s^2 \in \{0.9, 0.7, 0.5, 0.3, 0.25\}$.
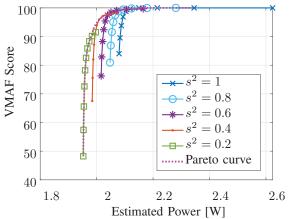

Fig. 4. Power-VMAF curves for the BQTerrace sequence and case A-Wi-Fi. Each curve represents a constant scaling factor as shown in the legend.

for the top-right markers all yield the maximum value 100, such that the Pareto curve does not begin with the maximum resolution and highest quality.

We take a closer look at the optimal combinations for the scaling factor and the crf $\{s^2, c\}$ by plotting the values corresponding to the Pareto curve in Fig. 5. The curve confirms that to obtain the highest possible VMAF quality score, it is not required to use both a maximum resolution and a minimum crf. In particular, we only observe one combination in which the crf is smaller than 22. Furthermore, it is more helpful to reduce the resolution than using a crf larger than 24. It is worth mentioning that with a scaling factor of $s^2 = 0.5$ and a crf of 22, we still obtain a very high VMAF score of more than 98. These results show that resolution scaling is a valid method to allow power efficient streaming at high visual qualities.
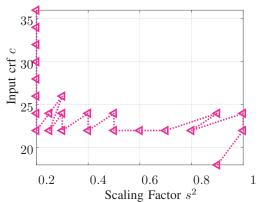
Fig. 5. $\{s^2, c\}$-combinations from the Pareto curve in Fig. 4 (BQTerrace sequence, case A-Wi-Fi). The marker at $\{s^2 = 0.9, c = 18\}$ corresponds to the top right position of the Pareto curve in Fig. 4.
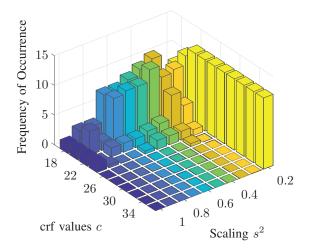

Fig. 6. Histogram on Pareto-efficient $\{s^2, c\}$-combinations collected for HD and 1280×720-sequences from Subsection III-A.

C. Derivation of Recommendations

In order to obtain a general recommendation for an optimal choice of the crf and the scaling factor, we perform a histogram analysis for all sequences introduced in Subsection III-A and evaluate them on all tested devices. The histogram is derived by collecting all $\{s^2, c\}$-combinations on the Pareto curve (Fig. 5). Checking the separate histograms for each test case (A-Wi-Fi, A-3G, and B-Wi-Fi), we only find minor differences such that we neglect them in the following.

At first, we consider HD sequences as well as sequences at the resolution 1280×720 (720p) and plot the histogram in Fig. 6. For each $\{s^2, c\}$-combination shown on the x- and y-axis, we count how often the combination occurs on the Pareto curves. For example, the combination $\{s^2 = 1, c = 18\}$ was only Pareto efficient in three cases. In contrast, $\{s^2 = 0.4, c = 20\}$ corresponds to a Pareto efficient combination in 21 cases (largest green block). The yellow bars at the right correspond to very low visual qualities (the green curve in Fig. 3 and 4)

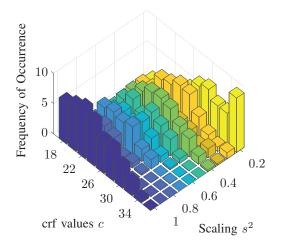| s^2	1	0.9	0.7	0.4	0.2
c	18	20	20	20	20
VMAF	100	99.997	99.88	97.67	89.6
Δp (720p)	0%	1.67%	2.66%	4.13%	5.13%
Δp (HD)	0%	7.14%	9.36%	12.66%	14.81%

Fig. 7. Histogram on Pareto-efficient $\{s^2, c\}$-combinations collected for low-resolution and screen content sequences from Subsection III-A.

such that we do not consider them in the following.

The distribution of the bars in Fig. 6 suggests that to save power, downscaling at a fixed crf of $c = 20$ is more beneficial than increasing c. As a consequence, we propose to use the combinations listed in Table IV in practice. For these combinations, we evaluate all sequences in all test cases. As criterion, we take the mean VMAF score as well as the mean relative power savings Δp with respect to $\{s^2 = 1, c = 18\}$ and also list them in Table IV. We can see that with negligible losses in quality (mean VMAF score of 99.88), power savings up to 10% can be achieved for HD sequences. Accepting mean VMAF scores of around 90, power savings up to 15% can be reached.

For the other sequences (screen content and low resolution (LR)), we plot the histogram in Fig. 7. We can see that the distribution is much wider and shifted towards higher crf values. The dark blue bars at $s^2 = 1$ correspond to screen content, which indicates that scaling is generally not beneficial if the sequence contains high spatial frequencies (e.g., text).

However, to determine potential power savings, we calculate the mean VMAF score and the mean relative power savings for some interesting combinations with a high frequency of occurrence, cf. Table V. Apparently, maximum power savings are below 5%, even accepting low qualities (below a VMAF score of 70). The reason is that due to the linear relation be-

TABLE V
MEAN VMAF SCORES AND RELATIVE POWER SAVINGS FOR DIFFERENT
COMBINATIONS OF $\{s^2, c\}$ WITH RESPECT TO $\{s^2 = 1, c = 18\}$. VALUES
ARE GIVEN FOR LOW-RESOLUTION SEQUENCES (LR) AND SCREEN
CONTENT.

s^2	1	1	1	0.7	0.25
c	18	22	28	22	26
VMAF (LR)	97.53	92.59	77.35	77.03	26.24
VMAF (screen)	100	100	99.23	95.75	68.31
Δp (LR)	0%	0.81%	1.40%	1.39%	2.35%
Δp (screen)	0%	1.01%	1.92%	2.47%	4.82%

tween the resolution, i.e., the number of pixels per frame, and
the power (cf. (3)), high power savings can only be obtained
by high-resolution sequences. For example, a resolution of
832×480 holds less than one 5th of the pixels per frame
as HD, which also prunes potential power savings by a factor
of 5. As a consequence, we conclude that spatial scaling is
not beneficial for screen content as well as for resolutions of
832×480 and below.

IV. Conclusions

This paper has shown that spatial scaling can be exploited
to optimize the power efficiency and the perceived visual
quality in mobile video streaming applications. In particular,
we showed that HD streaming and videoconferencing at high
resolutions can be optimized. With the help of the proposed
set of parameters, up to 10% of power can be saved for near-
optimal visual quality and up to 15% of power can be saved
for acceptable visual qualities.

In future work, a general model for the relation between
the perceived subjective quality and the resolution could be
established and exploited in rate-distortion optimization. Fur-
thermore, the sequence-specific frame rate could be considered
to obtain additional power savings at decent qualities.

Acknowledgment

References

[1] C. Herglotz and A. Kaup, "Decoding energy estimation of an HEVC hardware decoder," in *Proc. International Symposium on Circuits and Systems (ISCAS)*, Firenze, Italy, May 2018, pp. 1–5.
[2] *Advanced Video Coding for Generic Audio-Visual Services*. ITU-T Rec. H.264 and ISO/IEC 14496-10 (AVC), ITU-T and ISO/IEC JTC 1, Apr 2003.
[3] *High Efficiency Video Coding*. ITU-T Rec. H.265 and ISO/IEC 23008-2, ITU-T and ISO/IEC JTC 1/SC 29/WG 11 (MPEG), Apr 2013.
[4] statista. (2019, 5) Market share of top online video codecs and containers worldwide from 2016 to 2018. https://www.statista.com/statistics/710673/worldwide-video-codecs-containers-share-online/.
[5] C. Herglotz, S. Coulombe, S. Vakili, and A. Kaup, "Power modeling for virtual reality video playback applications," in *Proc. IEEE International Symposium on Consumer Technology (ISCT)*, Ancona, Italy, June 2019.
[6] R. Wang, C. Huang, and P. Chang, "Adaptive downsampling video coding with spatially scalable rate-distortion modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 11, pp. 1957–1968, Nov 2014.
[7] M. Afonso, F. Zhang, A. Katsenou, D. Agrafiotis, and D. Bull, "Low complexity video coding based on spatial resolution adaptation," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 3011–3015.
[8] M. Jenab, I. Amer, B. Ivanovic, M. Saeedi, Y. Liu, G. Sines, and S. Shirani, "Content-adaptive resolution control to improve video coding efficiency," in *Proc. IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2018, pp. 1–4.
[9] S. Hamis, T. Zaharia, and O. Rousseau, "Image compression at very low bitrate based on deep learned super-resolution," in *Proc. IEEE International Symposium on Consumer Technology (ISCT)*, Ancona, Italy, June 2019.
[10] L. Dragić, D. Hofman, M. Kovač, M. Žagar, and J. Knezović, "Power consumption and bandwidth savings with video transcoding to mobile device-specific spatial resolution," in *Proc. 9th International Symposium on Communication Systems, Networks Digital Sign (CSNDSP)*, July 2014, pp. 348–352.
[11] X. Li, Z. Ma, and F. C. A. Fernandes, "Modeling power consumption for video decoding on mobile platform and its application to power-rate constrained streaming," in *Proc. Visual Communications and Image Processing (VCIP)*, San Diego, USA, Nov 2012.
[12] C. Herglotz, A. Heindel, and A. Kaup, "Decoding-energy-rate-distortion optimization for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 172–181, 2019.
[13] (2018) Fast Forwards MPEG (FFmpeg). http://ffmpeg.org/. Accessed 201-11.
[14] x264: Encoder for H.264/MPEG-4 AVC video compression. x264.org. Accessed 2018-11.
[15] C. Herglotz, D. Springer, M. Reichenbach, B. Stabernack, and A. Kaup, "Modeling the energy consumption of the HEVC decoding process," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 1, pp. 217–229, Jan 2018.
[16] J. Ohm, G. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards - inlcuding high efficiency video coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1669–1684, Dec 2012.
[17] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2959–2965, Dec 1995.
[18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.
[19] F. Zhang, A. Mackin, and D. R. Bull, "A frame rate dependent video quality metric based on temporal wavelet decomposition and spatiotemporal pooling," in *IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 300–304.
[20] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," *The Netflix Tech Blog*, vol. 6, 2016. [Online]. Available: https://medium.com/netflix-techblog/toward-a-practical-perceptual-video-quality-metric-653f208b9652
[21] F. Bossen, "JCTVC-L1100: Common test conditions and software reference configurations," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Tech. Rep., Jan 2013.