

Perceptually Equivalent Resolution in Handheld Devices for Streaming Bandwidth Saving

Mateo Cámara¹, César Díaz¹, Juan Casal, Jorge Ruano, and Narciso García

Abstract—We present the description, results, and analysis of the experiments conducted to find the equivalent resolution associated with handheld devices. That is, the resolution from which users stop perceiving quality improvements if better resolutions are presented to them in such devices. Thus, it is the maximum resolution that it is worth considering for generating and delivering video, as long as sequences are not too intensively compressed. Therefore, the detection of the equivalent resolutions allows for notable savings in bandwidth consumption. Subjective assessments have been carried out on fifty subjects using a set of video sequences of very different nature and four handheld devices with a broad range of screen dimensions. The results prove that the equivalent resolution in current handheld devices is 720p as higher resolutions are not valued by users.

Index Terms—Adaptive streaming, bit rate, video quality, subjective assessment.

I. INTRODUCTION

VIDEO traffic accounts for more than 75% of all IP traffic today, and this percentage is expected to grow noticeably over the next years. Furthermore, smartphone traffic will reach approximately one third of total IP traffic by 2021 [1]. Additionally, despite the use of significantly more efficient video compression standards like H.264/AVC [2], H.265/HEVC [3], or VP9 [4], the bit rates required to provide sufficiently good quality content to end users are very high [5], due in part to increasingly higher user demands. Faced with these overwhelming data, service and content providers are constantly searching for means to provide quality of experience (QoE) at the lowest bit rate possible.

Considering HTTP/TCP-based adaptive bit rate streaming (ABR) techniques, the initial approach to generate content at different qualities is the use of general purpose fixed encoding ladders [5]. These ladders consist of tables that explicitly indicate the values of several parameters (e.g. image resolution,

framerate, bit rate...) to create a number of quality levels. Nevertheless, this approach cannot assure that bandwidth is used efficiently, as, on the one hand, bit rate can be wasted in encoding content of low complexity (e.g. news, cartoons, sitcoms...) or, on the other hand, not enough bit rate can be assigned for high complexity content (e.g. soccer, action movies...). These drawbacks can however be limited by using more flexible ladders that allow to specify ranges of values instead of predetermined ones [6] and/or by means of using not one but several subladders focused on a particular type of content (per-category encoding) or a targeted (type of) device (per-device encoding).

Regarding the per-category strategy, the approach is however insufficient, as sequences classified into the same category might not be of comparable level of complexity. This is why Netflix pioneered the implementation of per-title encoding (i.e., using different encoding settings per sequence [7]), recently evolved into even more specific per-shot or per-chunk encoding schemes, which are able to capture complexity variations of the content over time [8]. Meanwhile, most per-device implementations simply adapt the value of the encoding parameters according to the type of device in question or directly remove some of the quality levels. This is done, for instance, not including substandard or standard definition levels for Smart TVs or Ultra High Definition (UHD 4K) for mobile phones. So far, this has been done rather automatically, that is, without considering the actual specifications of the devices and how they impact the quality actually perceived by users. Moreover, the tendency is to generate increasingly specific subladders only targeting devices with screens of very similar dimensions. Considering these elements, it is desirable to introduce a perceptually-driven preparation of device-aware encoding ladders to properly assess which devices should use a common ladder and the characteristics of this ladder to improve efficiency rationally.

In this respect, this work describes the design, development, and analysis of experiments to find the so-called equivalent resolution actually perceived by users common to all handheld devices. The equivalent resolution is defined as the resolution from which users do not statistically significantly perceive an improvement if sequences with objectively better resolutions are presented to them. So we look to prove whether visual conditions limit the resolution needs, as stated in [9], and to validate our initial hypothesis that the maximum resolution that can be perceived by users in devices with screens equal or smaller than that of a tablet is statistically the same and so they all should share a common encoding subladder with the above mentioned equivalent resolution as the maximum one.

The letter is structured as follows. In Section II we describe the most significant works related to this topic. In Section III we present all the features of the conducted subjective assessment. Then, we present and analyze the results of the tests in Section IV. Finally, the conclusions are included in Section V.

II. RELATED WORK

Numerous works dealing with how users perceive audiovisual information are not just focused on studying the effect of variables strictly related to the data, but they also consider exogenous variables such as the type of display, the visualization environment or the behavior of the network. In particular, Ni *et al.* [10] analyzed the variation of the perceived quality of a mixture of media on screens with the display pixel density. The study verified that devices with a higher pixel density were better scored. Focusing on text, the work by Mayr *et al.* [11] enriched the environment including parameters related to the environment and comfort of the subject. Furthermore, the study by Lischke *et al.* [12] included, in addition to text, comparisons between images and video quality assessment and concluded that the three tasks have significantly different requirements in terms of pixel density. Additionally, considering specifically the video signal, other studies [13], [14] analyzed the assessment of video sequences including the possibility that the mobile device and the size of the screen may have some influence. Furthermore, the works by Ou *et al.* [15], and Laghari *et al.* [16] examined the influence of video characteristics and format in the user QoE, with a huge statistical analysis based on subjective assessment. In this respect, Jumisko *et al.* [17] analyzed the acceptance threshold, i.e., the point above which users statistically accept the quality provided by a service considering the terminal of the user. Additionally, considering other external aspects besides the device and the video features, the works by Tavakoli *et al.* [18] and Pérez *et al.* [19] presented studies where not only the impact in QoE of the encoding process, resolution and bit rate were explored but also that of the network behavior. Finally, regarding the best quality that viewers can perceive in different environments, as concluded from the previously cited works and more explicitly by Allison *et al.* [20], it is quite difficult to find a universal definition of what can be considered as “visually lossless” to users. The reason is that this attribute strongly depends on a number of elements besides the information characteristics, such as display features, viewing conditions, familiarity with the content and others.

Thus, considering this complex context, our work intends to shed light on one specific but significant aspect: the relation between video resolution and display screen size and its impact in the perceived quality, particularized for handheld devices and regardless of any other internal or external video variables. In this way, we can characterize these devices from the point of view of video resolution with the aim of creating effective encoding and transmission schemes to save bandwidth.

III. TEST FEATURES

All the procedures and considerations that have been taken into account during the different test phases are presented. Many of them are based on the operational guidelines included in

TABLE I
SRCs’ SPATIAL AND TEMPORAL COMPLEXITY

	Game	Venice	Football	India	Skate
Spatial Complexity	42.2	80.7	62.7	70.7	38.5
Temporal Complexity	14.2	11.7	18.0	13.3	16.4

Recommendations ITU-R BT.500-13 [21], ITU-T P.910 [22] and ITU-T P.913 [23].

A. Material

The test material consists of five 10-second-long very-high-quality UHD 4K Source sequences (SRCs), that is, 16:9, 2160p, acquired at 50 or 60 fps. They were chosen mostly among those recommended by the Video Quality Experts Group (VQEG) to include representative, varied, and habitual contents for observers. This material is characterized in terms of the average spatial and temporal complexity considering the spatial information (SI) and the temporal information (TI) indicators [24], [22]. As shown in Table I, the selected SRCs cover a wide range of complexities, from low to high values in both domains. Furthermore, all the SRCs include audio.

We have considered five Hypothetical Reference Circuits (HRCs): 16:9 aspect ratio at 1080p, 720p, 540p, 360p, and 270p. These set is close to a geometric progression with ratio $1/\sqrt{2}$, so each new frame size is close to half the previous one. All HRCs were AVC encoded at 30 fps, after conducting an exhaustive analysis of the state of the art [25], [26]. A target bit rate of 15 Mbps was selected to ensure that encoding parameters did not impact the quality presented to subjects. This bit rate is well above the highest bit rates used for TV sets [5], [7], therefore clearly higher than the bit rates considered for handheld devices. Thus, HRCs only differed in the picture resolution. Finally, we had a total of 25 Processed Video Sequences (PVSS) under evaluation.

B. Mobile Devices

We have selected three smartphones with very different screen sizes and native resolutions, and one tablet. In order to prevent the tablet from distorting subjects ratings, it was always shown either first or last, but never in between. In particular, the handheld devices chosen were the following:

- *Smartphone 1 (SP1)*: with a 4.3-inch screen and a resolution of 480×800 pixels (LG-P720).
- *Smartphone 2 (SP2)*: with a 4.7-inch screen and a resolution of 720×1280 pixels (Samsung Galaxy A3 (2017)).
- *Smartphone 3 (SP3)*: with a 5.1-inch screen and a resolution of 1440×2560 pixels (Samsung Galaxy S7).
- *Tablet*: with a 9.7-inch screen and a resolution of 1536×2048 pixels (iPad Air 2).

The same player has been used in all devices to assure a common up-sampling method. Whenever required, bicubic interpolation has been used for both luma and chroma.

C. Environment

The test room was set as comfortable as possible, with the aim of simulating to the greatest extent possible the habitual location where the subjects generally use their mobile devices to enjoy

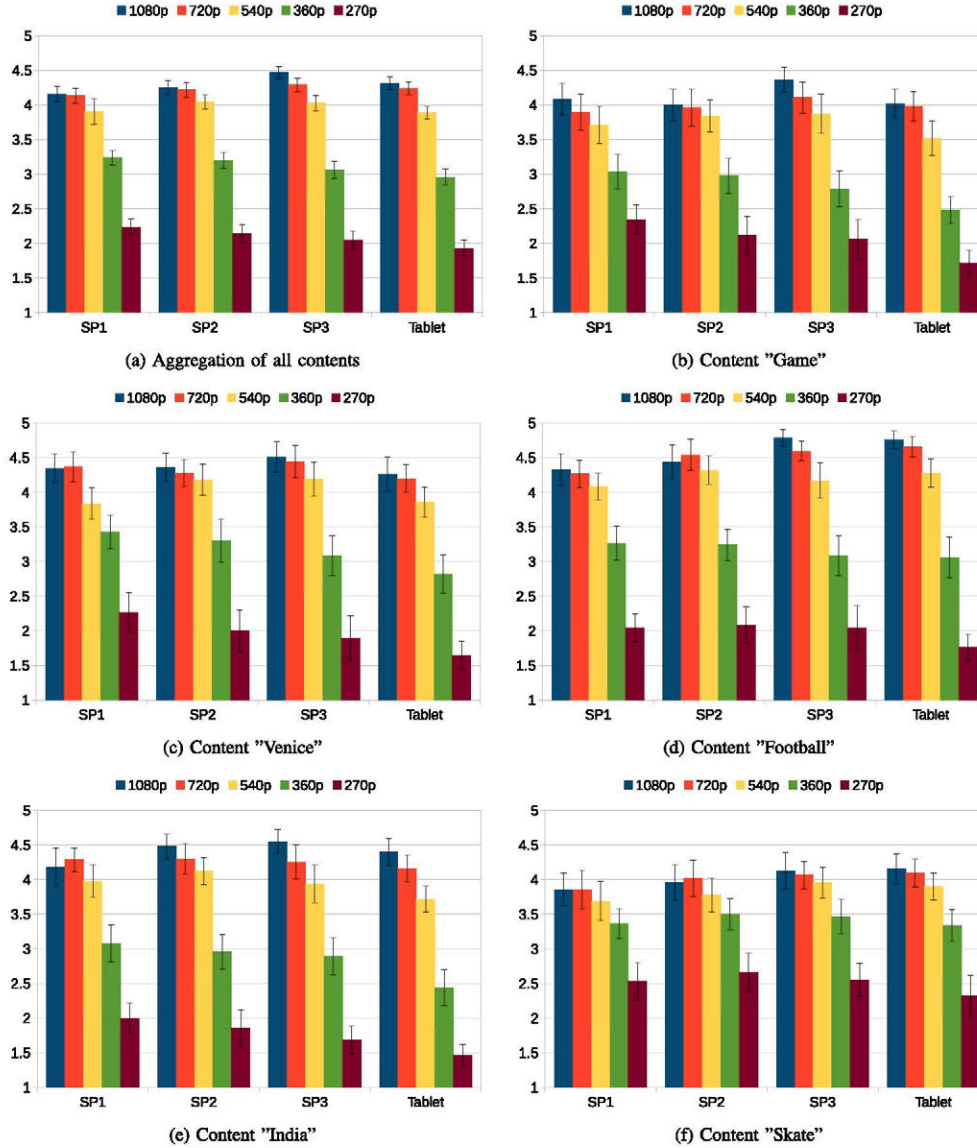


Fig. 1. MOS: global (a) and per-content (b to f), disaggregated per quality and device.

audiovisual content. In addition, the brightness was controlled so as to meet recommended values [22]: 69.6 Lx in front of the subjects, 80.6 Lx to their left, 94.9 Lx to their right, and 120.6 Lx above them.

Subjects were allowed to choose the viewing distance to the device they felt more comfortable with. Moreover, they could hold the mobile in their hands or leave it on the table. In this sense, we did not detect any correlation between viewing distance and screen size. Additionally, subjects used headphones to listen to audio.

D. Methodology

Before the test starts, the test designer reads out loud the guidelines on how the subjects should perform the tests. Next, before the use of each device, a rather short training session is carried out. Subjects are presented with two possible conditions: the highest and lowest qualities that will be shown, i.e., 1080p and 270p. To that end, a sixth sequence is used. In this way, subjects are more aware of the scale of qualities that they will encounter

and rate the sequences accordingly. Furthermore, this training session was also used to allow the subject to tune the audio level amplitude as he/she pleases.

The conducted tests consisted in sequentially presenting all the PVS's in random order to all subjects in all the considered devices. The subjects were asked to rate on paper each sequence immediately after its visualization. To help it, a four-second grey sequence was included between every two consecutive PVS's. The test method followed in the assessment is the Absolute Category Rating (ACR), where subjects have five possible answers to indicate the quality just perceived: "Excellent", "Good", "Fair", "Poor" and "Bad".

IV. TEST RESULTS

Fifty non-expert people participated in the test. Women and men on par, aged between 18 and 30 years, most of whom use mobile devices frequently to watch audiovisual content.

Fig. 1 depicts the mean opinion scores (MOS's) per resolution and mobile device. Results are presented both aggregated

(Fig. 1a) and per content (Figs. 1b to 1f). All bars include their 95% confidence intervals (CIs).

As expected, the MOS increases with the number of image pixels in all the cases. However, some CIs clearly overlap. In particular, the CIs associated with resolutions 1080p and 720p overlap for all the combinations. Furthermore, the CIs related to resolutions 720p and 540p overlap in a subset of the cases. Given these results, the main conclusion is then clear: subjects perceived no statistically significant difference between resolutions 1080p and 720p, regardless of the device. Furthermore, as described, in some of the combinations, they did not even perceive any statistically significant difference between other resolutions, in particular resolutions 720p and 540p, but this result cannot be extended to all the considered cases. Therefore, the equivalent resolution for handheld devices is 720p. Moreover, some interesting conclusions are that subjects tend to rate higher contents of lower resolutions on smaller-screen devices and contents of higher resolutions on greater screens devices. Additionally, we can observe that the smaller the screen of the mobile device is, the more the scores of all contents tend to homogenize. As the characteristics of the terminal screen improve, higher differences can be found. This shows that the demands in low-end mobiles are lower than in high-end mobiles, whose exigency is greater.

Regarding the results disaggregated per SRC, the trends mentioned above also occur for individual contents. Considering “India”, lower resolutions were rated slightly worse than for the rest of SRCs, presumably because at low resolutions, small objects that are key to perceiving a good quality in this specific content may “disappear”. Regarding “Skate”, the average rates given to the highest resolutions are quite low compared to the rest of the contents and many CIs overlap. Furthermore, the CIs are in general larger than they are for the rest of the SRCs, which means a weaker consensus of the subjects. Moreover, there seems to be a tendency that makes more extreme the scores given to contents with higher SI. That is, high-SI content high resolutions are better rated on average than low-SI content high resolutions and high-SI content low resolutions are rated worse in average than low-SI content low resolutions. By contrast, the TI indicator does not seem to have this kind of effect. Finally, it is also noteworthy mentioning that besides the objective characteristics of the sequences, result differences between contents might also be a result of the personal taste of subjects towards the specific material presented to them.

Additionally, to validate the randomization made to present the samples, we have analyzed the average score given to a particular resolution taking into consideration the resolution of the sequence visualized immediately before. Results related to 1080p resolution are presented in Fig. 2. The rest of the resolutions are not included since they yield similar results. As can be observed, results are in general random and CIs overlap. Only a small difference has been detected when subjects have no previous references. However, the number of occurrences of these cases is very low and non-statistically significant.

Finally, in Fig. 3, we show the results of the analysis of the influence of the order in which devices were presented to subjects on the final scores. We can see in the graph that the CIs

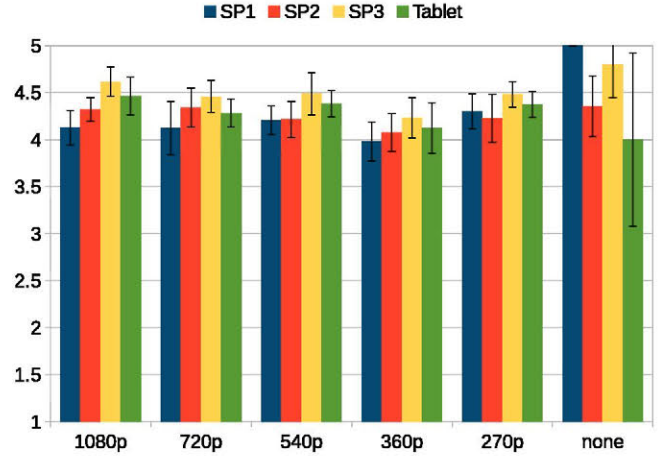


Fig. 2. MOS given to the 1080p resolution considering the previous one.

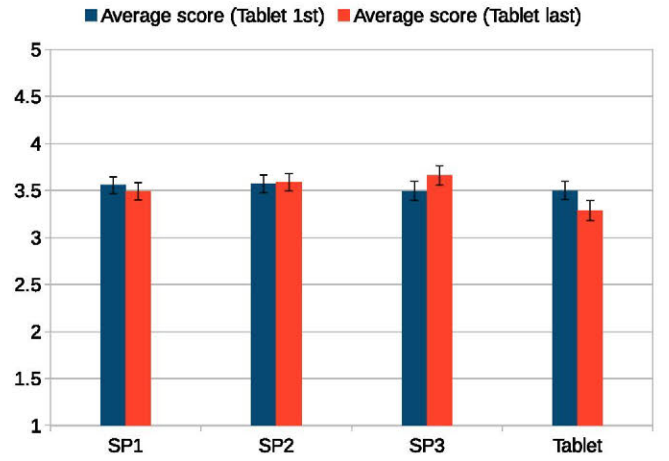


Fig. 3. Comparison between the average score given to each device when the Tablet was evaluated either first or last.

overlap for the mobile devices but there is a small but significant difference in the case of the tablet, indicating that it was rated better when it was the first device shown, presumably due to user expectations regarding quality.

V. CONCLUSION

We have presented the tests conducted to identify the so-called equivalent resolution in handheld devices, i.e., the maximum resolution that can be statically significantly perceived by users on such devices. This value has been found to be 720p, considering 16:9 aspect ratio videos and given that the video sequence is not damaged perceptively in the encoding process. This result can be used by service and content providers to set the maximum quality level that it is worth providing to users watching their content through mobile devices so that no bit rate is wasted. This conclusion is the main outcome of an assessment conducted on a large number of subjects (50 people of both genders alike and aged between 18 and 30 years), using sequences of unequal spatial and temporal complexities and four handheld devices with very different screen sizes.

REFERENCES

- [1] Cisco, “Visual Networking Index: Forecast and Methodology, 2016–2021,” 2017. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf>
- [2] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
- [3] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
- [4] A. Grange, P. de Rivaz, and J. Hunt, “VP9 bitstream & decoding process specification,” Mar. 2016. [Online]. Available: <https://storage.googleapis.com/downloads.webmproject.org/docs/vp9/vp9-bitstream-specification-v0.6-20160331-draft.pdf>
- [5] Apple, “HLS authoring specification for apple devices,” 2018. [Online]. Available: https://developer.apple.com/documentation/http_live_streaming/hls_authoring_specification_for_apple_devices
- [6] YouTube, “Live encoder settings, bitrates, and resolutions,” 2018. [Online]. Available: https://support.google.com/youtube/answer/2853702?hl=en&ref_topic=6136989
- [7] A. Aaron, Z. Li, M. Manohara, J. de Cock, and D. Ronca, “Per-title encode optimization,” 2015. [Online]. Available: <https://medium.com/netflix-techblog/per-title-encode-optimization-7e99442b62a2>
- [8] I. Katsavounidis, “Dynamic optimizer—A perceptual video encoding optimization framework,” 2018. [Online]. Available: <https://medium.com/netflix-techblog/dynamic-optimizer-a-perceptual-video-encoding-optimization-framework-e19f1e3a277f>
- [9] P. Le Callet and M. Barkowsky, “On viewing distance and visual quality assessment in the age of ultra high definition tv,” *VQEG eLetter*, vol. 1, no. 1, pp. 25–30, 2014.
- [10] T. Ni, D. A. Bowman, and J. Chen, “Increased display size and resolution improve task performance in information-rich virtual environments,” *Proc. Graph. Interface*, 2006, pp. 139–146.
- [11] S. Mayr, M. Köpper, and A. Buchner, “High pixel density on reading comprehension, proofreading performance, mood state, and physical discomfort,” *Displays*, vol. 48, pp. 41–49, 2017.
- [12] L. Lischke, S. Mayer, K. Wolf, A. Sahami, and N. Henze, “Subjective and objective effects of tablet’s pixel density,” in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, Apr. 2015, pp. 2769–2772.
- [13] W. Zou, J. Song, and F. Yang, “Perceived image quality on mobile phones with different screen resolution,” *Mobile Inf. Syst.*, vol. 2016, pp. 1–17, 2016.
- [14] A. Elmnsi, N. Osman, and I. Mkwawa, “Mobile devices pixel density and video quality,” in *Proc. 4th NAFOSTED Conf. Inf. Comput. Sci.*, Nov. 2017, pp. 325–330.
- [15] Y. Ou, Y. Xue, and Y. Wang, “Q-star: A perceptual video quality model considering impact of spatial, temporal, and amplitude resolutions,” *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2473–2486, Jun. 2014.
- [16] A. A. Laghari, H. He, A. Khan, and S. Karim, “Impact of video file format on quality of experience (QoE) of multimedia content,” *3D Research*, vol. 9, no. 3, 2018, Art. no. 191.
- [17] S. Jumisko-Pyykko, V. K. M. Vadakital, and M. M. Hannuksela, “Acceptance threshold: A bidimensional research method for user-oriented quality evaluation studies,” *Int. J. Digit. Multimedia Broadcast.*, vol. 2008, pp. 1–20, 2008.
- [18] S. Tavakoli, K. Brunnström, J. Gutiérrez, and N. García, “Quality of experience of adaptive video streaming: Investigation in service parameters and subjective quality assessment methodology,” *Signal Process.: Image Commun.*, vol. 39-B, pp. 432–443, Nov. 2015.
- [19] P. Pérez, J. Macías, J. J. Ruiz, and N. García, “Effect of packet loss in video quality of experience,” *Bell Labs Tech. J.*, vol. 16, no. 1, pp. 91–104, Jun. 2011.
- [20] R. S. Allison *et al.*, “Perspectives on the definition of visually lossless quality for mobile and large format displays,” *J. Electron. Imag.*, vol. 27, no. 5, pp. 1–23, Oct. 2018.
- [21] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, Rec. ITU-R BT.500-13, International Telecommunications Union, Geneva, Switzerland, Jan. 2012.
- [22] *Subjective Video Quality Assessment Methods for Multimedia Applications*, Rec. ITU-T P.910, International Telecommunications Union, Geneva, Switzerland, Apr. 2008.
- [23] *Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment*, Rec. ITU-T P.913, International Telecommunications Union, Geneva, Switzerland, Mar. 2016.
- [24] L. Janowski and M. Pinson, “The accuracy of subjects in a quality experiment: A theoretical subject model,” *IEEE Trans. Multimedia*, vol. 17, no. 12, pp. 2210–2224, Dec. 2015.
- [25] F. Li, F. Shuang, Z. Liu, and X. Qian, “A cost-constrained video quality satisfaction study on mobile devices,” *IEEE Trans. Multimedia*, vol. 20, no. 5, pp. 1154–1168, May 2018.
- [26] J. Wu, R. Tan, and M. Wang, “Streaming high-definition real-time video to mobile devices with partially reliable transfer,” *IEEE Trans. Mobile Comput.*, vol. 18, no. 2, pp. 458–472, Feb. 2019.