UVG Dataset: 50/120fps 4K Sequences for Video Codec Analysis and Development

Alexandre Mercat, Marko Viitanen, and Jarno Vanne
Ultra Video Group, Tampere University, Finland
{alexandre.mercat, marko.viitanen, jarno.vanne}@tuni.fi

ABSTRACT

This paper provides an overview of our open Ultra Video Group (UVG) dataset that is composed of 16 versatile 4K (3840×2160) test video sequences. These natural sequences were captured either at 50 or 120 frames per second (fps) and stored online in raw 8-bit and 10-bit 4:2:0 YUV formats. The dataset is published on our website (ultravideo.cs.tut.fi) under a non-commercial Creative Commons BY-NC license. In this paper, all UVG sequences are described in detail and characterized by their spatial and temporal perceptual information, rate-distortion behavior, and coding complexity with the latest HEVC/H.265 and VVC/H.266 reference video codecs. The proposed dataset is the first to provide complementary 4K sequences up to 120 fps and is therefore particularly valuable for cutting-edge multimedia applications. Our evaluations also show that it comprehensively complements the existing 4K test set in VVC standardization, so we recommend including it in subjective and objective quality assessments of next-generation VVC codecs.

CCS CONCEPTS

• Computing methodologies ~ Computer graphics ~ Image compression • Information systems ~ Data management systems ~ Database design and models

KEYWORDS

Open dataset, Raw video, Ultra High Definition (UHD), High Efficiency Video Coding (HEVC), Versatile Video Coding (VVC)

ACM Reference format:

Alexandre Mercat, Marko Viitanen, and Jarno Vanne. 2020. UVG Dataset: 50/120fps 4K Sequences for Video Codec Analysis and Development. In *Proceedings of ACM Multimedia Systems Conference (MMSys'20). ACM, Istanbul, Turkey, 6 pages.* https://doi.org/10.1145/3339825.3394937

MMSys'20, June 8–11, 2020, Istanbul, Turkey

ACM ISBN 978-1-4503-6845-2/20/06…$15.00
https://doi.org/10.1145/3339825.3394937

1 Introduction

Digital video has become ubiquitous in our everyday life thanks to a myriad of multimedia devices and services. Cisco [1] reports that 75% of total IP traffic was dedicated to video in 2017 and estimates it to grow to 82% by 2022. This explosive growth is mainly driven by a proliferation of numerous video applications and their ever-increasing quality expectations for immersive user experience.

In past decades, MPEG and ITU-T have addressed the needs of multimedia applications by announcing a series of video coding standards of which *High Efficiency Video Coding* (*HEVC/H.265*) [2] represents state-of-the-art. Currently, MPEG and ITU-T are also seeking coding gain beyond HEVC by developing a new standard called *Versatile Video Coding* (*VVC/H.266*) [3].

The main objective of digital video coding is to reduce the amount of bits needed to represent a video signal without deteriorating its quality. VVC and all prior international video coding standards are based on a classical hybrid video coding scheme that combines spatial/temporal (intra/inter) prediction, transform coding, quantization, and entropy coding.

Typically, video codecs are benchmarked with objective and subjective metrics to find out their *rate-distortion* (*RD*) performance, i.e., a trade-off between coding efficiency and loss of information. Conducting these quality assessments comprehensively calls for representative video datasets that address a broad range of video content (motion, texture, and illumination) and video parameters (resolution, frame rate, and bit depth).

In this paper, we present an open dataset made up of 16 raw 2160p (3840×2160) test video sequences. Eight of them were captured at 120 *frames per second* (*fps*) and the rest at 50 fps. The dataset is available online on our website at

http://ultravideo.cs.tut.fi/#testsequences

in 8-bit and 10-bit 4:2:0 YUV formats under a non-commercial Creative Commons BY-NC license. This work also characterizes all these sequences content-wise by their spatial and temporal perceptual information and performance-wise by their RD performance and coding complexity with the latest HEVC/H.265 and VVC/H.266 reference video codecs.

The rest of this paper is organized as follows. Section 2 gives an overview of the existing public video datasets. The proposed UVG dataset is presented in Section 3 and each UVG sequence is characterized in Section 4. Finally, Section 5 concludes the paper.

2 Existing Open Datasets

The existing open video datasets can be classified into 1) datasets made up of encoded video sequences; and 2) datasets containing original unprocessed video sequences in raw RGB or YUV format.

2.1 Encoded Datasets

These datasets are most commonly used in the *Quality of Experience* (*QoE*) measurements which are tolerant of quality degradations and visual artefacts introduced by dataset encoding. Therefore, typical applications of interest deal also with already encoded video, such as transmission and transcoding.

VideoSet [4] is composed of 220 5-second sequences, all in four resolutions: 1920×1080 (1080p), 1280×720 (720p), 960×540 (540p), and 640×360 (360p). Each of these 880 video clips were encoded using the AVC/H.264 codec [5] with *Quantization Parameter* (*QP*) values from 1 to 51. Authors provide the first three *just noticeable difference* (*JND*) points with 30+ subjects [6].

KoNViD-1k Video Database [7] contains 1200 panoramic 8-second 540p clips encoded with AVC/H.264 codec. The dataset was evaluated across blur, colorfulness, contrast, spatial information, temporal information, and video quality [8].

LIVE Netflix Video Quality of Experience database [9] consists of 112 distorted videos that were generated from fourteen 1080p videos at 24, 25, and 30 fps. The distortions are based on a set of eight different playout patterns including dynamic compression ratio of AVC/H.264 codec and rebuffering. This dataset was made for subjective quality assessment. Each video clip was assessed by 35 to 55 human subjects on a mobile device [10].

UGC dataset [11] includes 1500 20-second video clips sampled from millions of YouTube videos. They are split into 15 categories: animation, cover song, gaming, *high-dynamic-range* (*HDR*), how to, lecture, live music, lyric video, music video, news clip, sports, television clip, vertical video, vlog, and *virtual reality* (*VR*). Each video is available in YUV 4:2:0 format and in 360p, 480p, 720p, 1080p, and 2160p resolutions [12]. However, we classify this dataset into the first category, because it practically deals with encoded YouTube videos that were decoded and saved into YUV format. This approach introduces similar visual artifacts as found in the other encoded datasets.

2.2 Raw Datasets

This category is particularly essential in video codec evaluation and it is also better suited for many signal processing applications such as denoising. However, the acquisition of raw data requires professional camera equipment and extensive storage capacity because data integrity can only be preserved with lossless compression or without any compression.

Xiph.org Video Test Media [13] is a popular dataset for video compression. It contains 120 individual video clips of multiple resolutions (from 240p to 2160p), frame rates (from 25 fps to 60 fps), and content categories (e.g., movie and gaming) including nineteen 2160p sequences.

SJTU Media Lab proposed datasets for 4K [14] and HDR [15] video sequences. The 4K dataset includes fifteen 2160p sequences shot with Sony F65 camera at 30 fps. They are available in two formats: 10-bit YUV 4:4:4 and 8-bit YUV 4:2:0 [16]. The HDR dataset is made up of sixteen 2160p HDR sequences. The raw video data were recorded in Sony RAW 16-bit MAF format at 60 fps [17].

Joint Collaborative Team on Video Coding (*JCT-VC*) defined a dataset of 18 video sequences in the *Common Test Conditions* (*CTC*) [18] for HEVC standardization. The sequences were split into six classes (from A to F) according to spatial resolution and content type. The CTC dataset was updated by *the Joint Video Experts Team* (*JVET*) for VVC standardization [19]. Four sequences were removed, and nine new sequences were added including six 2160p sequences.

3 UVG Dataset Description and Characterization

The proposed UVG dataset contains 16 test video sequences. They were captured with Sony F65 video camera at either 50 or 120 fps in 16-bit F65RAW-HFR format and converted to 10-bit and 8-bit 4:2:0 YUV videos with FFmpeg [20]. Tables 1 and 2 tabulate these sequences in alphabetical order and depict their snapshots, essential parameters, content descriptions, specific features, and average RGB color histograms over all frames in each video.

Next, our dataset is first characterized with well-known *spatial perceptual information* (*SI*) and *temporal perceptual information* (*TI*) scores recommended by ITU [21]. Secondly, the RD characteristics of these sequences are evaluated with the latest releases of the HEVC/H.265 and VVC/H.266 reference encoders, i.e., the default 10-bit configurations of HM16.20 [22] and VTM8.0 [23], respectively. Finally, the sequences are assessed with the proposed complexity metric derived from [24]. In all these cases, the relevance of our UVG dataset is validated against the 2160p sequences of the VVC CTC dataset [19], which can be considered the most pertinent reference in current video coding research.

3.1 Spatial and Temporal Perceptual Information

SI indicates the maximum amount of spatial detail in a video. For a given N-frame sequence, SI is calculated by first filtering the luma samples (i,j) of each frame $F_n = F_0, F_1, \ldots, F_{N\text{-}1}$ with the Sobel filter, then computing the standard deviation over the Sobel filtered frame pixels, and finally taking the maximum of them [21] as

$$SI = \max_{0 \le n \le N-1} \{ \operatorname*{std}_{\substack{1 \le i \le W-2 \\ 1 \le j \le H-2}} \left[\sqrt{[Gh_n(i,j)]^2 + [Gv_n(i,j)]^2} \right] \}, \quad (1)$$

where W is the frame width and H is the frame height. $Gh_n(i,j)$ and $Gv_n(i,j)$ are the convolution results of the horizontal and vertical 3 × 3 Sobel edge detection kernel filters, respectively.

Correspondingly, TI stands for the maximum amount of temporal variation between successive frames $F_{n\text{-}1}$ and F_n [21] as

$$TI = \max_{1 \le n \le N-1} \{ \operatorname*{std}_{\substack{0 \le i \le W-1 \\ 0 \le j \le H-1}} \left[F_n(i,j) - F_{n-1}(i,j) \right] \}. \quad (2)$$

Table 1. Test sequence characteristics in the proposed UVG data set (sequences 1 - 8).

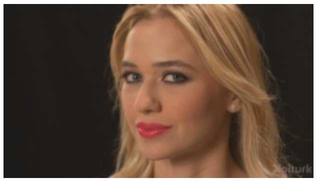
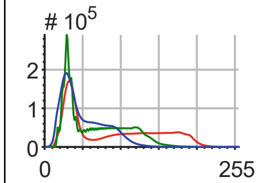
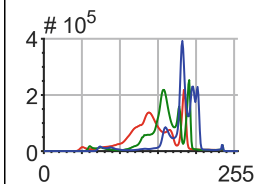
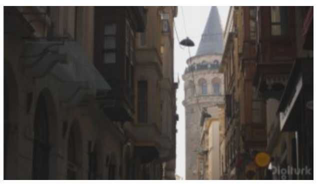
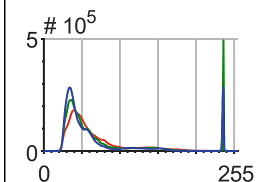
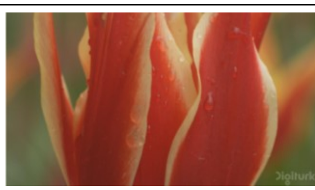
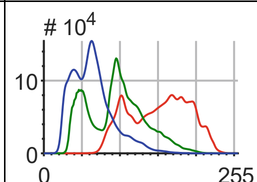
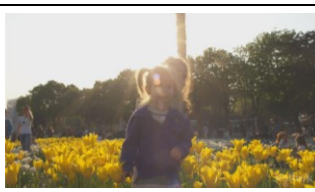
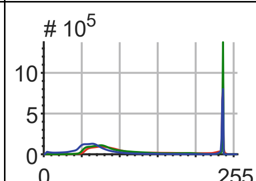
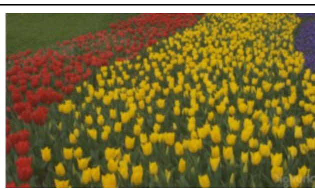
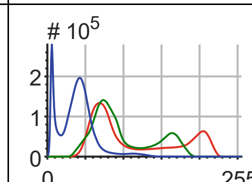
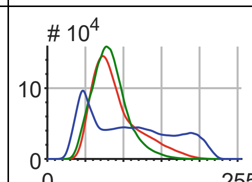
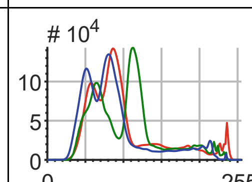
Snapshots	Characteristic and description	RGB color histogram
	**Name:** Beauty **Frame rate:** 120fps **Resolution:** 2160p **Frames:** 600 **Duration:** 5s **Bit depth:** 8 and 10 **Motion:** Slow / plenty of **Texture:** Complex **Contrast:** High **Description**: A close-up of female face with hair blowing in the wind. **Specific features:** High contrast between black background and female face; Noisy background.	
	**Name:** Bosphorus **Frame rate:** 120fps **Resolution:** 2160p **Frames:** 600 **Duration:** 5s **Bit depth:** 8 and 10 **Motion:** Slow / plenty of **Texture:** Complex **Contrast:** Low **Description**: A black boat cruising on the waterway. **Specific features:** Rippling water; foggy background composed of a hilly landscape, a seaside village, and a suspension bridge.	
	**Name:** CityAlley **Frame rate:** 50fps **Resolution:** 2160p **Frames:** 600 **Duration:** 12s **Bit depth:** 8 and 10 **Motion:** Slow / little **Texture:** Complex **Contrast:** High **Description**: A slow upward pan in a dark alley. **Specific features:** Detailed people on top of a tall tower at the end of the alley; high contrast between the alley and opening view.	
	**Name:** FlowerFocus **Frame rate:** 50fps **Resolution:** 2160p **Frames:** 600 **Duration:** 12s **Bit depth:** 8 and 10 **Motion:** Slow / little **Texture:** Smooth **Contrast:** Low **Description**: A close-up of a colorful red flower blowing in the wind. **Specific features:** Blurred green background; after panning up, the focus changes to flowers in the background.	
	**Name:** FlowerKids **Frame rate:** 50fps **Resolution:** 2160p **Frames:** 600 **Duration:** 12s **Bit depth:** 8 and 10 **Motion:** Slow / plenty of **Texture:** Complex **Contrast:** High **Description**: Two kids moving in a yellow flower field in a busy park. **Specific features:** Children close to the camera; camera shaking along the sequence; a lens flare caused by the sun in the background.	
	**Name:** FlowerPan **Frame rate:** 50fps **Resolution:** 2160p **Frames:** 600 **Duration:** 12s **Bit depth:** 8 and 10 **Motion:** Slow / little **Texture:** Complex **Contrast:** Low **Description**: A pan across a motionless colorful flower field. **Specific features:** Three bands of flowers: red, yellow, and indigo; flying insects on the flowers.	
	**Name:** HoneyBee **Frame rate:** 120fps **Resolution:** 2160p **Frames:** 600 **Duration:** 5s **Bit depth:** 8 and 10 **Motion:** Fast / little **Texture:** Complex **Contrast:** Low **Description**: A focus on a honeybee flying from flower to flower. **Specific features:** Fixed camera focused on the middle; blurred flowers in the foreground and the background.	
	**Name:** Jockey **Frame rate:** 120fps **Resolution:** 2160p **Frames:** 600 **Duration:** 5s **Bit depth:** 8 and 10 **Motion:** Fast / plenty of **Texture:** Smooth **Contrast:** High **Description**: A galloping racehorse with a jockey. **Specific features:** Camera tracks the horse causing fast moving background; the focus changes from the jockey to the legs of the horse.	

Table 2. Test sequence characteristics in the proposed UVG data set (sequences 9 - 16).

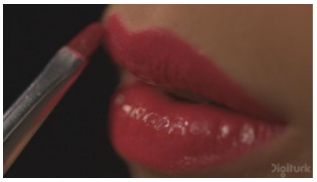
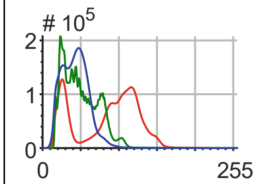
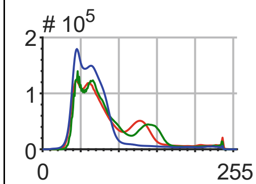
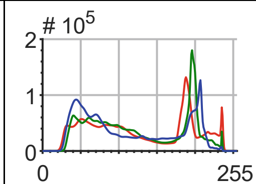
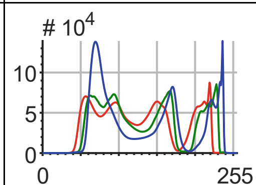
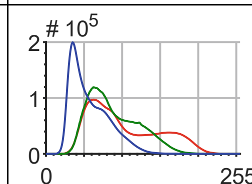
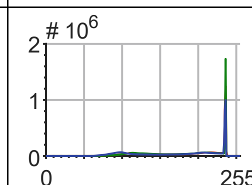
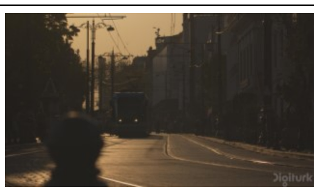
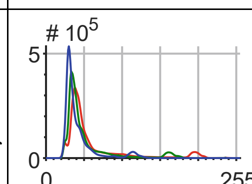
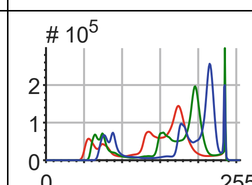
Snapshots	Characteristic and description	RGB color histogram
	**Name:** Lips **Frame rate:** 120fps **Resolution:** 2160p **Frames:** 600 **Duration:** 5s **Bit depth:** 8 and 10 **Motion:** Slow / little **Texture:** Smooth **Contrast:** High **Description**: A close-up of a female putting red lipstick with a brush. **Specific features:** Part of the face in the foreground is out of focus; high contrast between black background and female face.	
	**Name:** RaceNight **Frame rate:** 50fps **Resolution:** 2160p **Frames:** 600 **Duration:** 12s **Bit depth:** 8 and 10 **Motion:** Fast / plenty of **Texture:** Smooth **Contrast:** High **Description**: A group of horses and jockeys racing at night. **Specific features:** Camera tracks the horses in the first half of the sequence making the background move; static camera in the second half.	
	**Name:** ReadySetGo **Frame rate:** 120fps **Resolution:** 2160p **Frames:** 600 **Duration:** 5s **Bit depth:** 8 and 10 **Motion:** Fast / plenty of **Texture:** Complex **Contrast:** High **Description**: Horses go out of a starting gate at start of a horserace. **Specific features:** Camera tracks the horses creating occluded areas; sideways movement.	
	**Name:** RiverBank **Frame rate:** 50fps **Resolution:** 2160p **Frames:** 600 **Duration:** 12s **Bit depth:** 8 and 10 **Motion:** Slow / little **Texture:** Complex **Contrast:** High **Description**: A slow pan across riverbank. **Specific features:** Rippling water sparkling with the sun; background composed of detailed trees and blue sky.	
	**Name:** ShakeNDry **Frame rate:** 120fps **Resolution:** 2160p **Frames:** 300 **Duration:** 2.5s **Bit depth:** 8 and 10 **Motion:** Slow / plenty of **Texture:** Complex **Contrast:** Low **Description**: A dog shaking itself dry in a forest. **Specific features:** Water drops flying around; the dog leaves the static scene at the end of the sequence.	
	**Name:** SunBath **Frame rate:** 50fps **Resolution:** 2160p **Frames:** 300 **Duration:** 6s **Bit depth:** 8 and 10 **Motion:** Slow / plenty of **Texture:** Smooth **Contrast:** High **Description**: Tree branches swinging in front of the sun. **Specific features:** Part of the branches coming in and out of the scene; background overexposed from direct sunlight.	
	**Name:** Twilight **Frame rate:** 50fps **Resolution:** 2160p **Frames:** 600 **Duration:** 12s **Bit depth:** 8 and 10 **Motion:** Slow / plenty of **Texture:** Smooth **Contrast:** High **Description**: Arrival of a tram at twilight. **Specific features:** Dark scene; a man is crossing the scene at the end of the sequence.	
	**Name:** YachtRide **Frame rate:** 120fps **Resolution:** 2160p **Frames:** 600 **Duration:** 5s **Bit depth:** 8 and 10 **Motion:** Fast / plenty of **Texture:** Complex **Contrast:** High **Description**: A close-up of a boat cruising on the waterway. **Specific features:** Camera focused on foam and water sparkling with the sun; the boat leaves the scene at the end of the sequence.	

Table 3. Rate-Distortion characterization of the UVG and CTC datasets.

Index	Sequence	VTM8.0 – QP 32				HM16.20 vs. VTM8.0		
		Kbps	$PSNR_Y$	$PSNR_U$	$PSNR_V$	$BD\text{-}BR_Y$	$BD\text{-}BR_U$	$BD\text{-}BR_V$
U1	**Beauty**	7337	34.85	37.08	39.46	-22.99%	-34.39%	-34.13%
U2	**Bosphorus**	5323	41.02	47.62	46.55	-25.70%	-65.02%	-55.12%
U3	**CityAlley**	1050	40.49	48.10	47.18	-30.91%	-56.13%	-58.07%
U4	**FlowerFocus**	635	39.86	44.26	44.33	-50.15%	-55.42%	-61.45%
U5	**FlowerKids**	3729	39.98	43.77	45.27	-32.34%	-39.08%	-45.56%
U6	**FlowerPan**	2778	38.21	40.89	41.91	-38.92%	-38.44%	-40.05%
U7	**HoneyBee**	2756	38.90	42.70	42.51	-45.49%	-37.39%	-38.43%
U8	**Jockey**	4432	39.24	43.74	43.15	-46.58%	-51.95%	-50.70%
U9	**Lips**	3662	35.06	37.62	40.37	-46.60%	-74.29%	-53.50%
U10	**RaceNight**	4621	36.27	43.85	43.49	-30.81%	-46.22%	-38.20%
U11	**ReadySetGo**	11449	40.14	44.49	44.14	-32.12%	-41.82%	-40.90%
U12	**RiverBank**	7860	36.14	45.41	45.09	-23.55%	-69.34%	-71.26%
U13	**ShakeNDry**	13852	37.71	41.97	42.21	-24.12%	-34.61%	-41.98%
U14	**SunBath**	4065	42.83	46.85	48.87	-32.35%	-40.34%	-44.76%
U15	**Twilight**	1208	40.52	48.28	47.60	-31.94%	-62.63%	-58.28%
U16	**YachtRide**	19105	38.54	45.60	44.44	-27.18%	-63.71%	-60.65%
C1	Tango2	3662	38.86	47.58	45.07	-37.37 %	-54.83 %	-51.06 %
C2	FoodMarket4	3968	41.11	45.78	46.12	-36.28 %	-36.66 %	-38.50 %
C3	Campfire	7028	36.51	36.10	39.70	-38.34 %	-29.29 %	-48.19 %
C4	CatRobot	4186	38.34	40.8	41.25	-44.09 %	-52.79 %	-46.38 %
C5	DaylightRoad2	4151	36.41	43.97	41.38	-43.13 %	-49.11 %	-39.88 %
C6	ParkRunning3	19177	36.52	32.85	34.37	-37.29 %	-15.87 %	-14.33 %

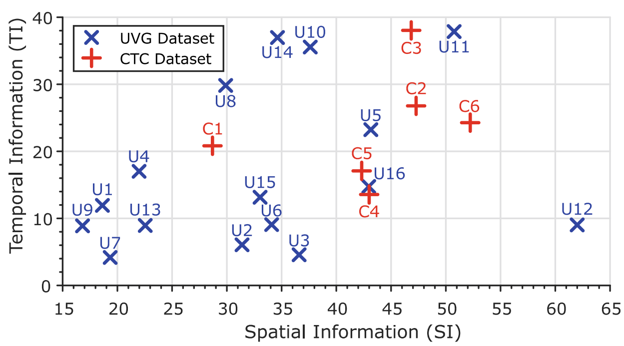


Figure 1: SI and TI for the UVG and CTC video sequences.

Figure 1 plots the SI and TI scores for the proposed UVG video sequences (U1-U16) in blue and the 2160p CTC sequences (C1-C6) in red. The indexes are linked to associated sequences in Table 3. The results illustrate that the proposed sequences are distributed across a wide range of SI and TI scores. This diversity makes the UVG dataset highly complementary to that of the CTC.

3.2 Rate-Distortion Characterization

Table 3 presents the RD characterization results of the UVG and CTC 2160p sequences with VTM8.0 and HM16.20. The four columns in the middle report the bitrate in *kilobits per second* (*Kbps*) and distortion in *peak signal-to-noise ratios* (*PSNRs*) for all sequences when encoding them with VTM8.0 and the QP value of 32. PSNR results (in dBs) are separately given for the luma (*Y*) and two chroma (*U* and *V*) components. The used coding configuration is *Random Access* (*RA*) specified in the CTC [18], [19].

Encoding the UVG dataset with VTM8.0 produced bit rates from 635 to 19105 Kbps and luma PSNRs from 34.85 to 42.83 dB for a QP value of 32. The respective metrics with the CTC dataset were from 3662 to 19177 Kbps and from 36.41 to 41.11 dB. The increased variances in rate and distortion attest to the content diversity of the UVG dataset in VVC encoding.

Furthermore, the rightmost three columns of Table 3 report the coding gain of VTM8.0 over HM16.20 for the same objective visual quality measured in PSNR. In practice, the bitrate gains were computed with a popular *Bjøntegaard Delta Bit Rate* (*BD-BR*) [25] metric under the RA coding configuration. BD-BR gives the difference in percent between the RD curves of VTM8.0 and HM16.20, which were interpolated through four RD points that correspond to the QP values of 22, 27, 32, and 37.

On average, VTM8.0 saves -39.42% (from -36.28% to -44.09%) of $BD\text{-}BR_Y$ over that of HM16.20 with the CTC dataset and the respective average is -33.86% (from -22.99% to -50.15%) with the UVG dataset. These results show that our dataset also provides edge cases (U1 and U4) beyond those of the CTC (C2 and C4) for coding efficiency comparison.

3.3 Spatial and Temporal Encoding Complexity

Our characterization is complemented by the *spatial encoding complexity* (S_{cplx}) and *temporal encoding complexity* (T_{cplx}) calculations based on [24]. They were yielded by encoding the sequences with HM16.20 using the QP values of 22, 27, 32, and 37 under the *All Intra* (*AI*) and RA configurations [18], respectively.

For a given sequence, S_{cplx} is calculated by averaging the bit counts over the four QP values in the AI case and normalizing the result (AI_{size}) with the bit count of the uncompressed 10-bit 4:2:0 YUV sequence (U_{size}) as

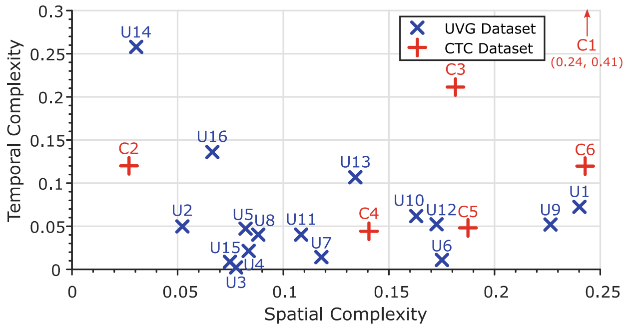


Figure 2: Complexity scores of the UVG and CTC sequences.

$$S_{cplx} = \frac{AI_{size}}{U_{size}}. \quad (3)$$

Correspondingly, T_{cplx} is obtained by first averaging the bit counts over the four QP values in the RA case and then computing the ratio between average bit counts of bi-predicted inter frames (B_{size}) and intra frames (I_{size}) as

$$T_{cplx} = \frac{\text{avg}(B_{size})}{\text{avg}(I_{size})}. \quad (4)$$

The intra period was set to 32, so every 32th frame was intra frame and the others were bi-predicted inter frames. The normalization by I_{size} value takes into account the correlation between B_{size} and I_{size} values. High I_{size} values tend to increase B_{size} values because the high spatial complexity increases residuals in inter prediction.

Figure 2 plots the S_{cplx} and T_{cplx} scores for the UVG and CTC 2160p sequences (see Table 3) in blue and red, respectively. Again, the UVG sequences are spread across the entire (S_{cplx},T_{cplx}) plane of encoding complexities. These results are well in line with our previous findings on the usefulness of the UVG dataset.

4 Conclusion

This paper presented our open UVG dataset that is composed of 16 versatile 4K (2160p) natural 50/120 fps video sequences in 8-bit and 10-bit 4:2:0 YUV formats. This online dataset is available on our website (ultravideo.cs.tut.fi). To the best of our knowledge, this is the first dataset that includes 4K 120fps raw videos, which make it particularly relevant for future video codec development, where the inevitable trend is towards higher resolutions and frame rates. Our dataset is also validated to complement the existing CTC 4K test set in terms of VVC and HEVC coding characteristics and complexity. Therefore, we believe it would diversify objective and subjective quality assessment in VVC standardization and related application development.

ACKNOWLEDGMENTS

This work was supported in part by the Academy of Finland (decision no. 301820). The authors would also like to thank Digiturk for providing video content for the proposed dataset.

REFERENCES

[1] Cisco, Cisco Visual Networking Index: Forecast and Trends 2017-2022, Dec. 2018.

[2] 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, Dec. 2012, pp. 1649-1668.

[3] B. Bross, J. Chen, S. Liu, and Y.-K. Wang, "Versatile Video Coding (Draft 8)," *Document JVET-Q2001*, Brussels, Belgium, Jan. 2020.

[4] H. Wang *et al.*, "VideoSet," *IEEE Dataport*, [Online]. Available: http://dx.doi.org/10.21227/H2H01C. Accessed: Apr. 03, 2020.

[5] T. Wiegand, G. J. 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, July 2003, pp. 560-576.

[6] H. Wang *et al.*, "VideoSet: A large-scale compressed video quality dataset based on JND measurement," *J. Vis. Commun. Image R.*, vol. 46, Jul. 2017, pp. 292-302.

[7] V. Hosu *et al.*, "The Konstanz Natural Video Database," [Online]. Available: http://database.mmsp-kn.de. Accessed: Apr. 03, 2020.

[8] V. Hosu *et al.*, "The Konstanz natural video database (KoNViD-1k)," *in Proc. Int. Conf. Quality of Multimedia Experience*, Erfurt, Germany, May-June 2017.

[9] C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik, "LIVE Netflix Video Quality of Experience Database," [Online]. Available: http://live.ece.utexas.edu/research/LIVE_NFLXStudy/index.html. Accessed: Apr. 03, 2020.

[10] C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik, "Study of temporal effects on subjective video quality of experience," *IEEE Trans. Image Process.*, vol. 26, no. 11, Nov. 2017, pp. 5217-5231.

[11] Y. Wang, S. Inguva, and A. Balu. "UGC Dataset," [Online]. Available: https://media.withyoutube.com/. Accessed: Apr. 03, 2020.

[12] Y. Wang, S. Inguva, and B. Adsumilli, "YouTube UGC dataset for video compression research," *in Proc. IEEE Int. Workshop Multimedia Signal Processing*, Kuala Lumpur, Malaysia, Sep. 2019.

[13] Xiph.org, "Video test media," [Online]. Available: https://media.xiph.org/video/derf/. Accessed: Apr. 03, 2020.

[14] SJTU Media Lap, "SJTU 4K Video Sequences," [Online]. Available: http://medialab.sjtu.edu.cn/web4k/index.html. Accessed: Apr. 03, 2020.

[15] SJTU Media Lap, "SJTU HDR Video Sequences," [Online]. Available: http://medialab.sjtu.edu.cn/HDR/index.html. Accessed: Apr. 03, 2020.

[16] L. Song, X. Tang, W. Zhang, X. Yang, and P. Xia, "The SJTU 4K video sequence dataset," *in Proc. Int. Workshop Quality of Multimedia Experience*, Klagenfurt am Worthersee, Austria, Jul. 2013.

[17] L. Song, Y. Liu, X. Yang, R. Xie, G. Zhai, and W. Zhang, "The SJTU HDR video sequence dataset," *in Proc. Int. Conf. Quality of Multimedia Experience*, Lisboa, Portugal, June 2016.

[18] F. Bossen, "Common test conditions and software reference configurations," *Document JCTVC-L1100*, Geneva, Switzerland, Jan. 2013.

[19] F. Bossen, J. Boyce, X. Li, V. Seregin, and K. Sühring, "Common test conditions and software reference configurations," *Document JVET-N1010,* Geneva, Switzerland, Oct. 2019.

[20] FFmpeg. [Online]. Available: https://www.ffmpeg.org/. Accessed: Apr. 03, 2020.

[21] *Subjective video quality assessment methods for multimedia applications*, document ITU-T Rec. P.910, Apr. 2008.

[22] JCT-VC, "HEVC reference software," [Online]. Available: https://hevc.hhi.fraunhofer.de/. Accessed: Apr. 03, 2020.

[23] JVET, "VVC reference software," [Online]. Available: https://jvet.hhi.fraunhofer.de/. Accessed: Apr. 03, 2020.

[24] C. Chen, S. Inguva, A. Rankin, and A. Kokaram, "A subjective study for the design of multi-resolution ABR video streams with the VP9 codec," *SPIE Electronic Imaging,* vol. 2016, no. 2, p. 1-5.

[25] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *Document VCEG-M33*, Austin, TX, USA, Apr. 2001.