Assessing video content coding complexity for live video streaming

Maria Martini and Shaymaa Al-Juboori
Kingston University London

VQEG meeting Dec 2022

Kingston University London

Upcoming workshop

IEEE ICASSP Workshop on

Humans, Machines and Multimedia - Quality of Experience and Beyond (HMM-QOE 2023)

Paper Submission Deadline: 24 February 2023

Motivation and Aim

- Compressing video sequences with different content complexity results in different compression bitrates for the same quality level or in different quality levels for the same bitrate
- To address this, per-title optimization is used recently (e.g., by Netflix) to generate appropriate rate-quality representations for different Video on demand (VoD) content to be streamed via adaptive video streaming.
- However, this cannot be adopted for live video streaming as it requires encoding (multiple times) each video content.
- Spatial Information (SI) and Temporal Information (TI) have been often used as an indicator of video complexity, for instance for preparing and describing content for video quality assessment tests, and for rate-distortion modelling.
- It has been questioned recently if different metrics could lead to a better estimation of ”compressibility” of video.

- We compare existing and proposed metrics in terms of their ability to estimate ”compressibility”.

Background

- Definition of content metrics for the estimation of bit-rate for event-based sensors

[N. Khan and M. G. Martini, “Bandwidth modelling of silicon retinas for next generation visual sensor networks,” *Sensors*, vol. 19, no. 8, p. 1751, 2019.

N. Khan and M. G. Martini, “Data rate estimation based on scene complexity for dynamic vision sensors on unmanned vehicles,” in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC),* Bologna, Italy, September 2018.]

– Analysis of SI and TI variation with compression, resolution, bit depth

[N. Barman, N. Khan, and M. G. Martini, “Analysis of spatial and temporal information variation for 10-bit and 8-bit video sequences,” in 2019 *IEEE 24th International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD)*, 2019.]

- Bit-rate model based on SI and TI

[Haseeb, A., Martini, M. G., Cicalo, S., & Tralli, V. (2012, June). Rate and distortion modelling for real-time MGS coding and adaptation. In *2012 Wireless Advanced (WiAd)* (pp. 85-89), June 2012.]

– Recent observations on image statistics (see tomorrow’s presentation)

Observations

1) Other filter types and pooling methods worked better than SI for other type of data (from event-based sensors). Can this apply also for compressed video?
2) The variance of pixel values in luminance frames (an index of contrast) appeared to be a key content element in another study (see tomorrow’s presentation) hence this is expected to correlate well with video coding complexity.

Performance evaluation method

Dataset 1

– BVI-HD

[F. Zhang, F. M. Moss, R. Baddeley, and D. R. Bull, “BVI-HD: A video quality database for HEVC compressed and texture synthesized content,” IEEE Transactions on Multimedia, vol. 20, no. 10, pp. 2620–2630, 2018.]

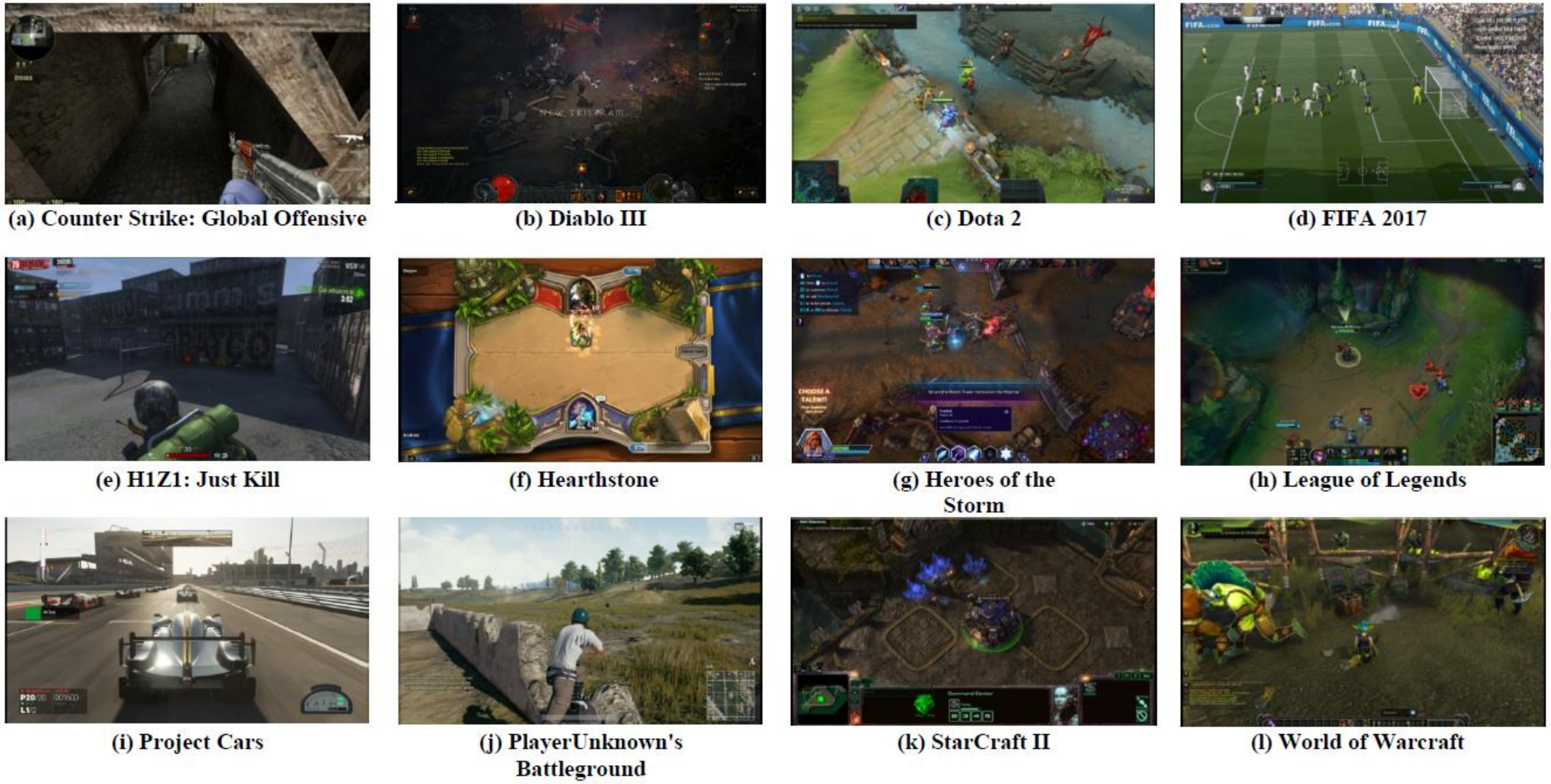
- 32 HD SRCs and 384 PVSs (compressed via HEVC), as well as subjective assessments.
- HRC: Six QP values for HEVC compression (from 22 to 47 with an interval of 5).
- Frames rates of 60 fps, 30fps, 50 fps (we considered the subset at 60 fps), duration truncated to 5s.
- We did two studies with this dataset: one considering the HEVC video sequences at 60fps (apart from a few where the scores reported were inconsistent), one only considering the sequences with no scene cuts

Dataset 2

– GamingVideoSet

[N. Barman, S. Zadtootaghaj, S. Schmidt, M. G. Martini, and S. Moeller, “GamingVideoSET: a dataset for gaming video streaming applications,” in 2018 16th Annual Workshop on Network and Systems Support for Games (NetGames)].

- 24 SRCs (gaming videos), 30 seconds (we cut them to 3s), 1080p, 30fps frames per second.
- Subjective quality evaluation scores for 90 PVSs (H.264/MPEG-AVC coding at three resolutions and five bitrates each).
- Objective scores for 576 PVSs (24 different resolution-bitrate pairs)



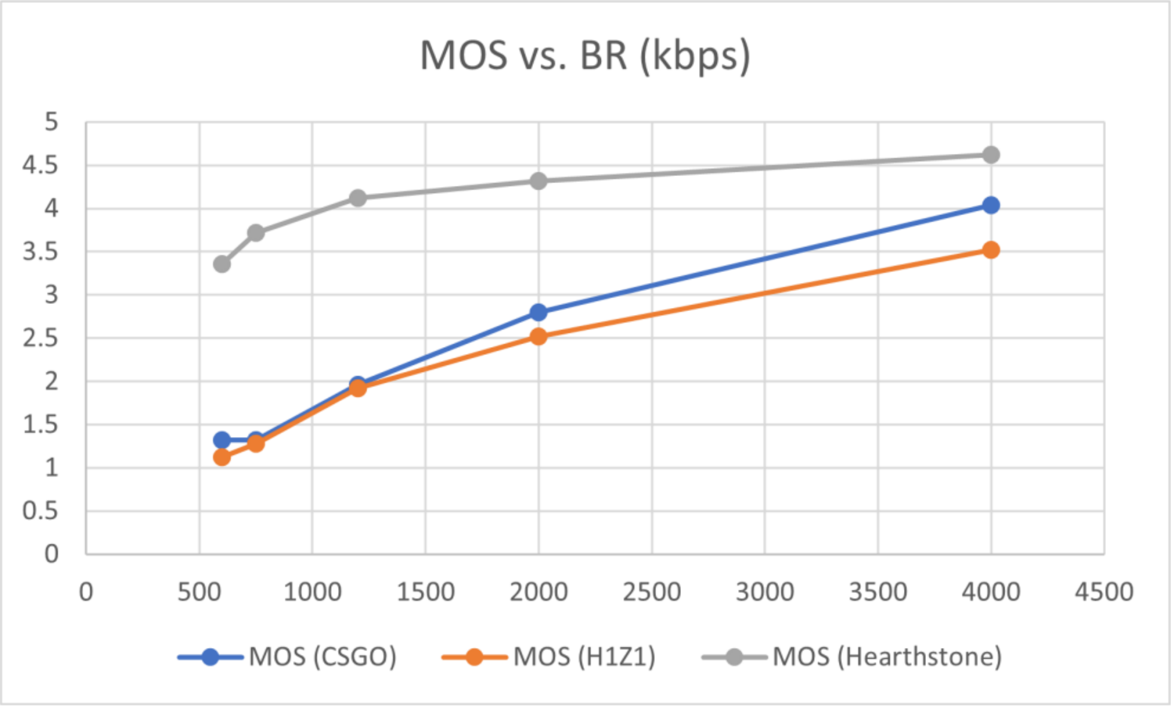
(a) Counter Strike: Global Offensive (b) Diablo III (c) Dota 2 (d) FIFA 2017
(e) H1Z1: Just Kill (f) Hearthstone (g) Heroes of the Storm (h) League of Legends
(i) Project Cars (j) PlayerUnknown's Battleground (k) StarCraft II (l) World of Warcraft

Performance indicators

– Video quality measures: SSIM, VMAF, MOS

– Area under the RD (Quality-Rate) curve K-SSIM K-VMAF K-MOS

– Correlation with coding complexity indices (PLCC, SROCC, KROCC)



Coding complexity metrics compared

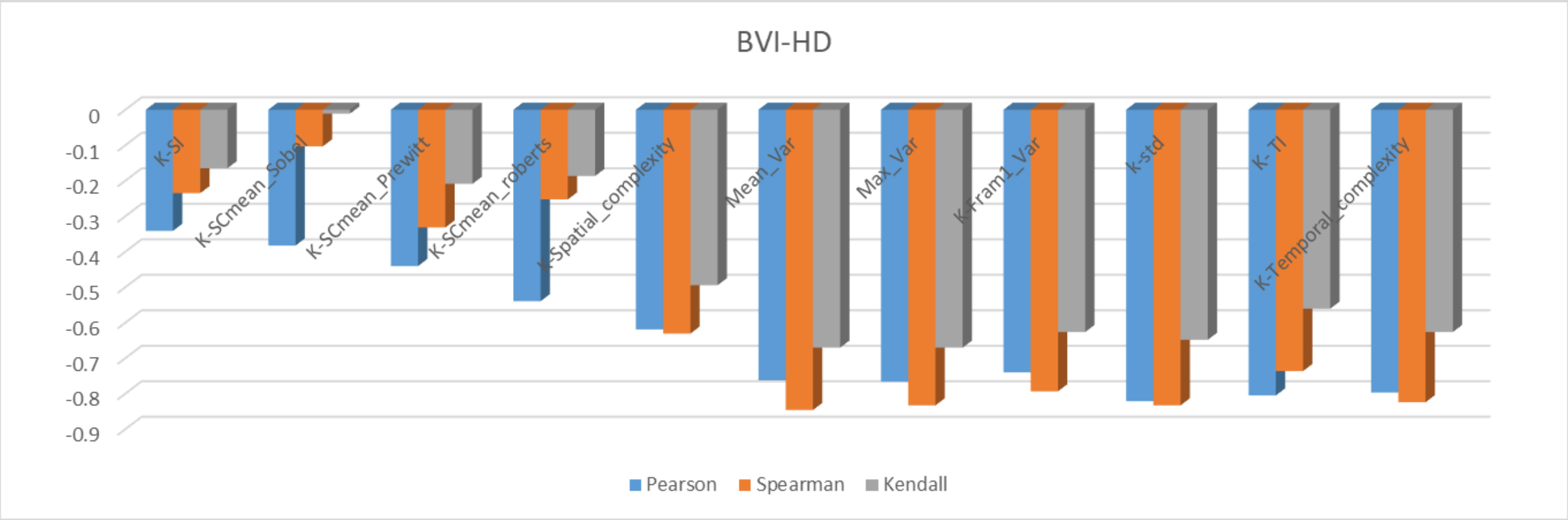
Metric	Definition
SI	$max_{time}\left\{SI_{std}\right\}$
$\mathrm{SC}_{meanSobel}$	$mean_{time}\{\frac{1}{M\cdot N}\sum_{j=1}^{N}\sum_{i=1}^{M}G_{i,j\ Sobel}\}$
$\mathrm{SC}_{meanPrewitt}$	$mean_{time}\{\frac{1}{M\cdot N}\sum_{j=1}^{N}\sum_{i=1}^{M}G_{i,j\ Prewitt}\}$
$\mathrm{SC}_{meanRoberts}$	$mean_{time}\{\frac{1}{M\cdot N}\sum_{j=1}^{N}\sum_{i=1}^{M}G_{i,j\ Roberts}\}$
E [20]	(8)
σ^2_{mean}	$mean_{time}\{\sigma^2_Y\}$
σ^2_{max}	$max_{time}\{\sigma^2_Y\}$
σ^2_{min}	$min_{time}\{\sigma^2_Y\}$
σ^2_1	$\sigma^2_{Y_1}$
σ^2_{std}	$std_{time}\{\sigma^2_Y\}$
TI	$max_{time}\left\{std[M_p^n]\right\}$
H [20]	(9)

[20] V. V. Menon, C. Feldmann, H. Amirpour, M. Ghanbari, and C. Timmerer, “VCA: Video complexity analyzer,” 2022. https://github.com/cd-athena/VCA

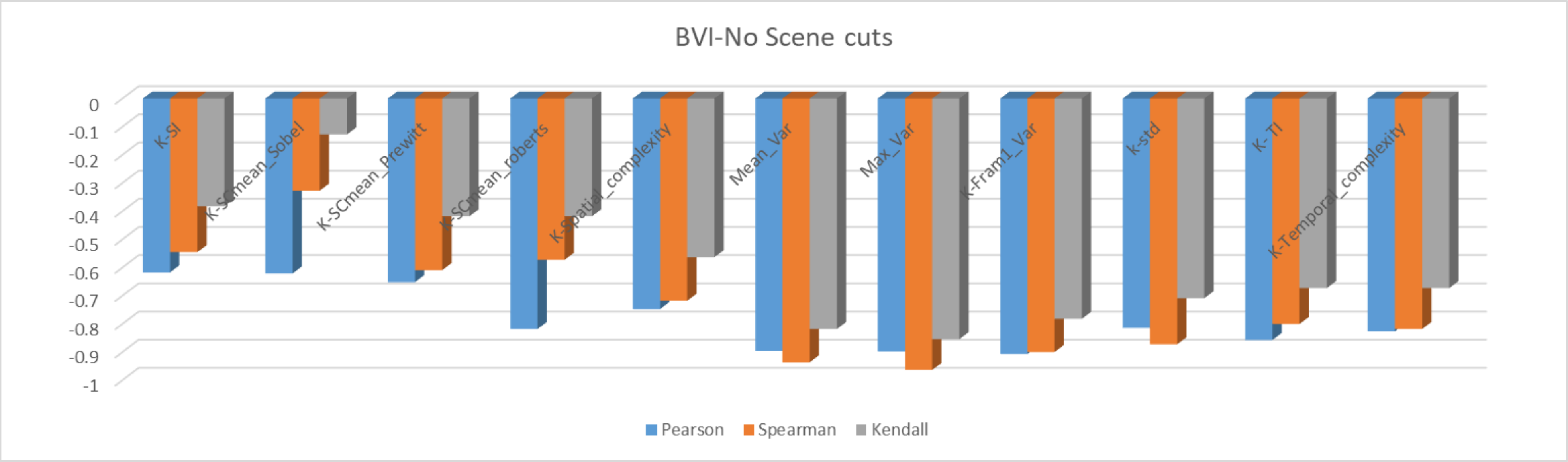
– Note: in addition, some tests have been done with MV related metrics – these did not overperform TI and H for the considered datasets hence results are not reported.

Results

BVI-HD dataset

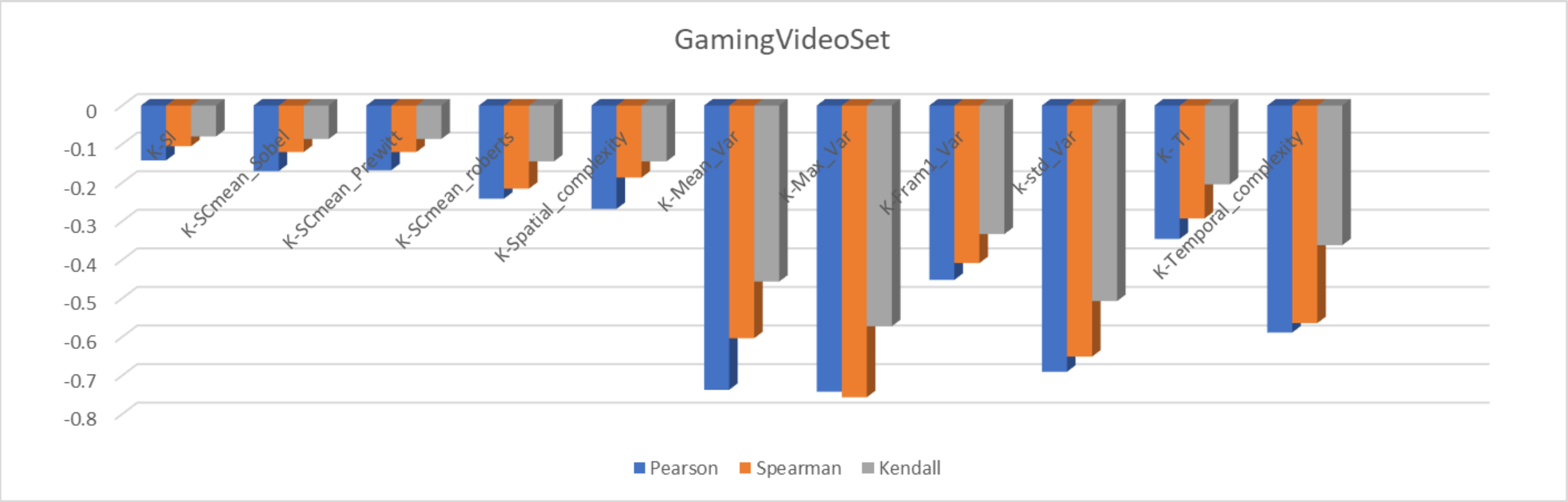


BVI-HD dataset no scene cuts



Kingston University London

GamingVideoSet



Towards a single coding complexity metric?

Possible combinations of TI and mean variance tested on BVI-HD dataset

(first one similar as for SI and TI in [Haseeb, A., Martini, M. G., Cicalo, S., & Tralli, V. (2012, June) Rate and distortion modelling for real-time MGS coding and adaptation. In *2012 Wireless Advanced (WiAd)* (pp. 85-89), June 2012.]

$$C = a + bx + cy \tag{10}$$

resulted in R-square 0.88;

$$C = a + bxy \tag{11}$$

resulted in R-square 0.87;

$$C = a + bx + cy^2 \tag{12}$$

resulted in R-square 0.91;

$$C = a + logx + cy \tag{13}$$

resulted in R-square 0.9.

Conclusion

- The variance of pixel values correlate well with coding complexity, with any pooling method considered
- Among temporal complexity metrics, TI and H perform best among the tested metrics.
- Combination of metrics are possible – combination of a metric representing spatial complexity and a metric representing temporal complexity is recommended.
- Tests with more datasets including different genres and more data are recommended and foreseen.

Link to associated paper