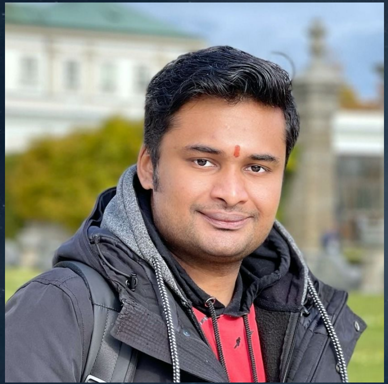
Video complexity analyzer (VCA) for streaming applications

December 14, 2021

Presenters: **Vignesh V Menon** & Hadi Amirpour

About Us



Vignesh V Menon

Hadi Amirpour

Researchers @ Christian Doppler Laboratory ATHENA, University of Klagenfurt, Austria

- Motivation for VCA
- Features
- Experimental Results
- Applications
- Future Roadmap

Motivation

Motivation

- We aim to develop online prediction systems tailor-made for live streaming applications.
- The state-of-the-art spatial and temporal complexity feature is SI-TI.

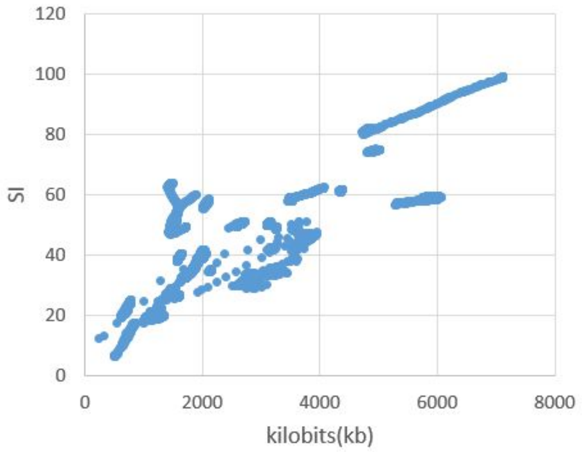


Fig.1: Correlation of SI feature with number of bits (in kb) per frame in IDR encoding with QP27 of x265 for 24 test sequences from MCML[1] and SJTU[2] dataset.

Pearson correlation coefficient (PCC) of SI with bits per frame is ~0.79.

[1] M. Cheon and J.-S. Lee, “Subjective and Objective Quality Assessment of Compressed 4K UHD Videos for Immersive Experience,”IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 7, pp. 1467–1480, 2018.
[2] L. Song, X. Tang, W. Zhang, X. Yang, and P. Xia, “The SJTU 4K Video Sequence Dataset,”Fifth International Workshop on Quality of Multimedia Experience (QoMEX2013), Jul. 2013.

Motivation

- Time taken to compute SI-TI features is very high!
  - ➢ ~0.05 seconds per frame for 1080p, ~0.2 seconds per frame for 2160p
  - ➢ Not suitable for live applications
  - ➢ Higher computational cost in VoD applications

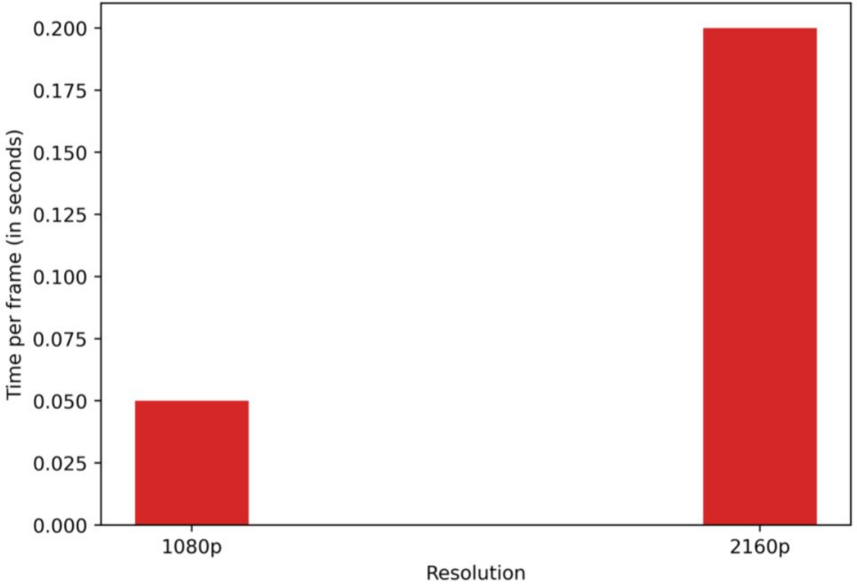


Fig.2: Time taken to compute SI-TI features [1] in Intel Xeon Gold 5218R

[1] SITI source code: https://github.com/Telecommunication-Telemedia-Assessment/SITI

Motivation

Video Complexity Analyzer (VCA) can be realized as a fast preprocessor which determines the spatial and temporal complexity of videos (segments) to aid the encoding process.

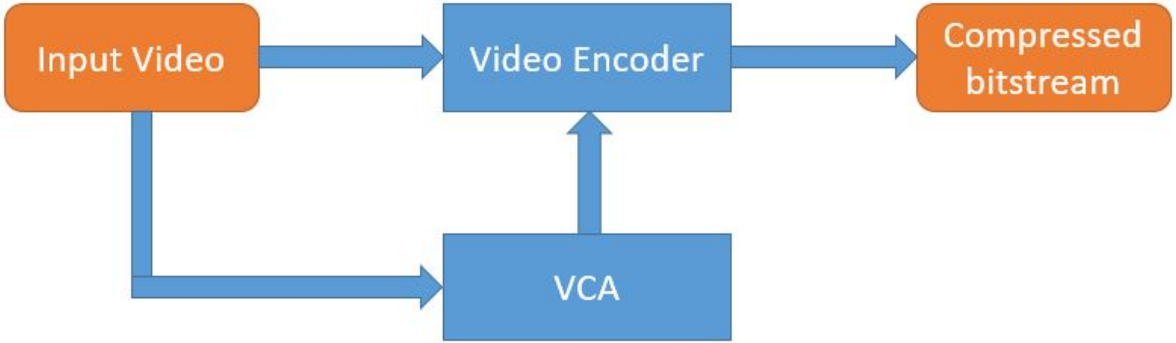


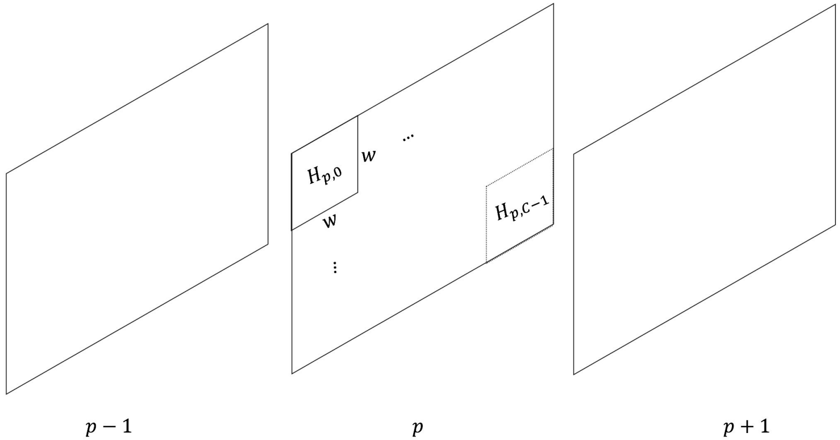
Fig.3: The proposed framework for streaming applications.

Features

Spatial complexity feature

$$H_{p,k} = \sum_{i=0}^{w-1} \sum_{j=0}^{w-1} e^{\left|\left(\frac{ij}{w^2}\right)^2 - 1\right|} |DCT(i,j)|$$

k is the block address in the pth frame,w×w pixels is the size of the block, and DCT(i, j) is the $(i,j)^{th}$ DCT component when i+j >1, and 0 otherwise.



Spatial complexity feature

$$E = \sum_{k=0}^{C-1} \frac{H_{p,k}}{C \cdot w^2}$$

C represents the number of blocks per frame.

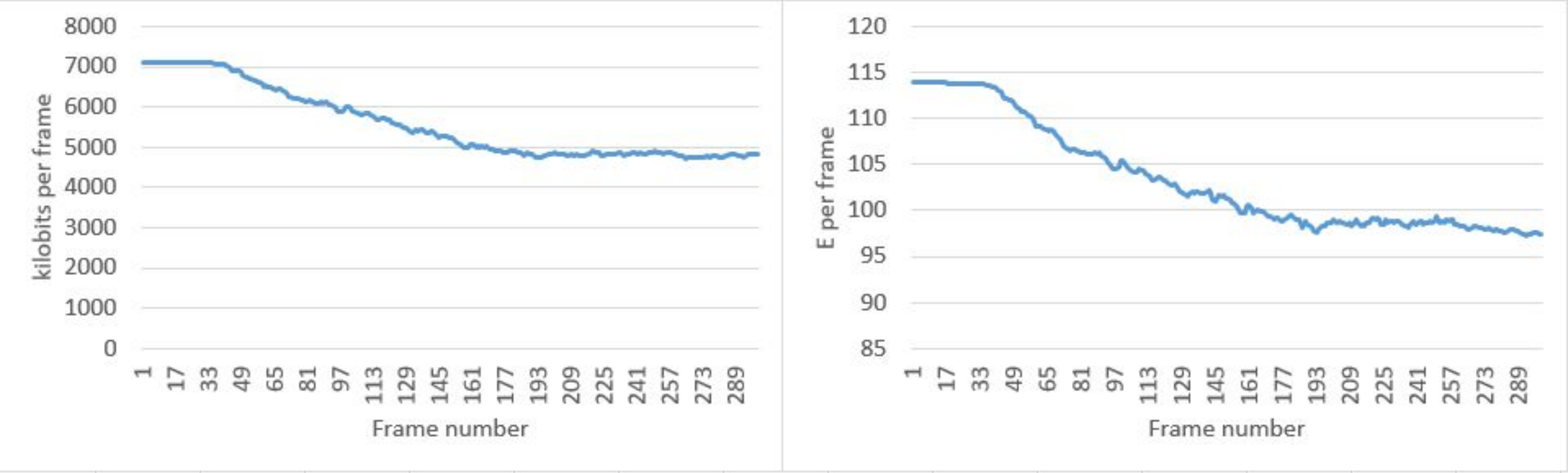
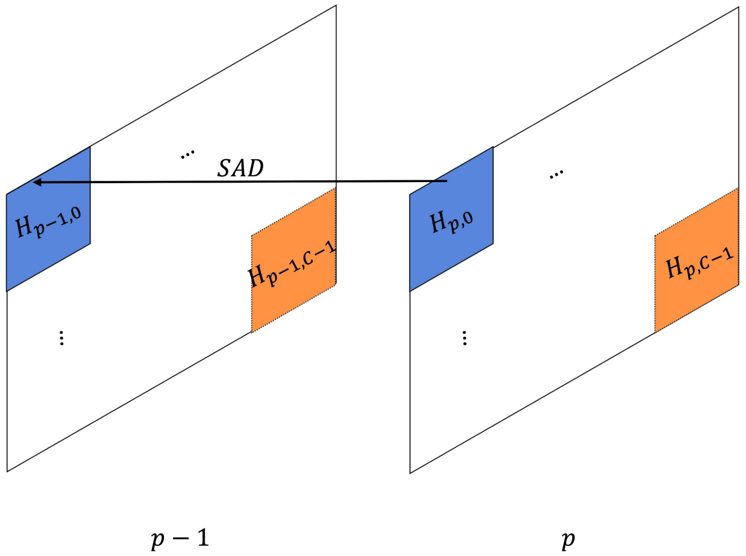


Fig. 4: Number of bits (in kb) per frame and E feature of Wood SJTU sequence.

Temporal complexity feature

The block-wise SAD of the texture energy of each frame (p) compared to its previous frame (p-1) is computed.



$$h = \sum_{k=0}^{C-1} \frac{SAD(H_{p,k}, H_{p-1,k})}{C}$$

Experimental Results

Results

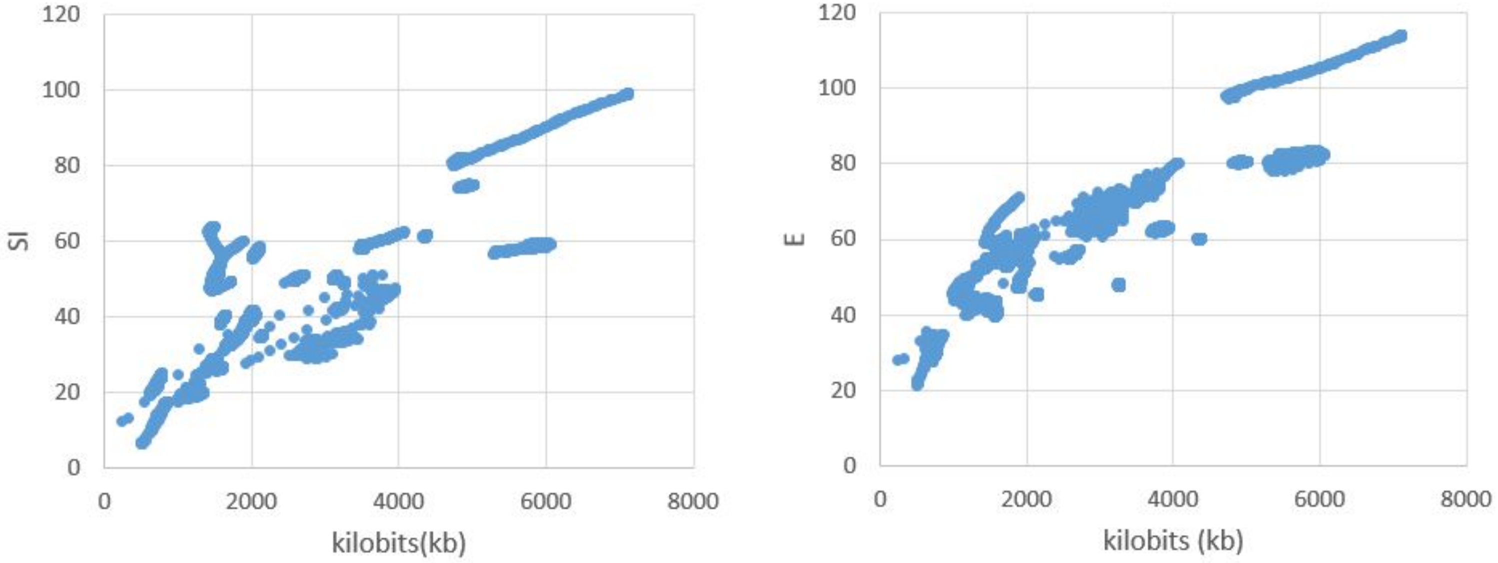


Fig. 5: Correlation of SI and E features with number of bits (in kb) per frame.

PCC(SI, Bits per frame) = 0.787
PCC(E, Bits per frame) = 0.856

Results

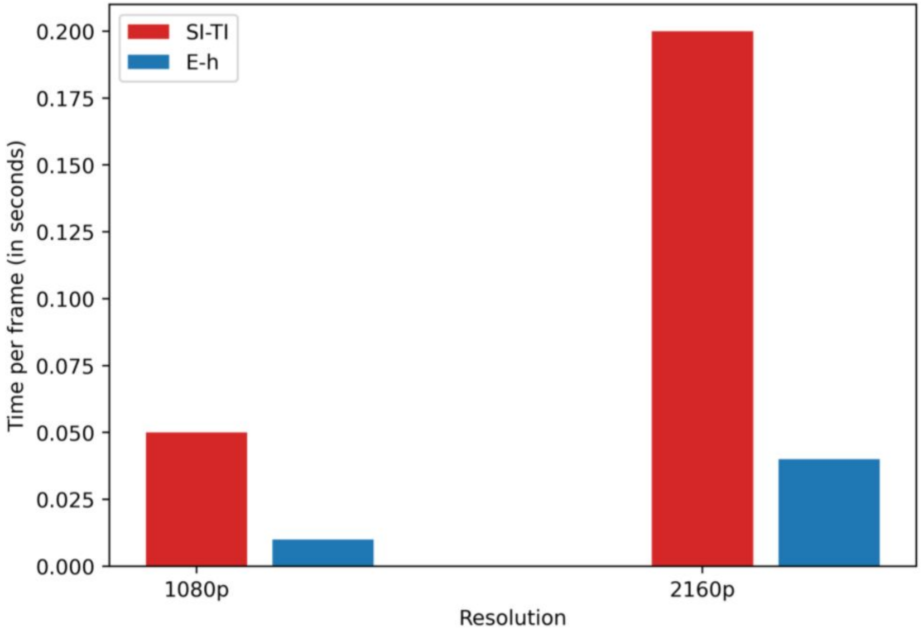


Fig. 6: Average time to compute E-h for various resolutions (with x86 SIMD).

Note: Presently, E-h computation is 5 times faster than SI-TI computation.

Applications

Shot Detection

We define the gradient of ‘h’ per frame ‘p’ as:

$$\epsilon_p = \frac{h_{p-1} - h_p}{h_{p-1}}$$

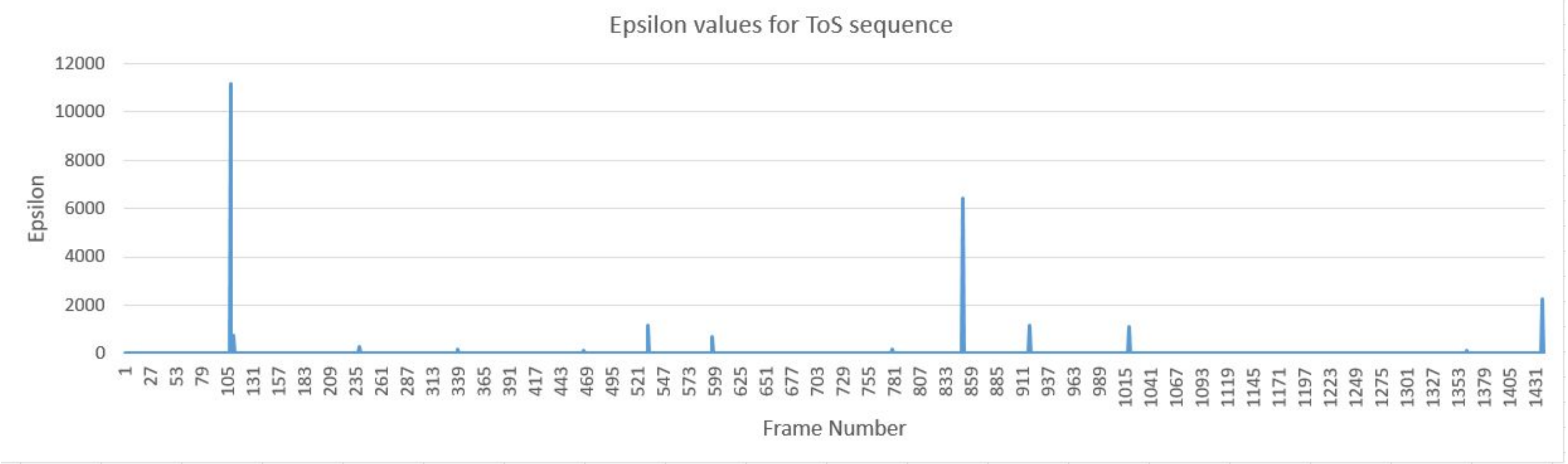


Fig. 7: Epsilon values for ToS sequence. Please note that shot transitions happen at frames: 107, 110, 238,338,465,531,596,778,850,917,1018,1361,1437.

Shot Detection

The algorithm is classified into two steps:

- Feature extraction
- Successive Elimination Algorithm

Video	Actual shot-cuts	Benchmark algorithm				Proposed algorithm			
		Accuracy	Precision	Recall	F-measure	Accuracy	Precision	Recall	F-measure
BigBuckBunny	10	99.88%	100.00%	80.00%	88.89%	**100.00%**	**100.00%**	**100.00%**	**100.00%**
Dinner	4	99.89%	100.00%	75.00%	85.71%	**99.89%**	**100.00%**	**75.00%**	**85.71%**
FoodMarket4	2	99.72%	-	0%	-	**99.86%**	**100.00%**	**50.00%**	**66.67%**
sintel_trailer	14	99.86%	100.00%	85.71%	92.31%	**99.93%**	**100.00%**	**92.86%**	**96.30%**
snow_mnt	3	99.47%	-	0%	-	**99.65%**	**100.00%**	**33.33%**	**50.00%**
Tears_of_Steel	13	99.93%	100.00%	92.31%	96.00 %	**100.00%**	**100.00%**	**100.00%**	**100.00%**
Busy City	11	99.64%	50.00%	18.18%	26.67%	**99.87%**	**100.00%**	**63.64%**	**77.78%**
FunOnTheRiver	12	99.60%	0%	0%	-	**99.80%**	**85.71%**	**50.00%**	**63.16%**

Benchmark is the default shot detection algorithm of x265.

Source: V. V. Menon, H. Amirpour, M. Ghanbari, and C. Timmerer,"Efficient Content-Adaptive Feature-Based Shot Detection for HTTP Adaptive Streaming," in 2021 IEEE International Conference on Image Processing (ICIP), 2021, pp. 2174–2178.

Future Roadmap

Future Roadmap

- The initial version will be released before March 1, 2022.
- Adding Multi-threading support
  - H computation for blocks in each frame can be realized concurrently.
  - ~6x speedup expected with 8 threads.
- Adding CUDA/ OpenCL support

Thanks for your attention!

Vignesh V Menon (vignesh.menon@aau.at)
Hadi Amirpour (hadi.amirpour@aau.at)