VCAv2.0: A green video complexity analysis

Vignesh V Menon
Christian Doppler Laboratory ATHENA, Alpen-Adria-Universität, Klagenfurt, Austria

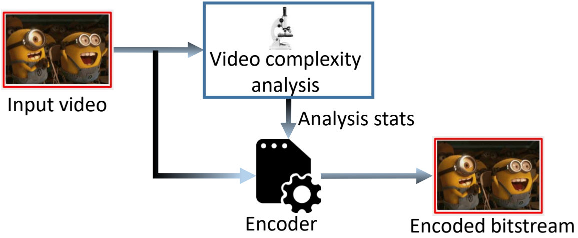
29 June 2023

Outline

Background

Why VCA?

- The optimal encoding parameters depend on the **video content complexity**.
- There is a need to extract content features that can represent the video content complexity to predict the optimal encoding parameters for that video content.



Texture Energy

Compute texture energy per block

A DCT-based energy function is used to determine the block-wise feature of each frame defined as:

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

where wxw is the size of the block, and $DCT(i,j)$ is the $(i,j)^{th}$ DCT component when $i + j > 0$, and 0 otherwise.

The energy values of blocks in a frame are averaged to determine the energy per frame.[1]

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

[1] Michael King, Zinovi Tauber, and Ze-Nian Li. "A New Energy Function for Segmentation and Compression". In: *2007 IEEE International Conference on Multimedia and Expo*. 2007, pp. 1647–1650. DOI: 10.1109/ICME.2007.4284983.

Texture energy gradient

h_p: SAD of the block level energy values of frame p to that of the previous frame $p-1$.

$$h = \sum_{k=0}^{C-1} \frac{\mid H_{Y,p,k} - H_{Y,p-1,k} \mid}{C \cdot w^2} \tag{3}$$

where C denotes the number of blocks in frame p.

Luminescence

The luminescence of non-overlapping blocks k of each frame p is defined as:

$$L_{Y,k} = \sqrt{DCT(0,0)} \tag{4}$$

where $DCT(0,0)$ is the DC component in the DCT calculation. The block-wise luminescence is averaged per frame denoted as L_Y as shown below.

$$L_Y = \sum_{k=0}^{K-1} \frac{L_{Y,p,k}}{C \cdot w^2} \tag{5}$$

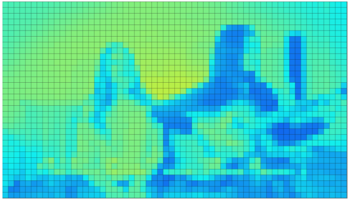
where C denotes the number of blocks in frame p.

Chroma features

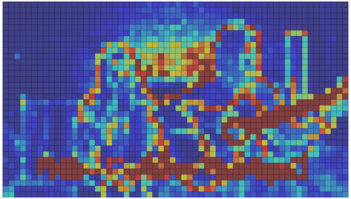
VCA also determines chroma texture energy E_U and E_V (for U and V planes), and the chrominance L_U and L_V (for U and V planes).

Video Complexity Features

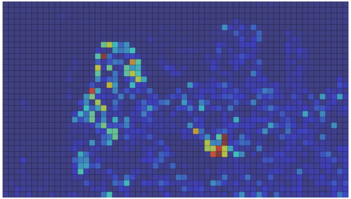
(a) Original video frame



(b) Heatmap of L



(c) Heatmap of E



(d) Heatmap of h

Figure: Example heatmap of Luminescence (L), spatial texture (E) and temporal activity (h) features of the 2^{nd} frame of *CoverSong_1080P_0a86* video of Youtube UGC dataset extracted using VCA.

Accuracy Analysis

Correlation of spatial complexity features with the ground truth

Bitrate in All Intra configuration[2] is considered as the ground truth of the spatial complexity.

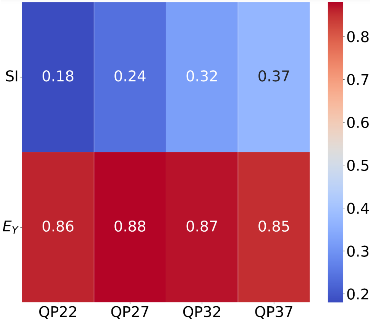


Figure: PCC between SI and E_Y, respectively, with bitrate in All Intra configuration with *medium* preset of x265 encoder for the VCD dataset.[3]

[2] Jill Boyce et al. *JVET-J1010: JVET common test conditions and software reference configurations*. July 2018.

[3] Hadi Amirpour et al. "VCD: Video Complexity Dataset". In: *Proceedings of the 13th ACM Multimedia Systems Conference*. MMSys '22. Athlone, Ireland: Association for Computing Machinery, 2022, 234–239. ISBN: 9781450392839. DOI: 10.1145/3524273.3532892. URL: https://doi.org/10.1145/3524273.3532892.

Accuracy Analysis

Correlation of complexity features with *bitrate* in the Low Delay P picture (LDP) configuration

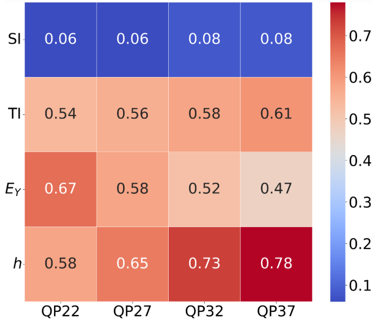


Figure: PCC between SI, E_Y, TI and h with *bitrate* in the Low Delay P picture configuration with *ultrafast* preset of x265 encoder for the VCD dataset.

E_Y and h strongly correlate with the encoding bitrate.

Performance Optimizations

Multi-threading optimizations

- Creates multiple threads within a VCA execution instance, which executes independently but concurrently, sharing process resources.
- Independent threads carry out DCT-energy computation per block.

x86 SIMD optimizations

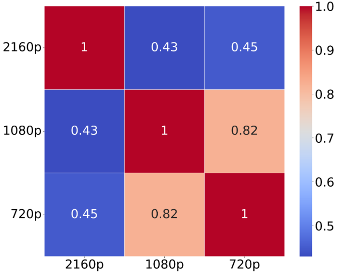
- SIMD optimization[4] of DCT functions implemented as intrinsic and assembly codes for x86 architecture.

[4] Praveen Kumar Tiwari et al. "Accelerating x265 with intel® advanced vector extensions 512". In: *White Paper on the Intel Developers Page* (2018).

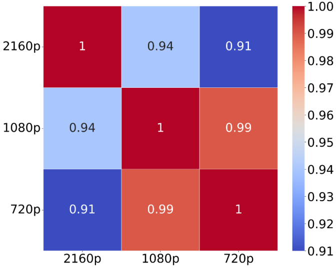
Performance Optimizations

Low-pass DCT optimization

- Unlike SI, the E_Y feature exhibits better correlation across resolutions.
- In VCA v2.0, the complexity features are evaluated on the video spatially downsampled by a factor of two.



(a)



(b)

Figure: PCC between the spatial complexity features (a) SI and (b) E_Y across multiple resolutions for the VCD dataset.

Performance Optimizations

Processing time results

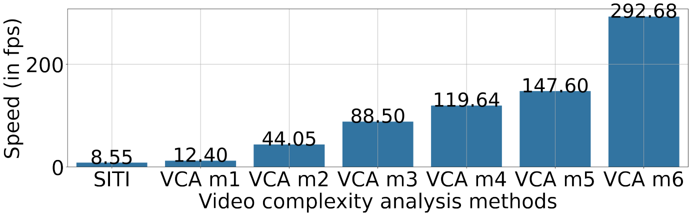


Figure: Processing time of video content complexity analysis methods.

m1: without any performance optimizations
m2: with SIMD
m3: with SIMD and low-pass DCT
m4: with SIMD , low-pass DCT and multi-threading (2 threads)
m5: with SIMD , low-pass DCT and multi-threading (4 threads)
m6: with SIMD , low-pass DCT and multi-threading (8 threads)

Performance Optimizations

Energy consumption results

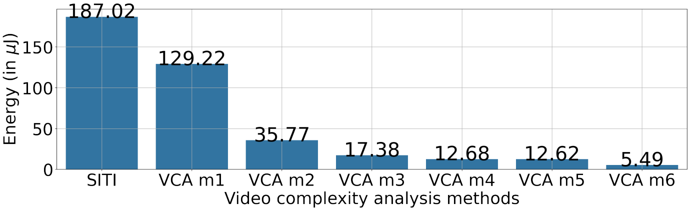


Figure: Energy consumption of video content complexity analysis methods.

m1: without any performance optimizations
m2: with SIMD
m3: with SIMD and low-pass DCT
m4: with SIMD , low-pass DCT and multi-threading (2 threads)
m5: with SIMD , low-pass DCT and multi-threading (4 threads)
m6: with SIMD , low-pass DCT and multi-threading (8 threads)

Conclusions

- VCA is an open-source video complexity analyzer library published under the GNU GPLv3 license.
- Low-complexity DCT-based energy features are extracted using VCA v2.0, which encoders use to derive decisions like bitrate-ladder, frame-type, block-partitioning, and much more.
- Multi-threading, x86 SIMD, and low-pass DCT optimizations improve the energy efficiency of the VCA implementation.
- Compared to the state-of-the-art SITI implementation of video complexity analysis, VCA v2.0 yields a better estimation of video complexity, with an energy consumption reduction of 97.06%.

Future steps

- Identify parallelizable code and perform SIMD optimization
- ARM optimization
- CUDA/ OpenCL optimization
- Open-source the prototypes of the VCA applications

Q & A

Thank you for your attention!

Vignesh V Menon (vignesh.menon@aau.at)

VCA