

Detection of Hard-To-Compress Video Sequences Based on Objective Quality Metrics

Filip Korus, Dawid Juszka, Michał Grega, Mikołaj Leszczuk, Michał Suchoń

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie AGH University of Krakow

tvhive, Krakow

03.07.2024



Large Volume & Demand

According to Cisco:

- 3-fold increase in IP traffic over the 5 years (from 2017)
- 82% of all IP traffic was IP video traffic (in 2022)



325.4 EB



Upload Increase



According to Cisco:

- the average internet user generated 84.6 GB of traffic per month (by 2022)
- up 194% from 28.8 GB in 2017

Fixed Consumer Internet Traffic per User per Month

84.6 GB





Gameplay and e-sport events

- A growing category of IP traffic is the streaming of gameplay and e-sports events.
- Netflix, YouTube or Twitch gains popularity, attracting fans of films and series, as well as computer game enthusiasts.
- E-sports events feature professional gamers competing against each other to win real money.
- The skill and determination of these professional gamers attract millions of viewers worldwide.
- Major e-sports events gather enormous amount of fans during live broadcasts as well as those who watch replays.



Intel Extreme Masters, Katowice 2018



Why encoding gameplay videos could be difficult?

- High level of details in the game environment
- Fast-paced action scenes
- Complex textures and patterns
- High frame rates
- Use of dynamic lighting and shadows
- Particle effects like explosions, smoke, water and fire



Counter-Strike: Global Offensive gameplay



Minecraft gameplay



Spatial Information (SI), Temporal Information (TI), criticality (o) metrics

• SI measures the amount of spatial detail in a video:

$$SI = max_{time} \{ std_{space} [Sobel(F_n)] \}$$

Fn is the n-th frame of the video

SI value greater than 200 means that video is very detailed*

• TI measures the amount of motion: $TI = max_{time} \{ std_{space} [F_n - F_{n-1}] \}$

high motion videos have TI higher that 60*

• Criticality is defined as the logarithm of the mean product of SI and TI values over time for a given frame: $o = log_{10} \{mean_{time}[SI(F_n) * TI(F_n)]\}$

criticality value under 2.5 means video is easy to encode* criticality value above 3.5 means video is hard to encode*



Experiment

- Database consisting of 1080p, 24 or 60 fps of gameplays, CCTV footage and Netflix series
- Every video was compressed using H.264 codec at bit-rates ranging from 100 kbps to 21 Mbps with step of 100 kbps
- VMAF was calculated for every compressed video (VMAF value greater than 90 means there is no loss of the quality)
- SI, TI and criticality metrics were computed

Results

- Most of the videos reached VMAF 90
- Several video sequences have never reached desired quality level defined by VMAF value of 90 (even using 21 Mbps bitrate)





Results

- Most of the videos reached VMAF 90
- Several video sequences have never reached desired quality level defined by VMAF value of 90 (even using 21 Mbps bitrate)

• Hard-to-encode videos have SI ≈ 100 and TI ≈ 10







Random Forest Classifier

- Binary classifier
- Input features: SI, TI, o, mean SI, mean TI
- Output class: 1 or 0 determining whether video reaches VMAF 90 or not
- Model accuracy is 92%, AUC is 0.98
- The ROC curve indicating a high true positive rate and a low false positive rate.





Feature Importance

- Using feature importance
- Defining which feature had the greatest impact on the prediction



Model Testing

- Normalized confusion matrix is used for presenting the results.
- Labels classify the video:
 - 1 (easy-to-encode) level of VMAF 90 reached
 - 0 (hard-to-encode) level of VMAF 90 not reached







Conclusions and Future work

 Described binary classification model could be used for choosing hard-to-compress videos

What next?

- A tool that would indicate shots (video fragments) that require higher bandwidth for encoding.
- It would be possible to temporarily change codec parameters to maintain constant quality.



Thank you for your attention!



Presented results come from the research project entitled "Technologia do bezpiecznego i niezawodnego dostarczania profesjonalnych przekazów kontrybucyjnych audio/wideo na żywo przy zachowaniu minimalnego możliwego opóźnienia." (eng. "Technology for safety and reliable delivery of professional live audio/video contributions while maintaining minimum possible delay.") co-financed by the National Fund Research and Development Center.

https://qoe.agh.edu.pl

agh.edu.pl