
# QUANTIFYING HUMAN PERCEPTION OF VIDEO QUALITY

**SOPHIA BATSI AND LISIMACHOS P. KONDI**

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, UNIVERSITY OF IOANNINA, GR-45110, IOANNINA, GREECE

GEC 2019 - Athens, Greece

# STATE OF PLAY

- Currently, QoE metrics are used to estimate human's perception of video quality
- One of these metrics is VMAF

## VMAF

Video Multimethod Assessment Fusion–VMAF, is a machine learning based model, which reflects human's perception of video quality.

Current VMAF implementation temporally pools the per-frame scores using the arithmetic mean, leading to an overestimation of the provided quality.

# EVALUATION STUDY

- Minkowski mean

$$OM_{Mink} = \left[\frac{1}{T}\sum_{t=1}^{T} OM^{P}(t)\right]^{1/p}, \qquad p \neq 0$$

Where $T$ is the number of frames in a video (frame sequence), $p$ is the Minkowski exponent and $OM$ stands for VMAF scores per frame

- Minkowski exponent, p, took the following values: -1 (Harmonic mean), 1 (Mean-VMAF default), 2 (Quadratic mean), 5, 8 and 10

- We used two popular datasets to validate the performance of the Minkowski mean for the selected values for the exponent p
  - ✓ A subset of the VQEG HD3 Dataset
  - ✓ The Netflix Video Dataset

# EXPERIMENTAL RESULTS

Measuring :

- Spearman Rank Correlation Coefficient (SRCC)
- Pearson Correlation Coefficient (PCC)
- Root-Mean-Square-Error (RMSE)

Best results were achieved when Minkowski exponent $p$=8 for both datasets.

| VMAF | | | | | | | |
|---|---|---|---|---|---|---|---|
| Pooling Method | | NETFLIX DATASET | | | VQEG HD3 | | |
| | | SRCC | PCC | RMSE | SRCC | PCC | RMSE |
| Minkowski | $p$=-1 | 0.914 | 0.926 | 12.249 | 0.923 | 0.934 | 0.387 |
| | $p$=1 (VMAF default) | 0.918 | 0.934 | 11.529 | 0.924 | 0.936 | 0.384 |
| | $p$=2 | 0.920 | 0.938 | 11.234 | 0.923 | 0.936 | 0.383 |
| | $p$=5 | 0.923 | 0.944 | 10.642 | 0.925 | 0.938 | 0.38 |
| | $p$=8 | **0.924** | **0.946** | **10.422** | **0.926** | **0.939** | **0.377** |
| | $p$=10 | 0.921 | 0.945 | 10.434 | 0.926 | 0.939 | 0.376 |

Minkowski mean as temporal pooling method:

- ✓ provides VMAF scores closer to subjective ones, compered to the arithmetic mean
- ✓ can be easily parametrized to fit in different video datasets or types of video impairments that affect QoE

# CONCLUSIONS

# THANK YOU