Measuring Video Quality with VMAF: Why You Should Care

Christos Bampis
Encoding Technologies, Netflix

AOMedia Research Symposium
San Francisco, October 15, 2019
Overview

- history and introduction to VMAF
- adoption
- challenges
- why is VMAF becoming more useful?
Need a better perceptual metric

<table>
<thead>
<tr>
<th></th>
<th>PSNR 29.1 dB</th>
<th></th>
<th>PSNR 29.3 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>19</td>
<td>Humans</td>
<td>69</td>
</tr>
</tbody>
</table>
VMAF

- accurately measures human perception of video quality
- consistent across content
- works well for picture artifacts relevant to adaptive streaming
  - compression artifacts
  - scaling artifacts
- open-source!

VMAF: Video Multimethod Assessment Fusion
The VMAF chronicle

- **2014**
  - Started collaboration with USC
  - Started collaboration with UT Austin
  - First VMAF running in prod @ Netflix

- **2015**
  - Started collaboration with U. Nantes

- **2016**
  - First public showing at ICIP
  - VMAF went live on Github; first VMAF techblog published

- **2017**
  - VMAF 0.6.1 published; added a phone model
  - libvmaf published; VMAF supported by FFmpeg

- **2018**
  - VMAF-enabled video optimization in prod @ Netflix
  - Speed optimization

- **2019**
  - Speed optimization
VMAF framework

- **Pixel Neighborhood**
  - spatial feature extraction (VIF, DLM)
  - temporal feature extraction (TI)

- **Frame Level**
  - within-frame spatial pooling

- **Training**
  - training with subjective data

- **SVM prediction**
  - “Fusion”
  - per-frame score

- **Human Visual System (HVS) modeling**
  - simulate low-level neuro-circuits

- **Temporal Pooling**
HVS modeling: contrast masking

- One signal (e.g. compression artifacts) becomes more difficult to be detected by human eye when it is superimposed on another masker signal (e.g. the pristine source) of similar spatial frequency and orientation.

[Source: HDR-VDP2, Mantiuk et al. 2011]
**VMAF framework**

**Pixel Neighborhood**
- spatial feature extraction (VIF, DLM)
- temporal feature extraction (TI)

**Frame Level**
- within-frame spatial pooling

**SVM prediction**
- trained model
- “Fusion”

**Training with subjective data**

**Temporal pooling**
- per-frame score

Machine learning: align features with subjective scores

---

Netlix
Lab test: collect subjective scores

Absolute Category Rating (ACR) Scale
Map ACR scale to VMAF scale

Absolute Category Rating (ACR) Scale

VMAF Scale
Demo time!
VMAF adoption examples

- industry
- research community
Integration in 3rd-party tools

10.107 libvmaf

Obtain the VMAF (Video Multi-Angle Factor) score for videos.

The obtained VMAF score is a numeric value in the range of 0-100. The higher the value, the better the quality of the video. It requires Netflix’s vmaf library. If you have the library installed, you can use it in your project. If no model path is specified, the vmaf v0.6.1.pkl model will be used.

The filter has the following options:

model_path
- Set the model path with a relative file path, e.g., "vmaf_v0.6.1.pkl",

log_path
- Set the log path with a relative file path, e.g., "vmaf_v0.6.1.pkl".

Why VQMT?
- Basic Information
- Why update?
- PRO Version (with command-line interface) benefits [Full List]
- Key Contributors

Support
- Setting & Screenshots (advanced files settings, visualization settings)
- Metrics Info (PSNR, NISNR, MSE, SSIM, QMEX, QUAD, VQM, MSU
  BPELM, YUV, JPEG, ISO
- FAQ (Frequently Asked Questions)
- Plugins & Plugin SDK
- MSU VQMT program performance
- Subjective quality metrics comparison

ZOND 265 — your indispensable tool

MSU Video Quality Measurement Tool

Projects, ideas: Dr. Dmitry Moskvin, Oleg Petrov, Sergey Pushlin, Sergey Grishin, Arsanoff, George

Video Clarity

Automated Testing Solutions

ClearView – Video Quality Measurement and Analysis Systems

- Resolution and frame rate independent picture quality analyzers with multiple models
- Versatile input/output with 12G-HDSDI, 10G IP, HDMI, IP stream decoder/capture
- Comprehensive media ingest with full featured file decoder application
- Quantitative picture quality metrics VMAF, NISNR, SSIM/MS-SSIM, PSNR, and Samoff JND (optional)
- Test for HDR color and noticeable differences with DETI
- Test for source sequence complexity using Temporal and Spatial metrics
- Performance testing for audio using a perceptual metric for audio quality with AV offset (lp-eqmc) measurement and a peak metric for standardized loudness testing
- Automatic temporal and spatial alignment, comparison playback of uncompressed video via HDSDI or HDMI outputs with interactive side-by-side views up to 8K and 4K @ 60 Hz
- Picture zoom up to 16x with pan and scroll provides unmatched inspection of compression artifacts during any play mode for a detailed view of one or a comparison of two videos
- Desktop window if comparison play modes provides for machine room or data center use without video output connectivity
VMAF in codec comparisons

[Source: JVET-O0451 Subjective Comparison of VVC and HEVC, JVET 15th meeting: Gothenburg, SE, 3–12 July 2019]
VMAF in research papers

VMAF reproducibility: Validating a perceptual practical video quality metric
R Rassool - 2017 IEEE International Symposium on ..., 2017 - ieeexplore.ieee.org
Measuring video quality with standard metrics ensures that operators can deliver to consumers the desired quality of experience (QoE) at an optimal cost. Such metrics also

Video Multimethod Assessment Fusion (VMAF) on 360VR contents
This paper describes the subjective experiments and subsequent analysis carried out to validate the application of one of the most robust and influential video quality metrics, Video

Practical Evaluation of VMAF Perceptual Video Quality for WebRTC Applications
B García, L López-Fernández, F Gortázar, M Gallego - Electronics, 2019 - mdpi.com
WebRTC is the umbrella term for several emergent technologies aimed to exchange real-time media in the Web. Like other media-related services, the perceived quality of WebRTC

NR-GVQM: A no reference gaming video quality metric
S Zadtootaghaj, N Barmann, S Schmidt... - ... on Multimedia (ISM), 2018 - ieeexplore.ieee.org
... NR-GVQM is designed by training a Support Vector Regression (SVR) with the Gaussian kernel using nine frame-level indexes such as naturalness and blockiness as input features and Video Multimethod Assessment Fusion (VMAF) scores as the ground truth ...
What are the challenges?

- design dimensionality
- dealing with noise
Design dimensionality

increased number of dimensions:

- different encoders: H.264/AVC, HEVC, VP9, AV1
- SDR vs. HDR, dark vs. bright scenes
- different viewing conditions (phone vs. TV, 1080 vs. 4K)
- key question: how to design a model that is extensible and consistent?
Dealing with noise

- VMAF underpredicts under noisy source
- assess film-grain synthesis tools (e.g. AV1)
Why is VMAF becoming more useful?

- newer codecs (e.g. AV1) add more perceptual tools to their arsenal and PSNR is not enough to evaluate them
- open-source and well-adopted: problems are easier to find
- we are committed to further improving VMAF’s accuracy and speed
Summary

- VMAF aims to fill the gap in perceptual video quality metrics
- adopted by industry and academia, but there is room for improvement
- becomes more relevant for new and future codecs (AV1, AV2), e.g., for codec comparison, encoding optimization
Questions?