Paper

Christopher Krämmer
Institute for Media Technology
TU Ilmenau
Email: christopher.kraemmer@tu-ilmenau.de

Serge Molina
Systèmes Robotiques et Interactifs
UPSSITECH
Email: serge.molina@kloumpt.net

Anton Schubert
Institute for Media Technology
TU Ilmenau
Email: anton.schubert@tu-ilmenau.de

***Abstract*—Abstract**
***Index Terms*—4k, videoquality, VMAF, bitrate ladder**

I. Introduction

Cool Introduction here

II. Test Preparation

A. Video Selection

1) Preconsiderations: There exists only a few public 4k video databases. They differ in content (scenario, field size, camera movement, illumination condition) as well as in technical aspects (resolution, frame rate, bit rate and color bit depth). To obtain reference video files for the subjective test, parts of the 4k content from the databases are used (also called source video files). For our test we want to have a large variety between the reference video files to evoke different encoding properties. All the reference videos should have a duration of 10 seconds with no cuts inside. To avoid the influence of judder, the smallest permitted frame rate of the source videos is 50 frames per second (fps). Furthermore the smallest resolution considered to be 3840x2160. If the source videos are not 4k (4096x2160 pixels) they will be upscaled before distortion. Moreover the frame rate of the reference video files is being adapted to consistent 50fps.

2) Dataset Preparation: For obtaining a good variety of video sequences three data bases are used: The Harmonic [**?**] which contains 18 different video files, Cable Labs [**?**] with 9 relevant 4k contents and the Blender Foundation [**?**] for receiving the cartoon Big Buck Bunny. They are partial under the creative commence license and all of them are available in the ProRes format, except of Big Buck Bunny where the 4k video content can be downloaded only as a compressed video file, however there also exist high quality PNGs for all the frames of the cartoon. In order to generate a high quality reference video with the frames of Big Buck Bunny, an automatisation script is used to find a sequence with a duration of 10 seconds, to download the respective images and to encode them as a video file. Generally, the challenge is to extract 10 seconds sequences from the original video files with no cut inside because there exists only a few.

One problem of the video sequences from the database is that even they have a high bit rate and are stored as ProRes, no conclusion referring to the contained visual quality can be done. For this reason a preselection with a very powerful system, able to play 4k content in ProRes with high bit rates and connected to a 4k screen, was applied. The remaining reference videos are 6 source clips, each with a duration of 10 seconds. All the files have got a resolution of 3840x2160 pixels and are stored in the ProRes format. Because this format is a 10 bit codec only, the original color depth of each video file can not be determined. Further specific informations about the reference videos are shown in Table I.

Sequence Name	Frame Rate in fps	Bit rate in Mbps	Timestamps m:s.ms	Source
Air Show	59.94	1703	00:48.500	H
Big Buck Bunny	60	2304	05:47.000	B
Fjord	50	1469	00:21.000	H
Moment of Intensity	59.94	1822	02:16.000	C
Snow Monkeys	59.94	1750	00:17.000	H
Streets of India	50	2094	00:00.000	H

TABLE I
META DATA OF THE VIDEO FILES. THE TIMESTAMPS ARE THE START POSITIONS OF THE EXTRACTION OF 10 SECOND SEQUENCES FROM THE SOURCE FILES, WHERE M STANDS FOR MINUTES, S FOR SECONDS AND MS FOR MILLISECONDS RESPECTIVELY. FURTHER SHORTCUTS: H = HARMONIC, B = BLENDER FOUNDATION, C = CABLE LABS

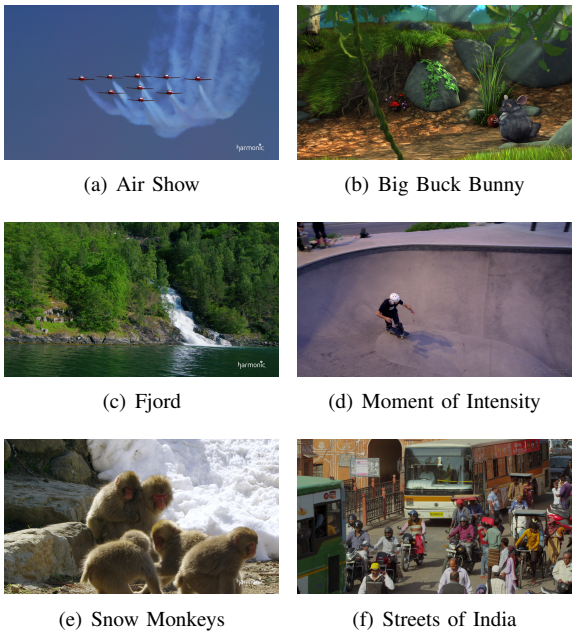
They contain a wide range of high-level features (animation, camera motion, people, water) and low-level characteristics (brightness, contrast, texture, motion, color variance, sharpness) as can be seen in Figure 1.

B. Encoding Parameters

1) Selection of Bitrates: Video Multi-Method Assessment Fusion (VMAF) is a full reference metric for estimating human perception of video quality [1].

To estimate relevant HEVC bitrates for our source content we sample the VMAF scores for different resolutions. The reference sequences are resampled to a fixed 50 frames per seconds to avoid frame rate differences, while the distorted sequences are downsampled, encoded with CBR rate control and upsampled to 4k again using lanczos resampling. The resulting VMAF scores show an overlap between different resolutions as seen in Fig. 3 and the final encoding bitrates are chosen near those intersections.

2) Encoding Presets: Two different presets are used for the sequence encodings. The first is a simple CBR-encoding and the second a 2-pass encoding with a Q-CTRL pass followed by a B-CTRL pass. Every sequence is encoded with both presets at the 3 resolutions and 3 bitrates. The resulting VMAF-scores for the encoded sequences can be seen in figure **??**.



(a) Air Show (b) Big Buck Bunny

(c) Fjord (d) Moment of Intensity

(e) Snow Monkeys (f) Streets of India

Fig. 1. Selection of frames contained in the reference videos to show the variety of the content.

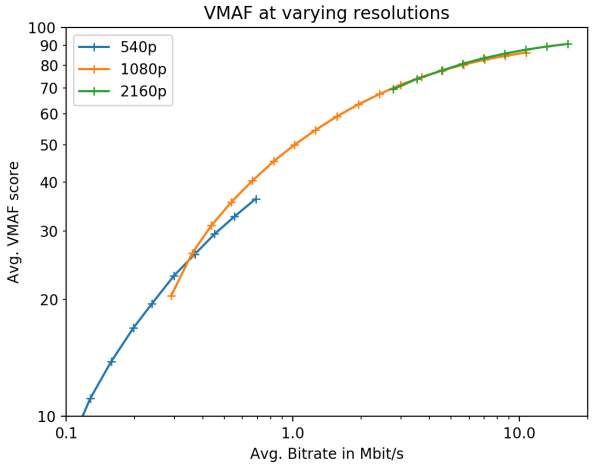


Fig. 2. VMAF scores for 25 different bitrates at 3 resolutions

C. Test Setup

Several steps have been followed in order to enable the acquisition of meaningful data from the test participants. Theses steps included defining the test environment, choosing a rating framework and creating additional question that may help inferring informations about the participants and their way of rating content.

1) Test Environment: In order to make the results of our research reproducible the ITU recommendations [2] have been followed. The key point being:

- Viewing distance
- Peak luminance of the screen

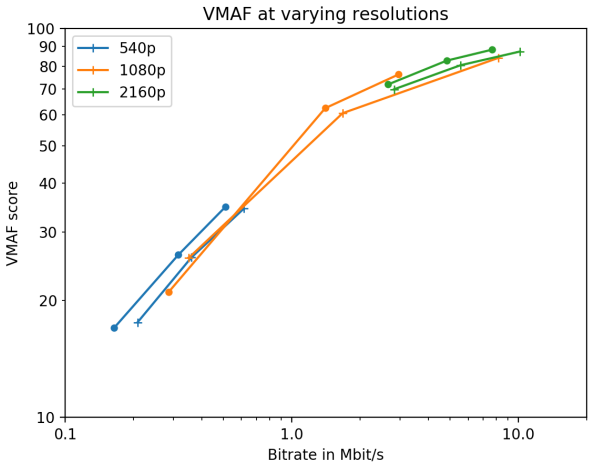


Fig. 3. VMAF scores of encoded videos

- Ratio of luminance of the screen in different conditions
- Ratio of luminance of background behind picture monitor to peak luminance
- Chromaticity of background
- Background room illumination

As performing test following these specifications is common in the Ilmenau's Technical University, a room meeting the requirements was already available and therefore used.

2) Rating Framework: The testing procedure which seems the more suited to the case was the ACR [2] where different versions of an original sequence are showed to a test participant.

For each sequence the participant issues categorical ratings from any of these 5 answers:

- Excellent
- Good
- Fair
- Poor
- Bad

The workflow of this rating framework is the following (see fig. 4):

1 Rating of reference videos
2 Submission of participants personal info
3 Rating of 108 videos
4 Submission of feedback questions (see "Additional collected data")

3) Additional collected data: In order to gain more knowledge about the users behaviour, additional data is collected from the participants.

One of the suggestions of the ITU recommendation paper being the usage of several test questions in addition to the ratings [2], the following questions have been asked at the end of each rating session:

- Presence of blocky artefacts
- Visible bands of colour
- Smoothness of the playback

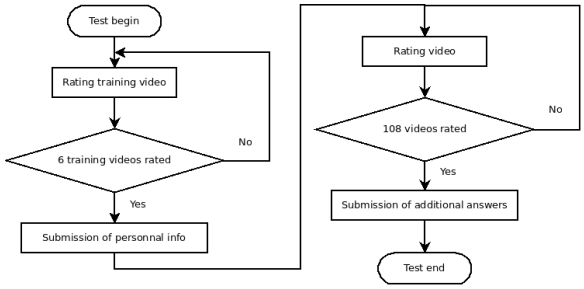


Fig. 4. Rating workflow

- Have you seen 4K content before?
- Have you seen content on a 4k screen before?
- How sure were you about the rating that you provided?

These idea behind these question is dual as these questions may translate how users percieve/are sensitive to video features as well as how users may clearly express their perceptions. As it has already been noticed in previous experiments and in discussions with test participants: users may still rate using a different scales, thus spreading the final MOS or falsely being classified as outlier when their ratings may only represent a shift from the overall population.

Moreover, mouse interactions have been collected during the rating of each sequence. The intent behind this is that as the MOS scale doesn't allow detailed answers some participants may hesitate between two answers and change their answers or hover with their mouse around some answers. Also, answering speed, which could be an indicator of a participant skipping answers or to the contrary being strongly confident of his answers. We believe that several informations can be extracted from these kind of behaviours:

- Confidence in the participant sequence rating
- Confidence in the participant overall rating
- Intermediate scores (eg: 4.5)

III. Test Results

A. Ratings

(MOS Results, Plots and stuff)

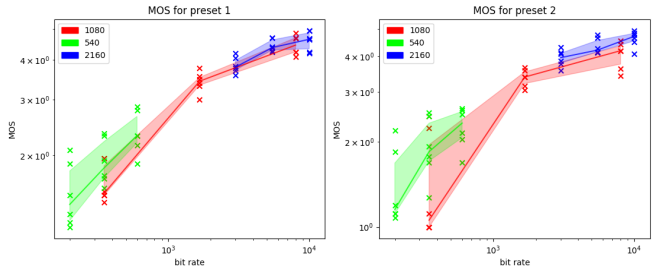


Fig. 5. Correlation between bitrate and MOS (needs improvement)

B. Features

(Mouse tracking, Additional Questions)

Fig. 6. Correlation between mouse related features and rating confidence

Fig. 7. Correlation between feedback related features and rating confidence

IV. Conclusion

References

[1] J. Y. Lin, T. J. Liu, E. C. H. Wu, and C. C. J. Kuo, "A fusion-based video quality assessment (fvqa) index," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, Dec 2014, pp. 1–5.

[2] I. REC, "P. 910," *Subjective video quality assessment methods for multimedia applications*, 1998.

Fig. 8. Correlation between bitrate and MOS after correction based on mouse features

Fig. 9. Correlation between bitrate and MOS after correction based on feedback features

Fig. 10. Correlation between bitrate and MOS after correction based on mouse and feedback features