

Updates on testing activities in MPEG VQA

Mathias Wien
RWTH Aachen University

ISO/IEC JTC 1/SC 29/AG 5
MPEG Visual Quality Assessment

VQEG meeting Klagenfurt, 2024-07-01/05

MPEG VQA Activities: Outline

- Testing activities with JVET
 - Multilayer VVC, Film Grain Synthesis
 - Performance comparison of ECM and VTM:
on-site expert viewing, lab expert and naïve viewing
- CfP for MPEG AI-PCC
- Upcoming: CfP for MPEG Lenslet Video Coding
- CVQM: Dataset for compressed video for study of quality metrics
- ITVVM: Interactive testing of visual volumetric media
- Study on MPEG VQA Remote Expert Viewing tests

Testing activities with JVET

- Verification Test on Multilayer VVC
 - Dual layer spatial scalability in VVC VT
 - Comparison:
 - Full resolution single-layer coding
 - Upscaled base resolution
 - Dual-layer scalable coding
 - Discussion on test scenarios to be used
- Subjective quality testing of FGC SEI message
 - Reconstructed quality over wider rate range
 - Quality of reconstructed video at high quality rate point
 - Evaluate characteristics of film grain with (FG) experts

Second Performance evaluation of ECM vs. VTM

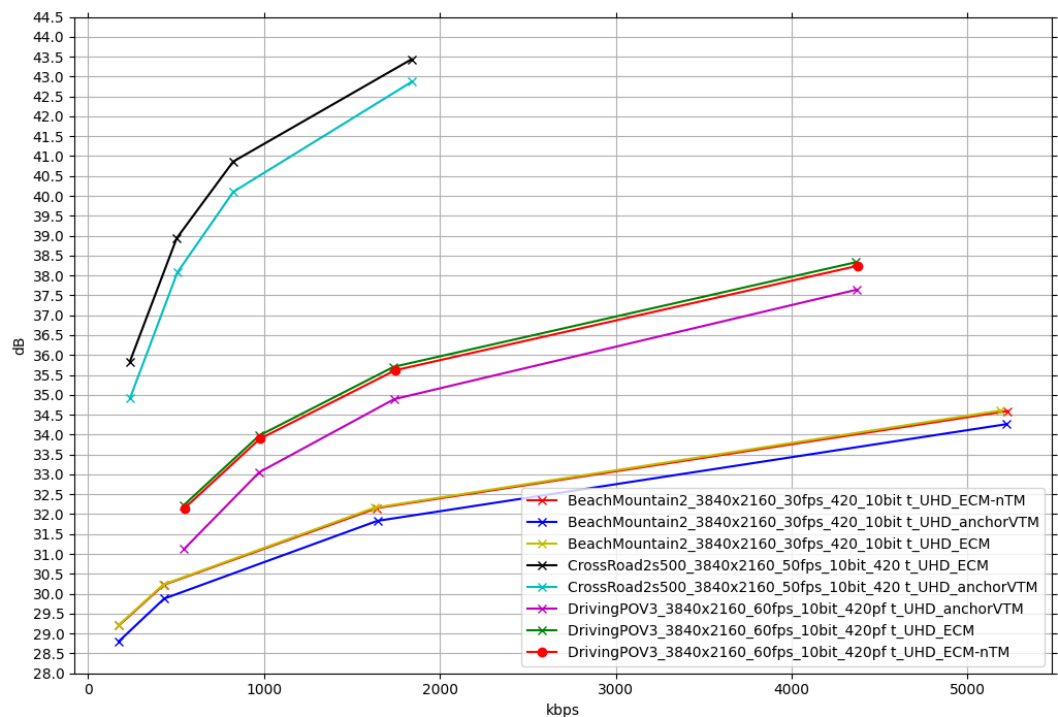
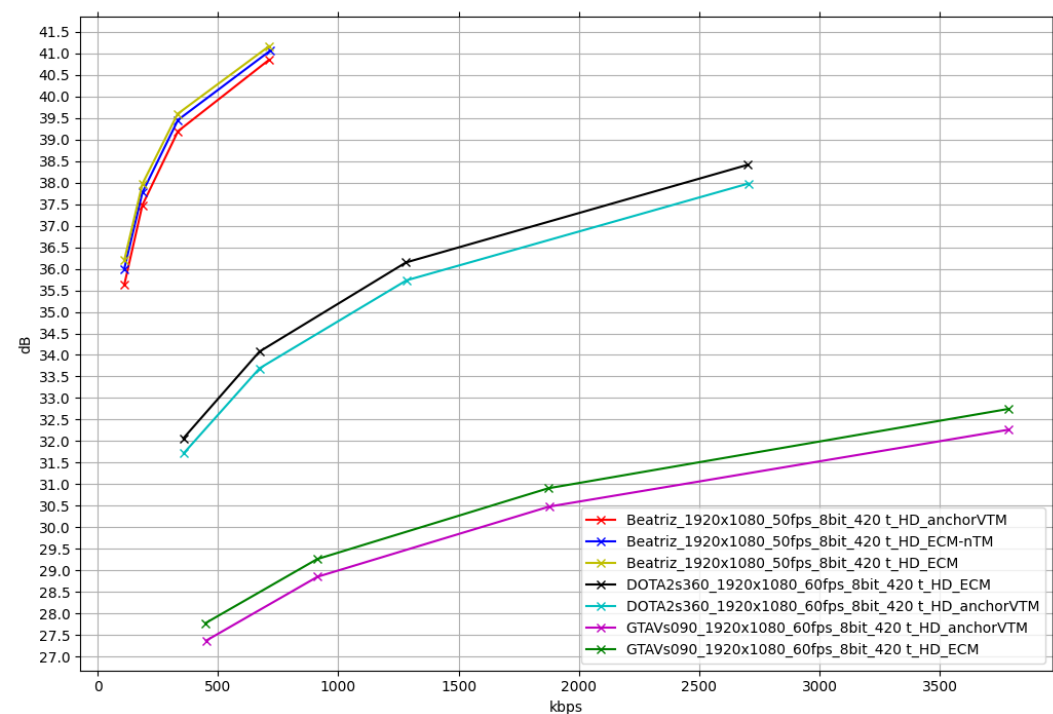
Report: JVET-AH2029

- Comparison points:
 - VTM-11ecm-12 as anchor
 - ECM 12.0
 - ECM-nTM (git hash 6da105bb)
Configuration using **no Template Matching** tools
- 3 UHD sequences (RA cfg), 3 HD sequences (LDB cfg)
- Rate matching:
ECM rate $\leq$ VTM rate (dev. max 2%)
ECM rate $\leq$ ECM-nTM rate (dev. max 2%)
- UHD: 38 PVS
- HD: 34 PVS

	Sequences	rate point s	co de cs	# PVS
UHD VTM/ECM	CrossRoad2s500	4	2	8
UHD VTM/ECM/nTM	DrivingPOV3 BeachMountain2	4	3	24
HD VTM/ECM	DOTA2 GTAVs090	4	2	16
HD VTM/ECM/nTM	Beatriz	4	3	12

Test set

PSNR-Y BD rate numbers



UHD sequences	ECM	ECM-nTM
BeachMountain2	-20.05	-18.98
CrossRoad2s500	-19.00	
DrivingPOV3	-24.39	-21.90
AVG	-22.22	-20.44

HD sequences	ECM	ECM-nTM
Beatriz	-14.85	-9.33
DOTA2s360	-12.18	
GTAVs090	-16.88	
AVG	-14.85	-9.33

Test logistics in Rennes



10	Imperceptible
9	
8	Slightly perceptible
7	
6	Perceptible
5	
4	Clearly perceptible
3	
2	Annoying
1	
0	Severely annoying

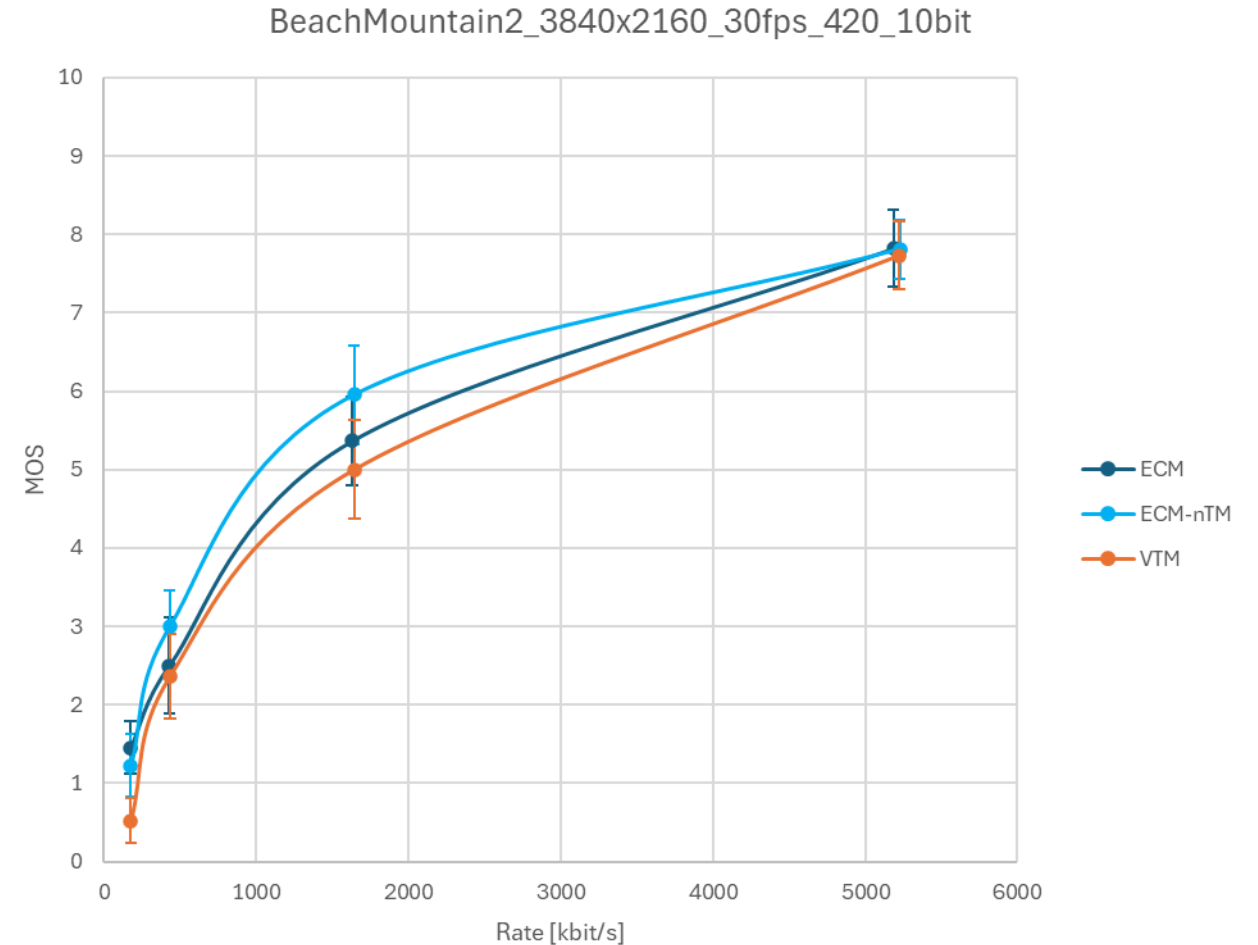
- PC with playout to HDMI via video board
- Distribution to displays via HDMI splitter
- 4× 65“ display Samsung TQ65S90CAT
- Confirmed 4K 60fps 10 bit playout

Viewers

- 24 JMET experts, confirmed for visual acuity and normal color vision
- Efficiency: 3 viewers at 1.5H distance per screen
- HD displayed without scaling centered on UHD screen with gray background

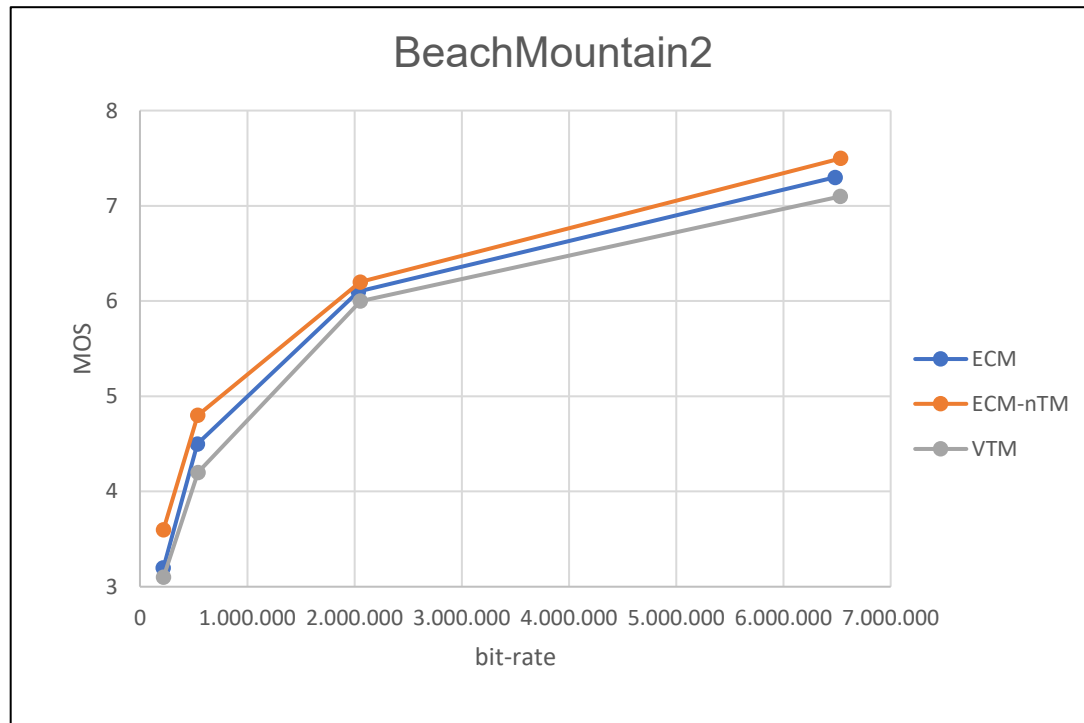
Results: JVET experts in Rennes

- BeachMountain2
- Assessment of VTM, ECM and ECM-nTM
- Indicative BD-rate savings from MOS (partially overlapping CIs)
 - ECM: -12.7%
 - ECM-nTM -29.7%

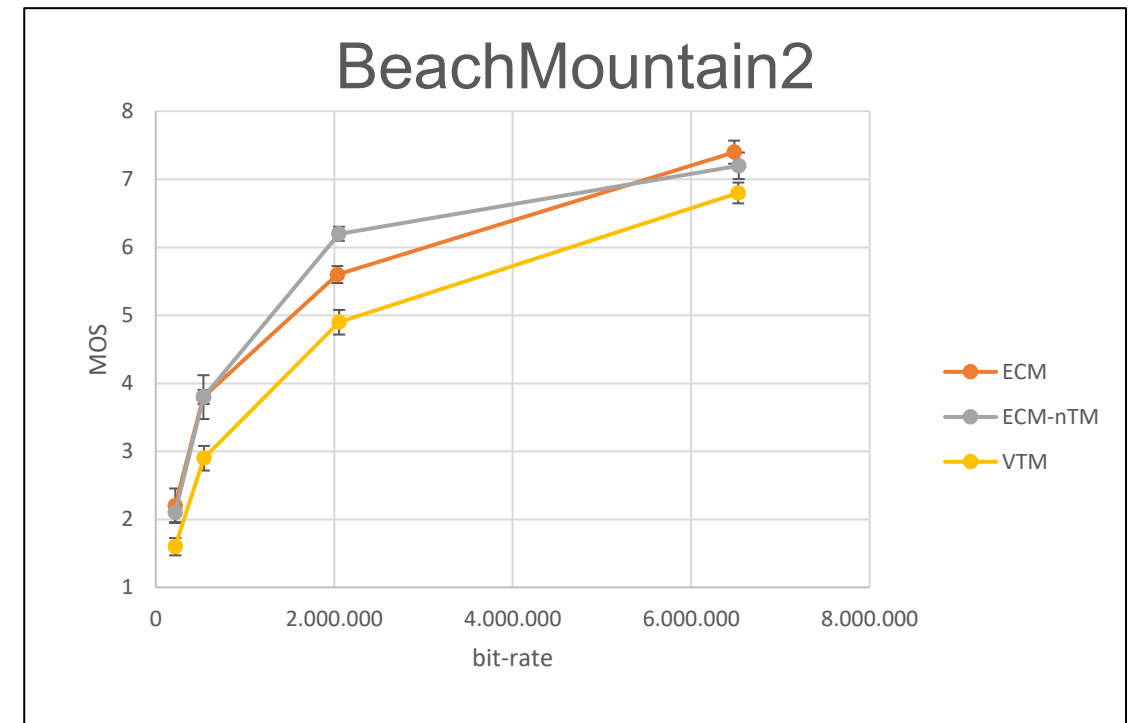


Results: Viewing experts in Rome

Laboratory: Expert viewers (GBTech)

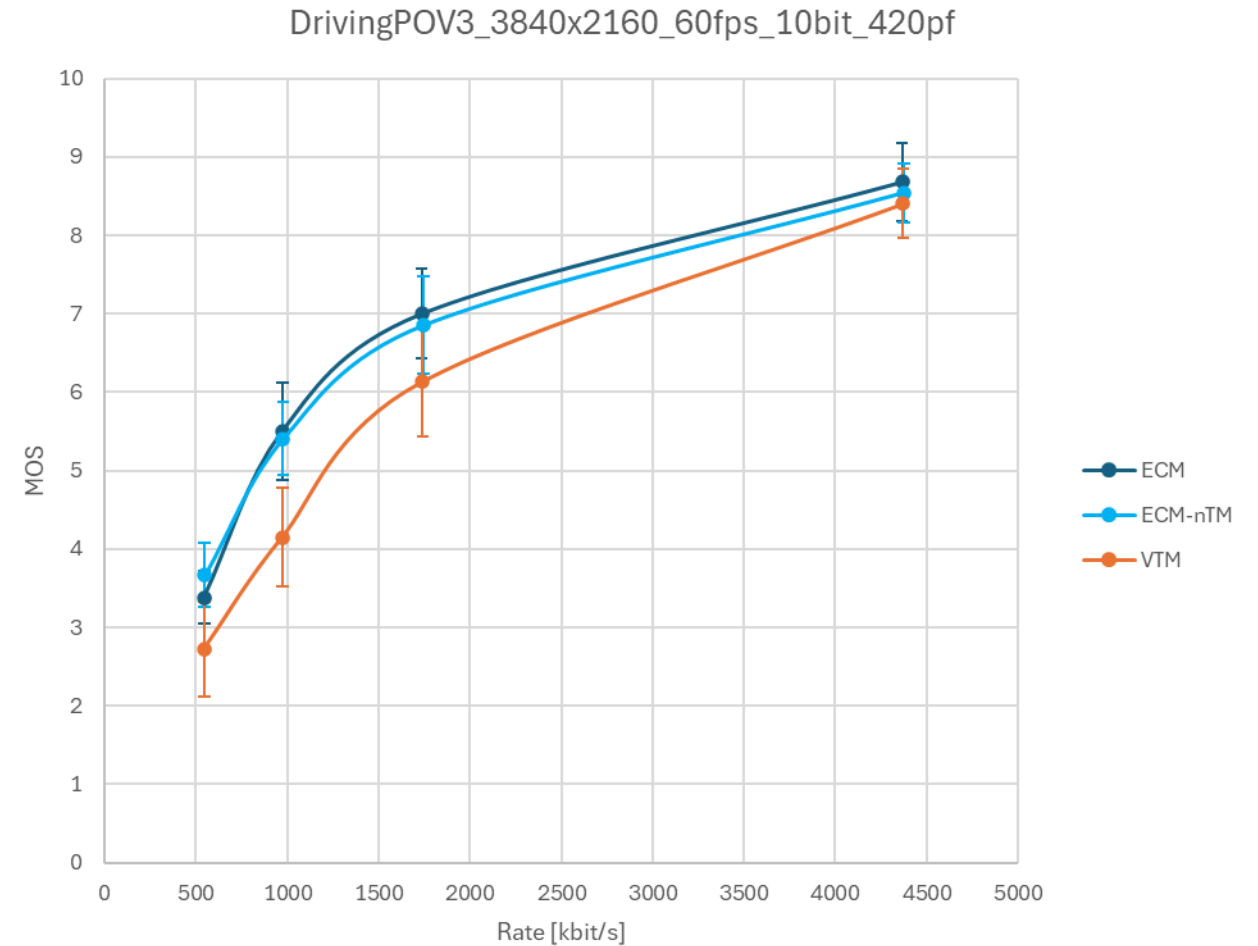


Laboratory: Naive viewers (VABTech)



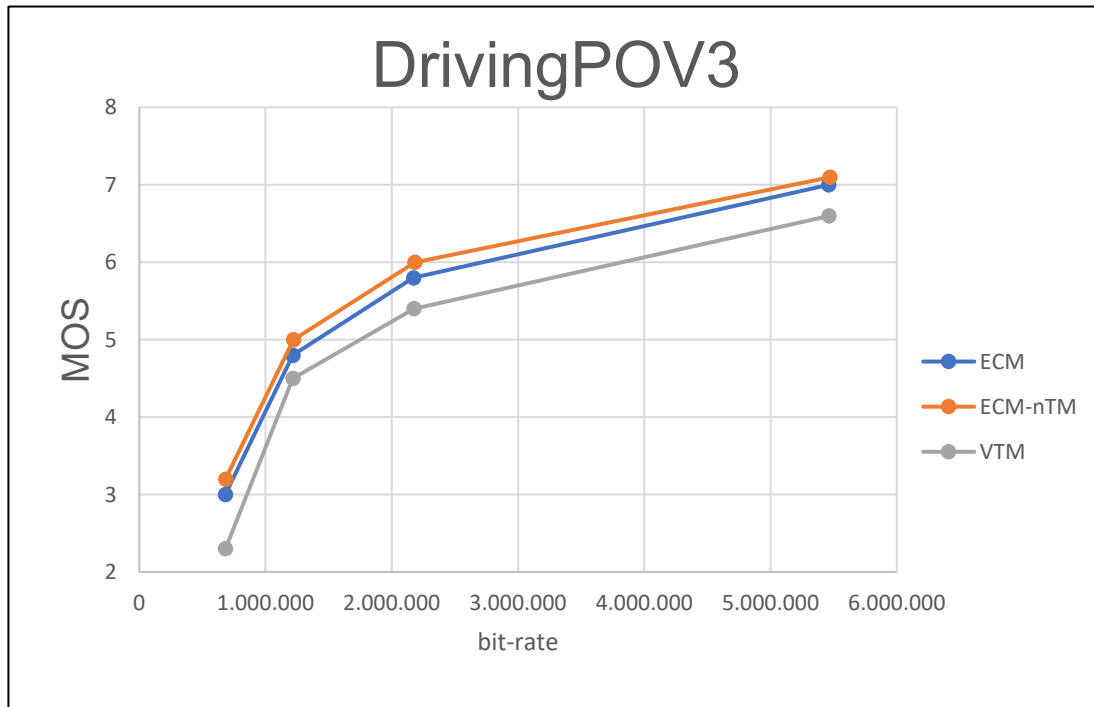
Results: JVET experts in Rennes

- DrivingPOV3
- Assessment of VTM, ECM and ECM-nTM
- Indicative BD-rate savings from MOS (partially overlapping CIs)
 - ECM -28.8%
 - ECM-nTM -26.1%

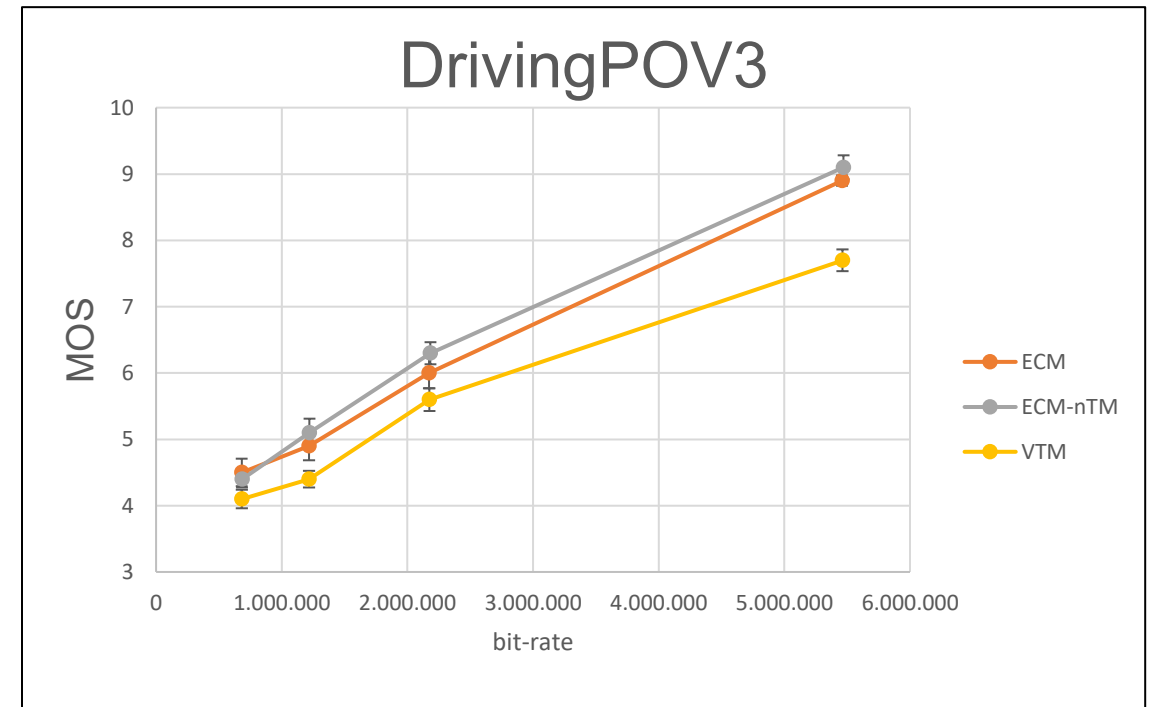


Results: Viewing experts in Rome

Laboratory: Expert viewers (GBTech)



Laboratory: Naive viewers (VABTech)



MPEG AI-PCC CfP

Pre-experiment Test material: m65442

- Total of 60 points under test
 - 3 test sequences
(s1: Model, s2: Basketball_Player, s3: Thaidancer)
 - 3 compression methods under test (m1, m2, m3)
 - 3 rate points each (r1, r2, r3)
1 rendering of uncompressed sequences
 - Rendering: YUV 4:2:0 1920x1080 30Hz 8bit
Simple camera path
 - (a) geometry only
 - (b) geometry plus attributes

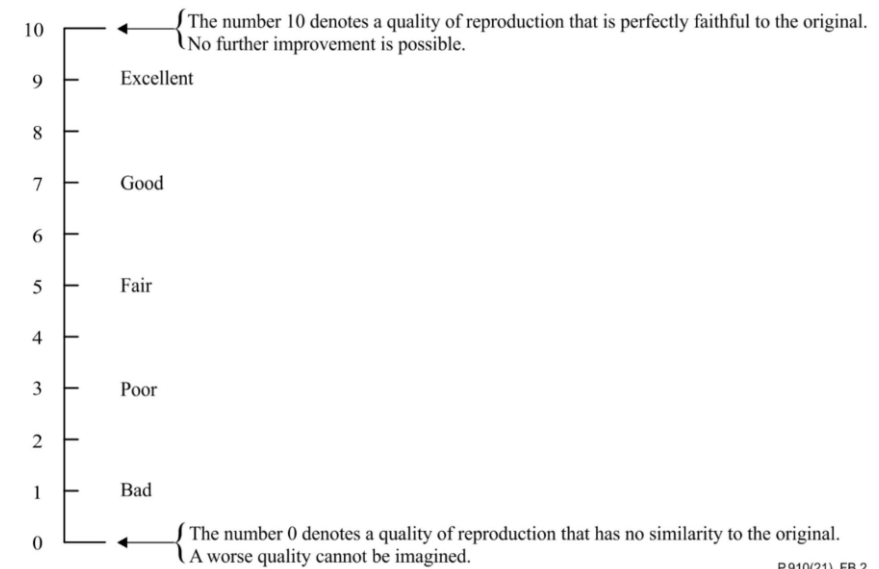


MPEG AI-PCC CfP: Expert viewing test

- **Test protocol: ACR-HR**

Absolute Category Rating with Hidden Reference

- For time efficiency reasons, viewing in groups of 5:
 - 3 experts seated at ~3H
 - 2 experts standing at ~3H
- 23 participants (4 female, 19 male)
 - Requested to confirm acuity and color vision
- 2 test sessions
 - 1st: Geometry only, 2nd: Geometry+Attribute
 - 36 PVS each (incl. stabilization and trapping)
 - Trap: originals shown twice, criterion: divergence ≤ 2
 - Separate training each: followed approach described in JVET-AF0311, no reiterations needed



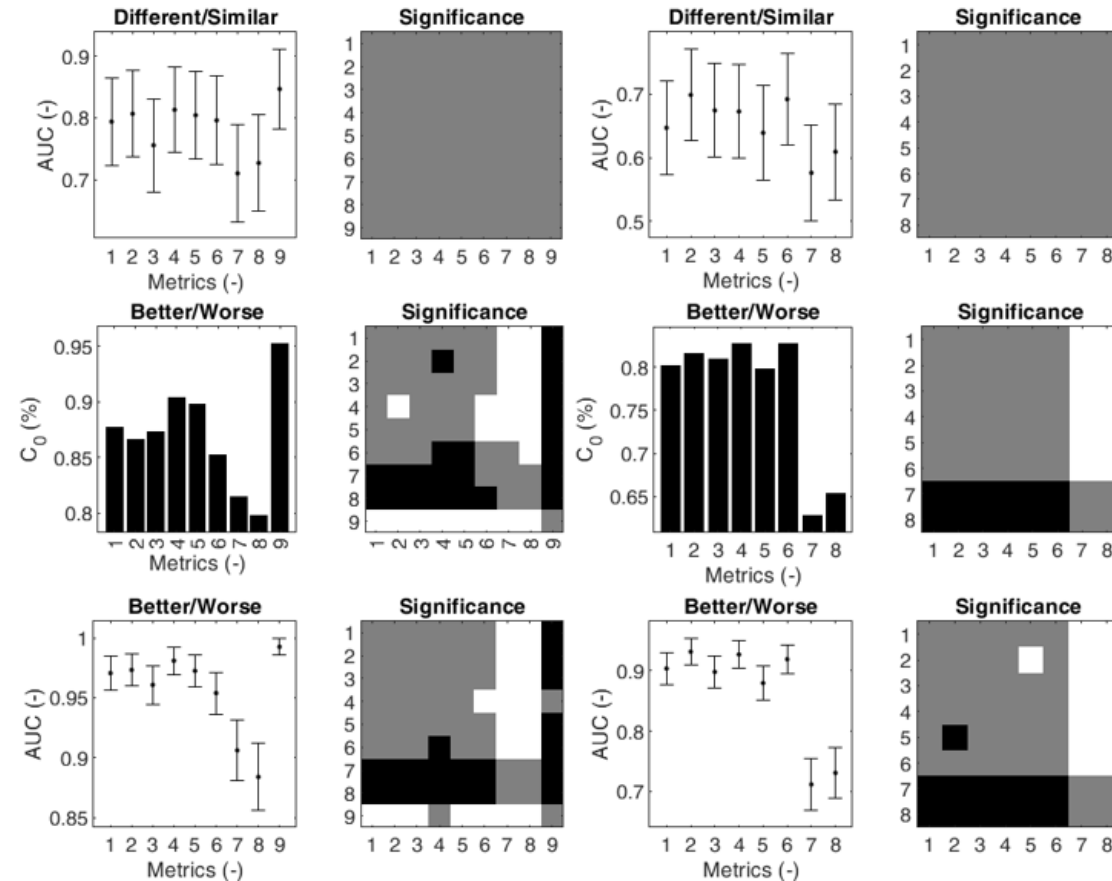
[Doc. m65832](#)

AI-PCC: Analysis of metrics

- Study of objective metrics (image-based, PC-based)

1: <u>ib</u> -3SSIM	2: <u>ib</u> -MSSSIM	3: <u>ib</u> -PSNR	4: <u>ib</u> -SSIM	5: <u>ib</u> -VMAF
6: <u>ib</u> -VQM	7: D1	8: D2	9: PSNRY	

- AUC**: Distinguishing different and similar observed qualities across the test set?
- For the set of test point pairs with different qualities
 - Classification rate** and **AUC**: Correct classification of better or worse quality
- Findings: PSNRY surprisingly effective. Gap for geometry-based metrics D1/D2

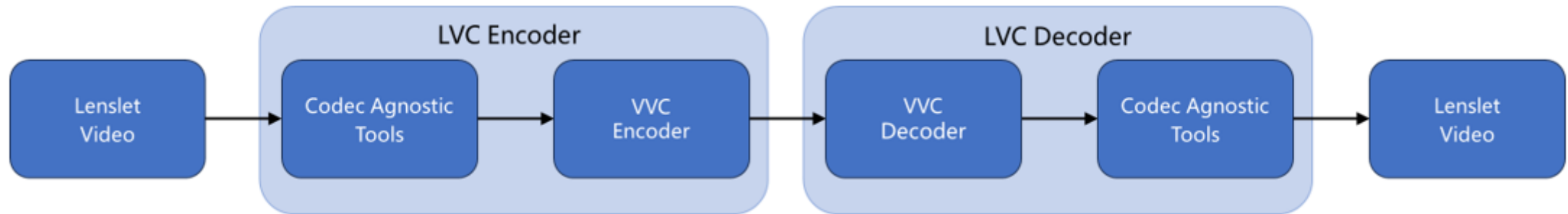


AUC: Area under Receiver Operating Characteristics

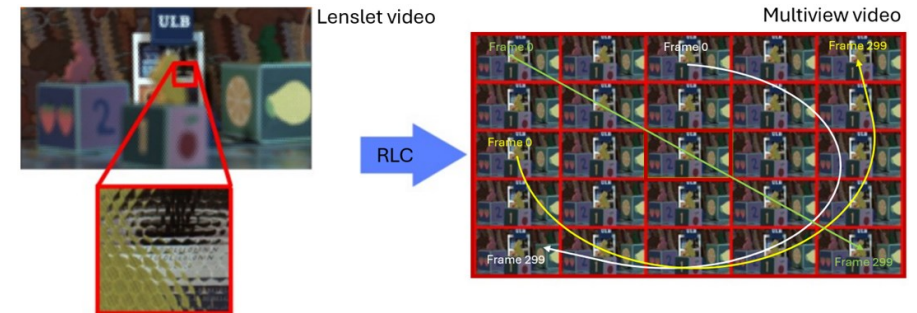
L. Krasula, K. Fliegel, P. Le Callet, and M. Klima, "On the accuracy of objective image and video quality models: New methodology for performance evaluation," in 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), 2016-06. doi: 10.1109/qomex.2016.7498936.

M. Wien and J. Jung, "Performance of point cloud- and image-based quality metrics on geometry-only point clouds and textured point clouds," submitted 2024.

Upcoming: CfP for Lenslet Video Coding



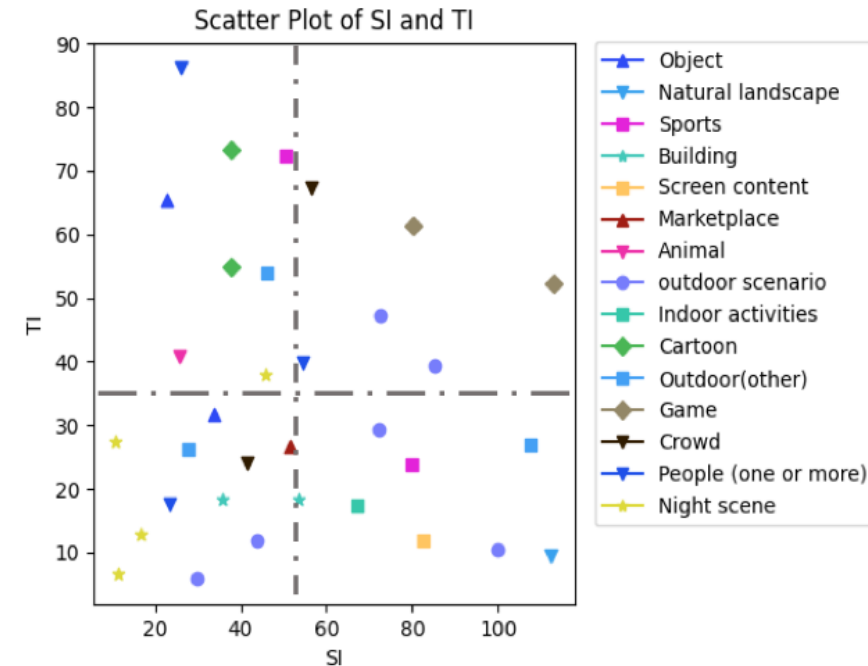
- LVC: towards plenoptic acquisition of light field
 - Enables immersive interaction with LF video
- Subjective evaluation in context of CfP
 - Rendering with a reference method (to be used for all proposals)
 - Viewing of 3D video at approx. HD resolution (1080 lines)
 - Evaluation using DCR/DSIS protocol



Figures from WG4N0488 “Common test conditions for lenslet video coding”, Apr. 2024

CVQM - Dataset of compressed video for study of quality metrics

- 33 sequences covering a wide range of characteristics, 8 rate points
- Conventional codecs: VTM, HM, VVenC, x265
- NN-based codecs (phase 1)
 - NYCU + LUH: MaskCRT
 - USTC + Alibaba: NVC
- Phase 1: HD test sequences
 - 4 rate points for visual evaluation



CVQM - Dataset of compressed video for study of quality metrics

- Status of work
 - $6 \text{ codecs} \times 4 \text{ rate points} \times 8 \text{ sequences} = 192 \text{ test points}$
 - Four volunteering laboratories for subjective tests: RWTH, VABTech, HHI, USTC
 - Pre-experiment: trial viewing test with reference data for cross-check of lab procedures
 - Common instructions for participants
 - Common instructions for laboratories
 - Design of test sessions by test coordinators
 - Common training session (18 BTC)
 - One sync test session (29 BTC, common)
 - Two regular test session (29 BTC each, non-overlapping)
- Plan: completion of tests at all laboratories by mid July
 - Start of analysis and evaluation phase

ITVVM: Interactive testing of visual volumetric media

- **Challenge: reliable, efficient, and fast assessment of immersive video**

- Use cases for immersive video include interaction
- Large variety of applications, end devices

- **Investigation in AhG of AG 5**

- Methodologies for interactive subjective quality assessment of immersive video
- Identify real-time renderers for the purpose of MPEG immersive video visual quality assessment
- Demo at MPEG146 meeting

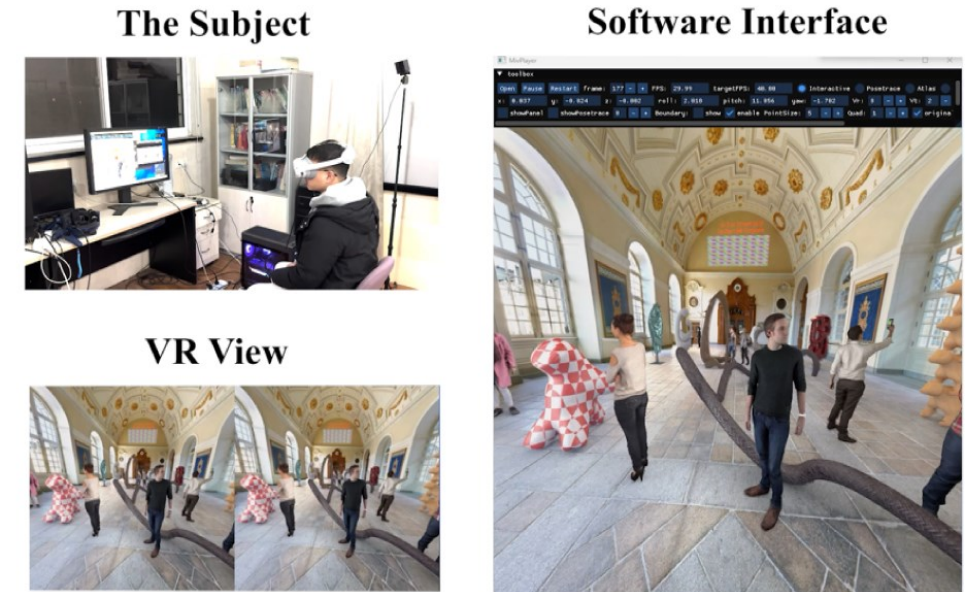


Figure from MPEG doc: Jiaqi Zhang, Siyuan Zhu, Yuxiao Bai, Lu Yu, "Interactive Subjective Visual Quality Evaluation Method for Volumetric Video," MPEG doc. M66437, ISO/IEC JTC 1/SC 29/AG 5, Jan 2024.

Study: Remote Expert Viewing Tests

- Data from remote expert viewing tests conducted during Corona times in JVET and MPEG
 - REV tests, based on MPEG VQA guidelines (AG5N40)
 - Formal laboratory test (DCR)
 - 13 objective metrics, incl. conventional and learning-based metrics
- Analysis w.r.t suitability for decision making in standardization
- Result:
 - Remote expert viewing found to be reliable, good discriminative power
 - Best conventional metrics: PSNR, SSIM, MS-SSIM
 - Best learning-based metrics: VMAF, AVQT



M. Wien and J. Joel: „Remote expert viewing, laboratory tests or objective metrics: Which one(s) to trust?“, EURASIP Journal on Image and Video Processing
DOI : 10.1186/s13640-024-00630-7. May 2024

Conclusions

- **JVET verification testing and subjective evaluations**
 - Multilayer VVC and film grain synthesis studies
- **JVET performance evaluation of ECM**
 - On-site expert viewing test, laboratory tests with expert and naïve viewers
- **CVQM (dataset for compressed video for study of quality metrics)**
 - Data acquisition for phase 1 completed soon
 - Next steps: analysis and consolidation of subjective results, study of objective quality metrics
- **CfP: AI-PCC**
 - Study on objective metrics for point clouds with and without textures
- **CfP: LVC**
 - preparations in progress
- **ITVVM: interactive testing for visual volumetric media**
- **Remote expert viewing**
 - Study performance compared to DCR laboratory tests and objective metrics



AG 5 Contact and Communication Pointers

- Mailing list: mpeg-vqa@lists.rwth-aachen.de
Subscribe at: <https://lists.rwth-aachen.de/postorius/lists/mpeg-vqa.lists.rwth-aachen.de/>
- Convenor: Mathias Wien, wien@lfb.rwth-aachen.de
- Chair for SDR: Vittorio Baroncini, baroncini@gmx.com
- Chair for HDR: Andrew Segall, asegall@amazon.com
- Chair for 360: Yan Ye, yan.ye@alibaba-inc.com
- Chair for MIV: Joel Jung, joeljung@tencent.com

