1. CONTEXT & CONTRIBUTIONS

Keywords / context:
- Cloud Gaming
- Automatic quality evaluation using bitstream derived features (qp, mv, ...)
- Very low complexity and very low delay, no access to decoded pixels

Framework: P.BBBG: Parametric bitstream-based Quality Assessment of Cloud Gaming Services
Work item of ITU-T Study Group 12 Question 14

New family of metrics, maximizing classical performance indicators vs MOS, and jointly maximizing a new performance indicator reflecting ability of metrics to detect sudden abrupt quality changes at frame level, that occur frequently in gaming content

Contributions:
1. Evaluation of existing metrics on heterogeneous gaming dataset (different codecs, bitrates, ...)
2. Design of three new very low-delay and very low-complexity learning-based models
3. Proposition of a new frame level performance indicator to consider gaming content characteristics
4. Proposition of a new training approach to optimize models on all performance indicators

2. GAMING DATASETS

KUGVD dataset
- Kingston University
- 30 videos, 6 games, 1080p@30fps
- H.264, 600 kbps to 4 Mbps
- Lab tests, ACR

CGVDs dataset
- TU Berlin
- 39 videos from 13 games, 1080p@60fps
- H.264 (NVENC), 2 to 6 Mbps

3 datasets merged: 239 PVS and 53 different scenes
Difficulty increased / Realism increased

3. STATE OF THE ART EVALUATION

7 state of the art metrics tested (references in the paper):
- PSNR, SSIM, VMAF: complex, full reference, with access to decoded pixels
- DNNNetGaming: complex, no-reference, NN based, trained on gaming content
- DBCNN: complex, no-reference, DNN based
- P.1203.1, P.1204.3: low-complexity bitstream based models

Dataset split: training (186 PVS) - testing (53 PVS)
Linear mapping applied for RMSE computation (based on training set) [ITU-T P.1401]

Achieving good correlation with MOS on segments of several seconds is insufficient

4. THREE NEW MODELS & RESULTS: learning-based / low-complexity / low-delay

VQMCG: Video Quality Metrics for Cloud Gaming:
- New learning-based models: no-reference, no-access to decoded pixels
- VQMCG.a: weighted linear combination of features, weights learnt on training set with gradient descent
- VQMCG.b: Support Vector Regression (SVR), supervised learning algo mapping the features space with MOS by finding a hyperplane on the training set
- VQMCG.c: Multi-Layer Perceptron (MLP), NN with fully connected layers, weights initialized with Glorot, trained with back-propagation with Adam optimizer (4 layers with 100, 50, 25 and 10 neurons, activated by a ReLU function)

VQMCG.a better than VMAF and NDNetGaming
VQMCG.b and VQMCG.c outperform existing methods on “classical” indicators

Why high performance for the learning-based models? Games (CG) made of repetitive visual characteristics (motion pattern, color diversity, backgrounds, ...)
Similar repeating scenes, spatial and temporal similarities: gaming content adapted to learning-based models

5. DIFFICULTY OF GAMING CONTENT

Gaming content characteristics:
- sudden & fast rotations
- explosions
- large and abrupt/sudden quality changes...
- not reflected by low-complexity metrics

code “stress” (especially low delay): motion prediction issue, intra blocks, ...

FVM: Frame Variation Match: measures ability of a model to reflect large and sudden quality variations when a reference metric reports similar kind of variation
FVM counts % of time when the model has a quality change varMod above a threshold th, when a reference metric also has a quality change varRef above th, and in the same direction, in the same window W

Reference metric: VMAF (any reliable full-reference metric can be used)
Matching a full-reference metric with low-complexity models at the frame level needed

6. NEW FRAME LEVEL CORRELATION INDICATOR

VQMCG.a vs VQMCG.b vs VQMCG.c

VQMCG.a better than VMAF and NDNetGaming
VQMCG.b and VQMCG.c outperform existing methods on “classical” indicators

VQMCG.a vs VQMCG.b vs VQMCG.c

Proposed training process:
- Reference objective scores used at a frame level
- MOS used as an offset applied to the objective scores models trained on VMAF shifted, centered on MOS

7. FINAL RESULTS WITH “GAMING FRIENDLY” LEARNING & CONCLUSION

Proposed training process:
- Reference objective scores used at a frame level
- MOS used as an offset applied to the objective scores models trained on VMAF shifted, centered on MOS

FVM vs VMAF @ frame level improved
AND other indicators vs MOS preserved
Excellent correlation on all indicators