A VMAF Model for 4K

Zhi Li
Netflix

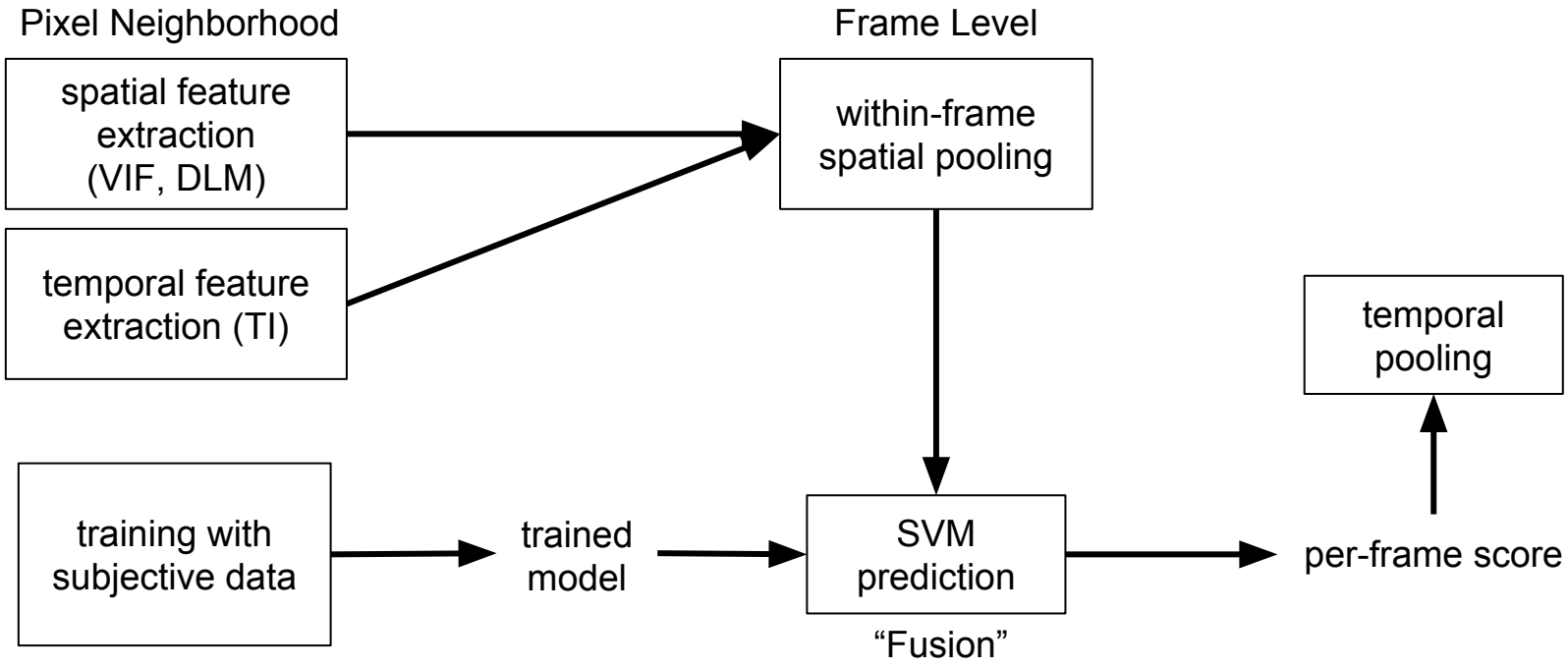
Toinon Vigier, Patrick Le Callet
University of Nantes

VQEG Madrid 2018

Acknowledgements

- Special thanks to Lukas Krasula and Yoann Baveye of University of Nantes, for helping design and conduct the subjective experiment.

VMAF framework



Why do we need a 4K model?

- VMAF v0.6.1 seems to work fine in terms of subjective data correlation

Pearson Correlation

	PSNR (Lin)	VMAF (Lin)	SSIM (poly)	MS-SSIM (poly)	VIF (exp)	VQM (poly)
Class A	0.7212	**0.9389**	0.8873	0.8319	0.8887	**0.9019**
Class B	0.663	**0.9216**	0.5755	0.601	0.6779	**0.8163**

Spearman Correlation

	PSNR (Lin)	VMAF (Lin)	SSIM (poly)	MS-SSIM (poly)	VIF (exp)	VQM (poly)
Class A	0.7756	0.9539	0.8997	0.8376	0.8764	0.8997
Class B	0.7276	0.958	0.5578	0.6185	0.667	0.8125

MSE

	PSNR (Lin)	VMAF (Lin)	SSIM (poly)	MS-SSIM (poly)	VIF (exp)	VQM (poly)
Class A	0.6652	0.3304	0.4429	0.5328	0.4403	0.4147
Class B	0.7188	0.3728	0.7852	0.7674	0.7245	0.5546

Class A: 3H for HD videos
Class B: 1.5H for UHD videos

* VMAF framework performance on UHD videos,
by Jesus Gutierrez et al, VQEG Meeting 08/05/2017

Why do we need a 4K model? (Cont'd)

- VMAF v0.6.1 is trained on 1080p, for 1080p
  - Video source: mix of 4K and 1080p
  - PVS: mix of encoding resolutions 1080p, 720p, 480p etc.
  - Trained on subjective data collected on 1080p device at distance 3H
  - Mapped to score range of [0, 100]
- When applied to 4K videos:
  - Predicts quality at 60 pixels/degree - 1.5H for 4K display
  - What it does NOT capture:
    - Viewing angle (1.5H for 4K has wider viewing angle)
    - Not calibrated to viewer expectation of 4K experience

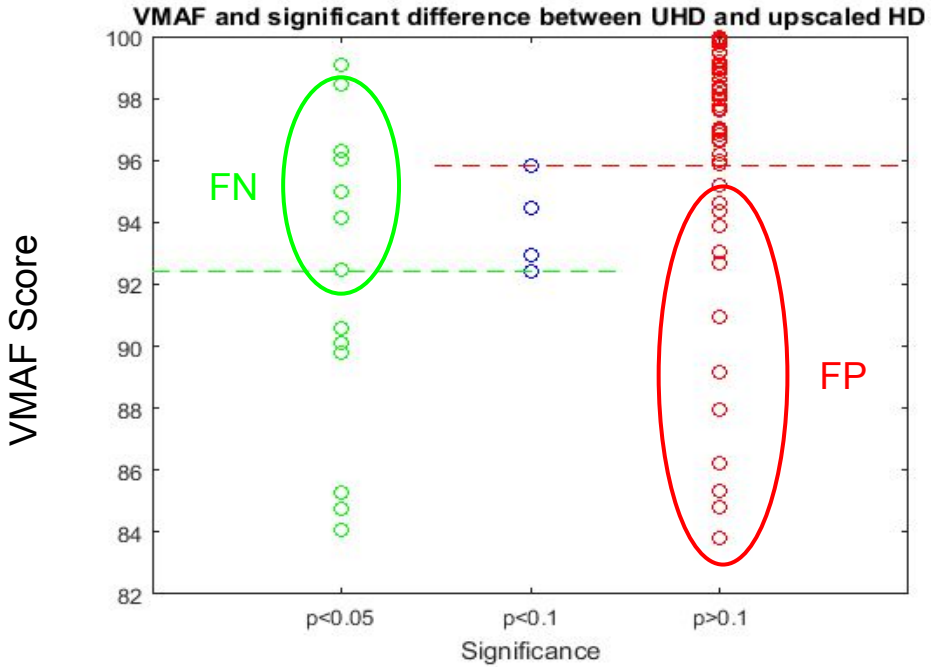
Subjective experiment setup

- Content selection
  - 8 new clips based on **false positive / false negative analysis** of v0.6.1
  - 18 old clips reused from training VMAF v0.6.1
  - In total, 26 clips, 13 are 4K source, 13 are 1080p source
- Impairment generation
  - X264 main profile, 2-sec GoP, 3 CRF values: 21, 25, 29
  - Resolutions 3840x2160, 1920x1080, 1280x720, 960x540, 640x360, 480x270, 384x216, PAR 1:1
  - Encoded video to be upscaled to 4K using bicubic before display
  - With hidden reference, total #PVS 13 * 22 + 13 * 19 = 533

Subjective experiment setup (Cont'd)

- Experiment size
  - Methodology: ACR with hidden reference
  - 533 PVS and 24 scores / PVS
  - Fit into 7 sessions, each approximately 20 minutes
- Lab setup
  - 43" 4K TV (Sony FW-43XD8001)
  - Standardized room environment
  - Controlled lighting (ITU-R BT.500-13)
  - Viewing distance 1.5H
  - Two parallel sessions (with two TVs) to speed up data collection

False positive / false negative analysis of VMAF v0.6.1



Paired Comparison on 79 videos: UHD vs. upscaled FHD

FP: subjects do not see a significant difference in quality but VMAF yields a low score (thus falsely detects some artifacts)

FN: subjects see a significant difference in quality but VMAF yields a high score (thus falsely ignores some artifacts)

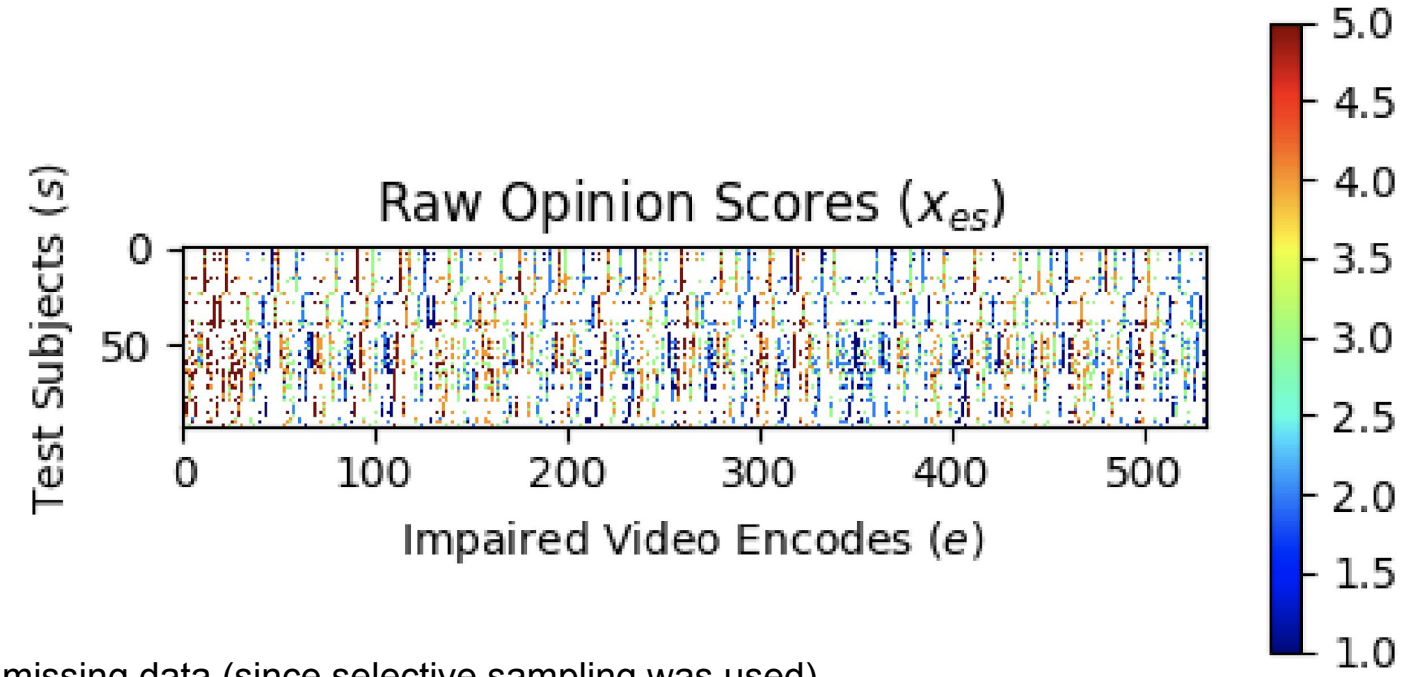
False positives

Sample images removed due to intellectual property issue

False negatives

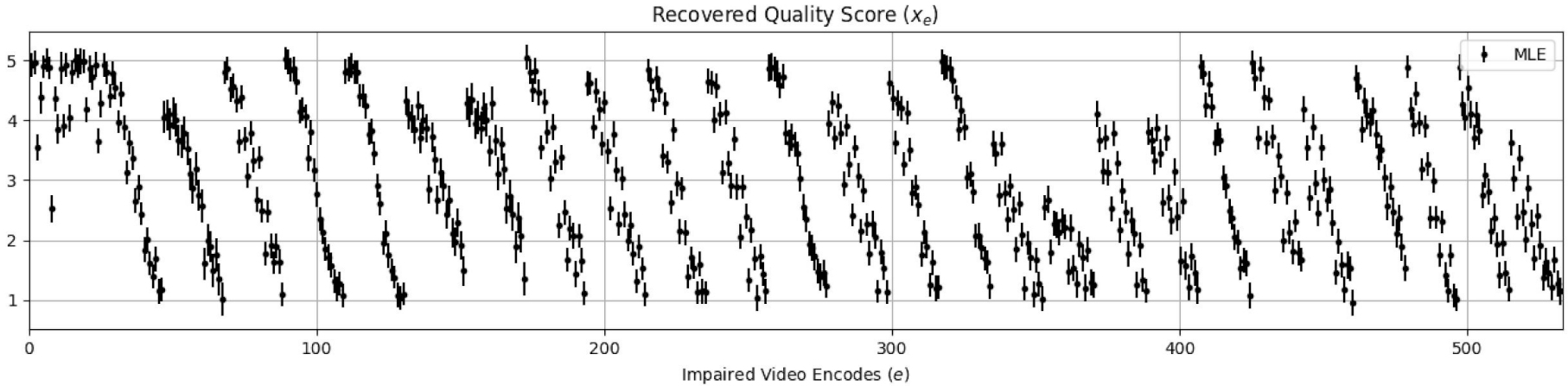
Sample images removed due to intellectual property issue

Raw opinion scores



*white: missing data (since selective sampling was used)

MOS scores recovered by maximum likelihood estimation



Cross-dataset validation (SROCC)

Training	
LIVE Video	0.686
LIVE Mobile	0.832
CSIQ-VQA	0.788
NFLX	0.828
MPEG SHVC	0.762
VQEG HD3	0.830
EPFL	0.776
VMAF v0.6.1	**0.838**
VMAF 4K	**0.841**

*Training on one dataset and test on the rest, then report the average SROCC

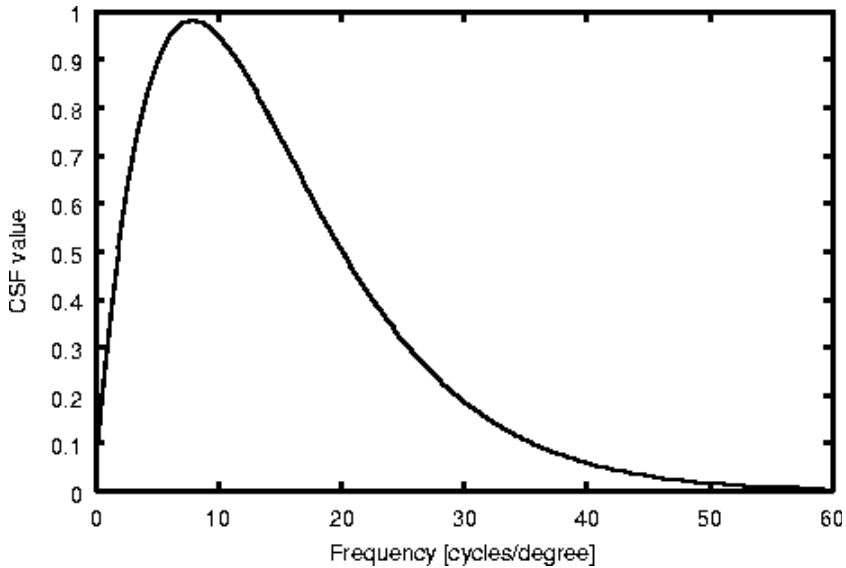
Coming up next to VMAF open-source repo

- Release of new 4K model (this talk)
- Confidence interval of VMAF model (Wednesday's talk)
- Enhanced temporal features
- New HDR model

Backup Slides

NETFLIX

Critical viewing distance per resolution



Optimal horizontal viewing angle, optimal viewing distance in picture heights (H)

Image system	Reference	Aspect ratio	Pixel aspect ratio	Optimal horizontal viewing angle	Optimal viewing distance
720 x 483	Rec. ITU-R BT.601	4:3	0.89	11°	7 *H*
640 x 480	VGA	4:3	1	11°	7 *H*
720 x 576	Rec. ITU-R BT.601	4:3	1.07	13°	6 *H*
1,024 x 768	XGA	4:3	1	17°	4.5 *H*
1,280 x 720	Rec. ITU-R BT.1543 Rec. ITU-R BT.1847	16:9	1	21°	4.8 *H*
1,400 x 1,050	SXGA+	4:3	1	23°	3.3 *H*
1,920 x 1,080	Rec. ITU-R BT.709	16:9	1	31°	3.2 *H*
3,840 x 2,160	Rec. ITU-R BT.1769	16:9	1	58°	1.6 *H*
7,680 x 4,320	Rec. ITU-R BT.1769	16:9	1	96°	0.8 *H*