

Objective and subjective quality assessment for gaming videos

Nabajeet Barman, Maria G. Martini
Kingston University, London

Introduction: Motivation

- Gaming Videos: Increasing in popularity
- Increasing number of Gaming OTT Providers: Twitch.tv, YouTube Gaming, Hitbox.tv
- Twitch.tv alone consists of approximately two million streamers, nine million daily active users
- Twitch.tv Ranked 4th in terms of total traffic during peak hours

Introduction: Motivation ctd...

- Gaming videos are different from traditional videos
- Synthetic and Artificial content
- Streaming Requirements: Real-time, CBR, 1-pass

**H.264/MPEG-AVC,
H.265/MPEG-HEVC,
VP9**

**Codec Comparison for Live Gaming Video
Streaming**

Source Sequences



(a) Counter Strike



(b) Diablo



(c) Dota2



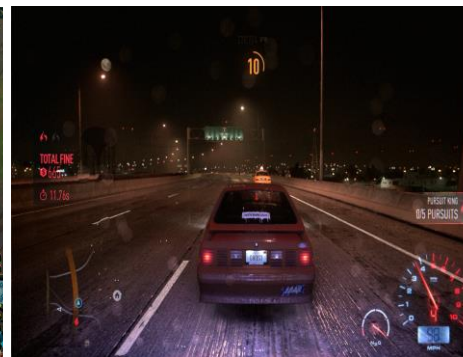
(d) FIFA16



(e) Hearthstone



(f) League of Legends



(g) Need for Speed



(h) World of Warcraft

Encoding Parameters

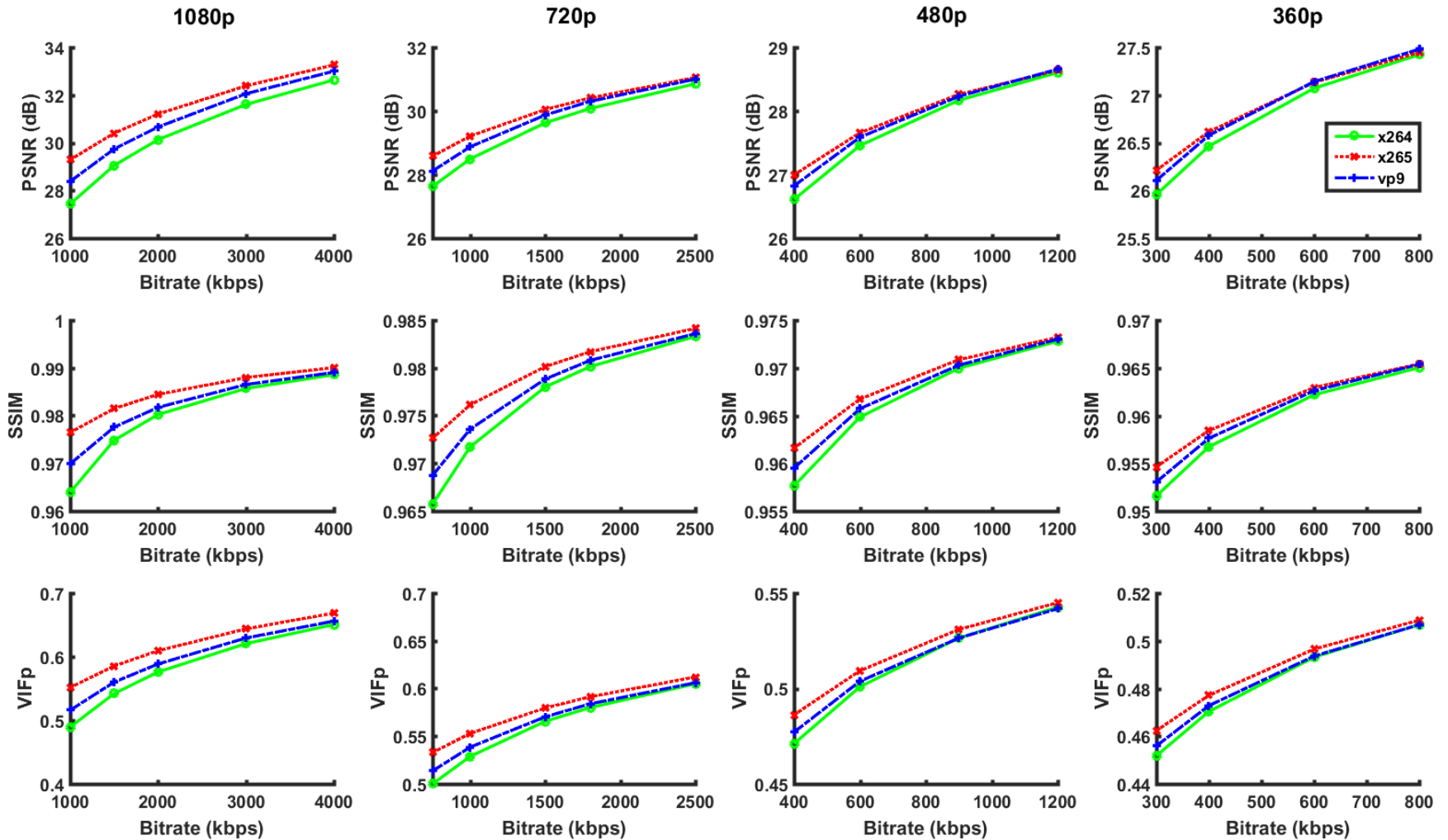
Table I: Resolution-Bitrate pairs

Resolution	Bitrate (kbps)
1920x1080 (1080p)	1000, 1500, 2000, 3000, 4000
1280x720 (720p)	750, 1000, 1500, 1800, 2500
640x480 (480p)	400, 600, 900, 1200
480x360 (360p)	300, 400, 600, 800

Table II: Encoding Settings Summary

Encoder	Settings
libx264, libx265	preset=veryfast, profile=main, level=4.0
libvpx-vp9	deadline=realtime, quality=realtime
libx264, libx265	single pass, buffer=bitrate
libvpx-vp9	closed gop=48(2s), CBR

Results: Objective Quality vs. Bitrate for three codecs for one of the sample videos (CSGO).



Results: Bjontegaard-Delta Results (PSNR-based) for all videos

Sequence/ Resolution	x265 vs. x264				vp9 vs. x264				x265 vs. vp9			
	1080p	720p	480p	360p	1080p	720p	480p	360p	1080p	720p	480p	360p
CSGO	-30.63	-20.75	-12.02	-10.66	-13.31	-6.40	-0.10	-1.27	-18.33	-14.24	-11.67	-9.18
Diablo	-36.21	-31.67	-21.36	-21.70	-14.22	-10.18	-1.76	1.49	-25.90	-23.48	-20.41	-23.38
DOTA2	-16.69	-21.38	-15.40	-13.50	21.95	19.08	19.64	19.78	-31.56	-33.49	-29.57	-28.02
FIFA16	-43.11	-40.47	-27.95	-19.26	-9.06	-10.02	-0.39	8.76	-36.58	-28.06	-23.96	-23.45
Hearthstone	-33.54	-33.55	-21.64	-21.94	14.97	9.35	17.82	18.42	-42.17	-39.28	-33.36	-34.22
LOL	-8.29	-11.07	-0.82	2.60	35.01	24.23	33.69	38.85	-32.87	-27.79	-25.82	-25.90
NFS	-30.90	-23.39	-16.78	-12.99	-9.22	-5.33	0.74	0.20	-20.04	-16.75	-16.70	-12.42
WoW	-14.17	-23.67	-15.92	-11.70	28.86	25.16	33.58	35.65	-34.94	-39.71	-37.82	-35.26
Average	-26.69	-25.74	-16.48	-13.64	6.87	5.73	12.90	15.23	-30.29	-27.85	-24.91	-23.97
Total Average	-20.64				10.18				-26.76			

Results: Encoding Run Times (in seconds) for 1080p resolution videos (total duration=10 sec)

Sequence/ Bitrate (kbps)	x264					x265					vp9				
	1000	1500	2000	3000	4000	1000	1500	2000	3000	4000	1000	1500	2000	3000	4000
CSGO	2.65	2.75	2.83	2.96	3.04	8.43	9.09	9.59	10.92	11.47	12.72	14.39	15.51	17.09	18.47
Diablo	3.05	3.05	3.20	3.36	3.33	6.95	7.60	7.18	8.82	9.46	9.85	10.48	11.38	12.64	13.34
DOTA2	4.09	4.14	4.19	4.11	4.37	6.12	6.53	6.68	7.50	7.87	9.22	10.19	10.79	11.41	12.21
FIFA16	2.90	3.03	3.18	3.32	3.09	5.76	6.26	6.42	7.20	7.98	10.49	10.99	11.43	11.63	12.35
Hearthstone	2.86	2.88	2.95	2.72	2.91	5.23	5.10	5.73	5.8	6.05	6.97	8.02	7.97	8.83	9.21
LOL	1.79	1.84	1.89	1.91	1.93	5.15	5.25	5.63	6.11	6.38	8.41	8.96	9.09	9.69	9.97
NFS	2.74	2.86	2.90	3.04	3.16	9.07	10.37	10.83	11.99	12.63	11.06	12.67	13.48	15.34	16.85
WoW	1.94	2.05	2.14	2.15	2.24	5.72	5.92	6.24	6.49	6.96	9.83	10.48	10.83	11.60	11.71
Average	2.75	2.82	2.91	2.95	3.01	6.56	7.02	7.29	8.10	8.60	9.82	10.77	11.31	12.28	13.02
Total Average	2.89					7.51					11.44				

Conclusion

In terms of BD-BR analysis using PSNR as the objective quality metric:

- H.265 performs better than both H.264 and VP9 for all videos, resolutions-bitrate pair considered
- H.264 vs. VP9 performance is highly dependent on the content type
 - For half of the videos, VP9 performs better than H.264

In terms of Encoding Time: H.264 is almost 2.6 times faster than H.265 and 4 times faster than VP9

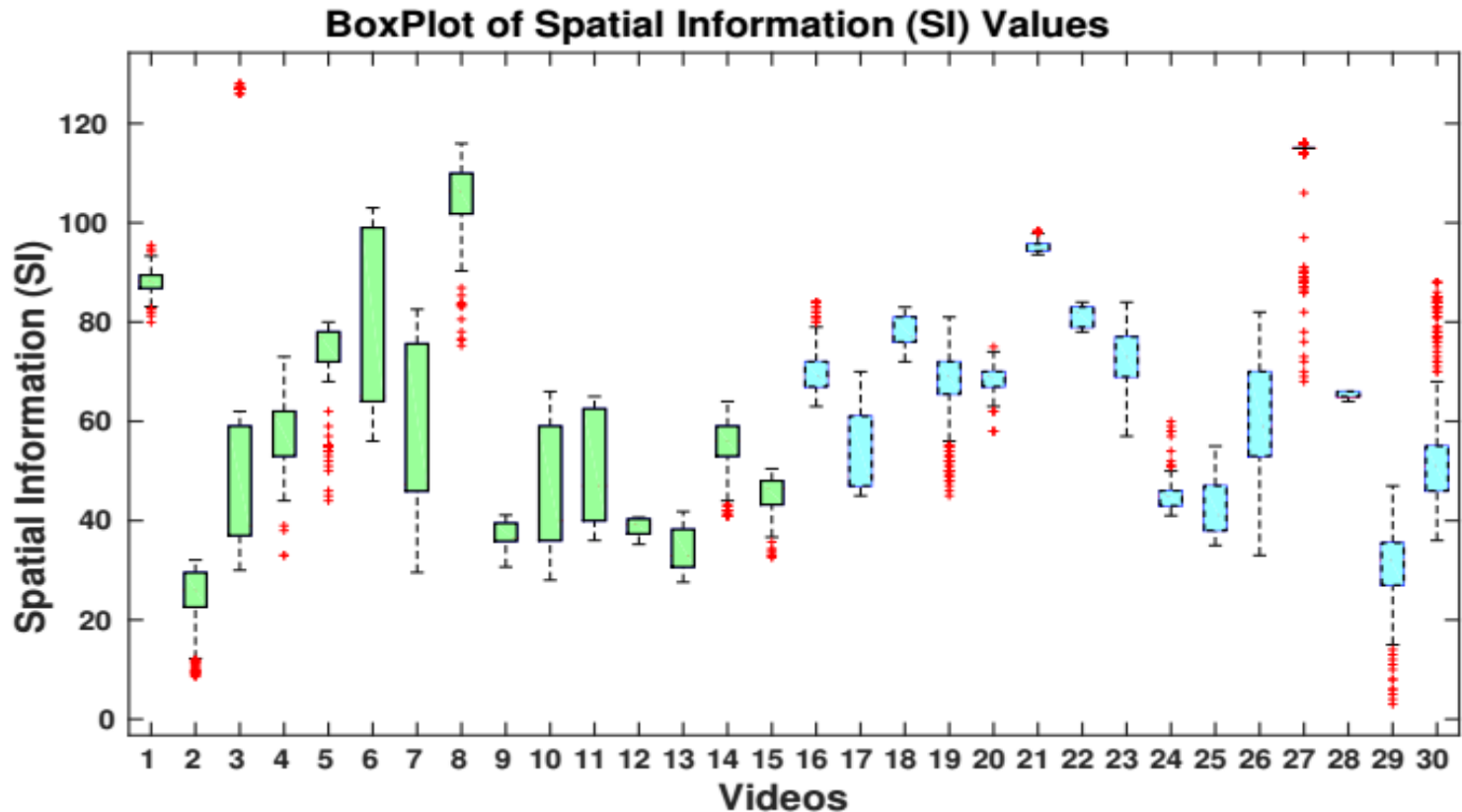
A Comparative Quality Assessment Study for Gaming and Non-Gaming Videos

N. Barman, S. Zadtootaghaj, M. G. Martini, S. Möller, and S. Lee

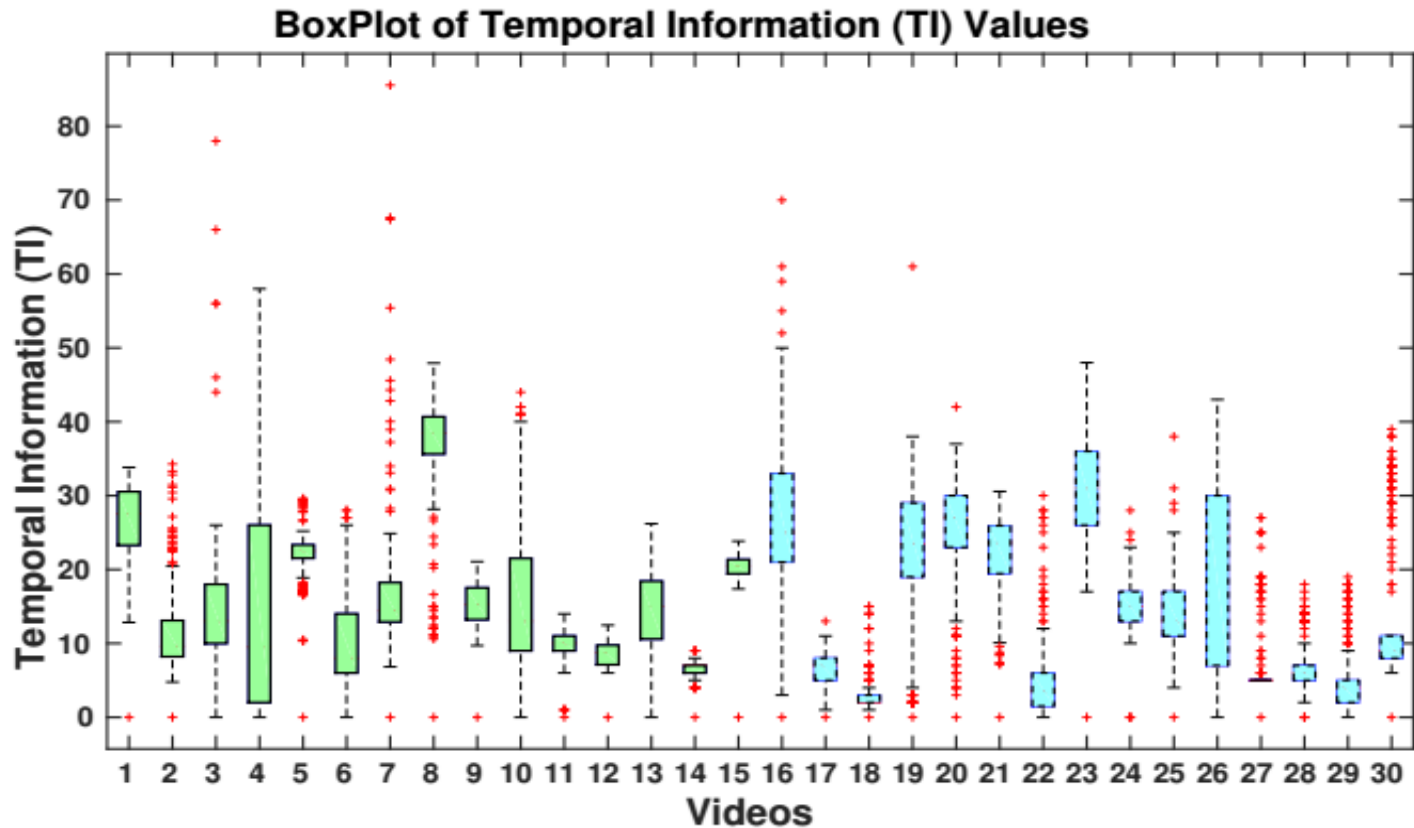
Summary of video encoding parameters

Parameter	Value
Duration	10 sec
Resolution	1080p
Frame Rate	24
Number of Reference Videos	30
Total Quality Levels	19
Number of Encoded Videos	570
Encoder	FFmpeg
Encoding Mode	CRF
Video Compression Standard	HEVC
Objective Quality Metrics	PSNR, SSIM [5], VIFP [6], VMAF [7]

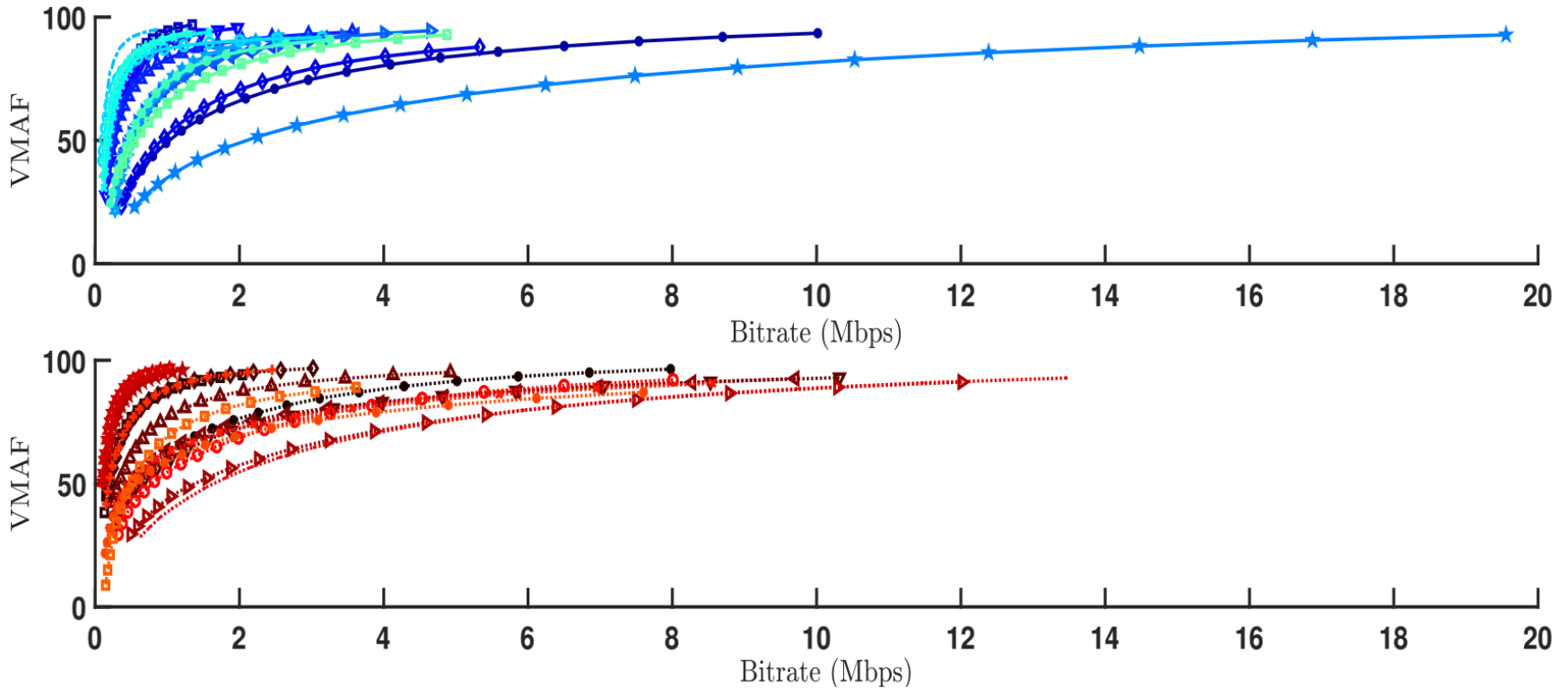
Box Plot for **SI** values of the 30 reference videos (Videos 1-15, Green: non-gaming videos; Videos 16-30, cyan / dashed boxes: gaming videos).



Box Plot for **TI** values of the 30 reference videos (Videos 1-15, Green: non-gaming videos; Videos 16-30, cyan / dashed boxes: gaming videos).



Quality (VMAF)-Bitrate curves for the 15 non-gaming videos (top) and the 15 gaming videos (bottom) encoded at 19 different CRF values



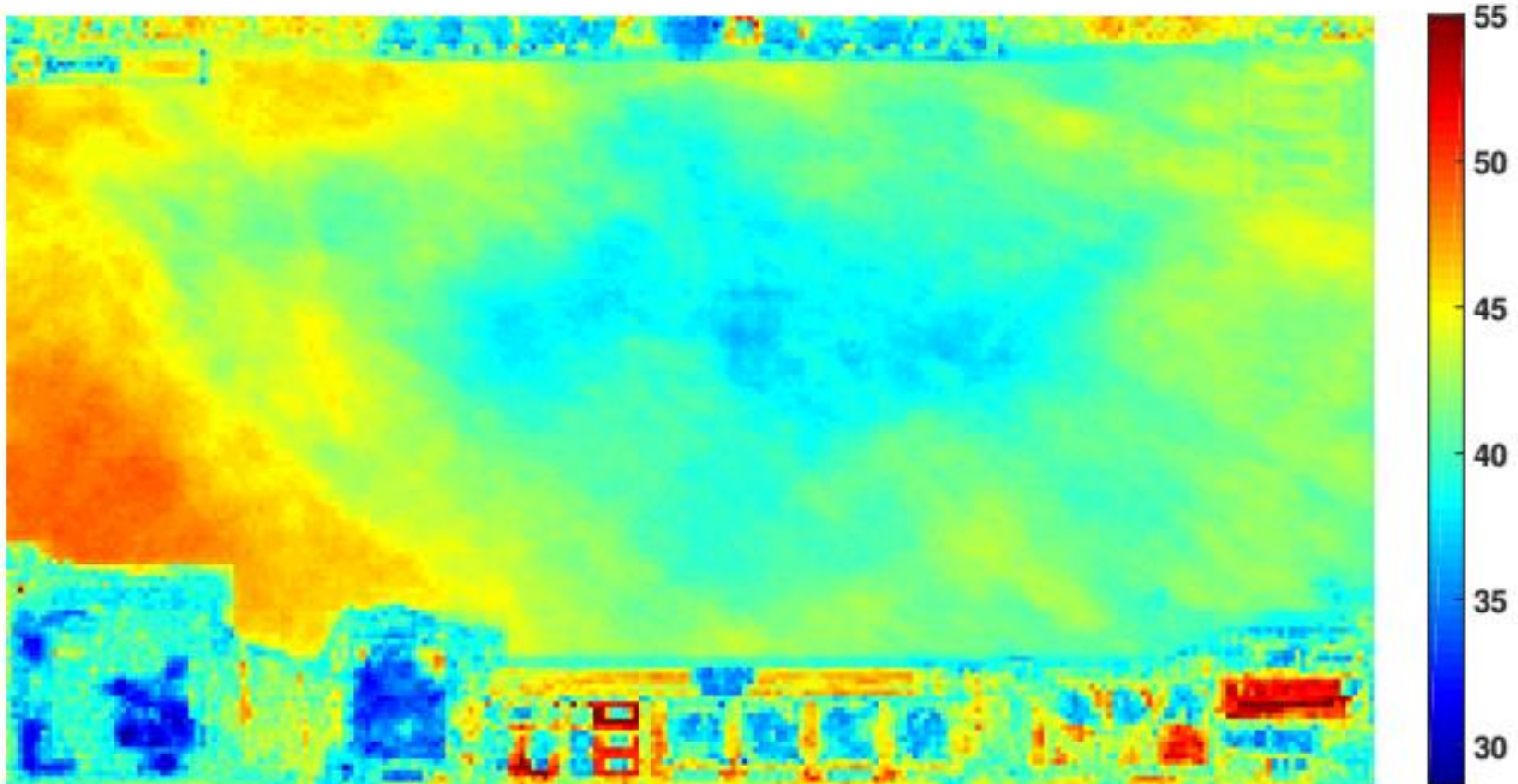
PLCC and SROCC values of four objective scores with respect to MOS scores

Objective Metrics	PLCC		SROCC	
	Non-gaming Videos	Gaming Videos	Non-gaming Videos	Gaming Videos
PSNR	0.7779	0.6980	0.7096	0.7153
SSIM	0.8189	0.5150	0.8936	0.5231
VIFP	0.8591	0.6693	0.8765	0.6567
VMAF	0.9270	0.8810	0.9422	0.8928

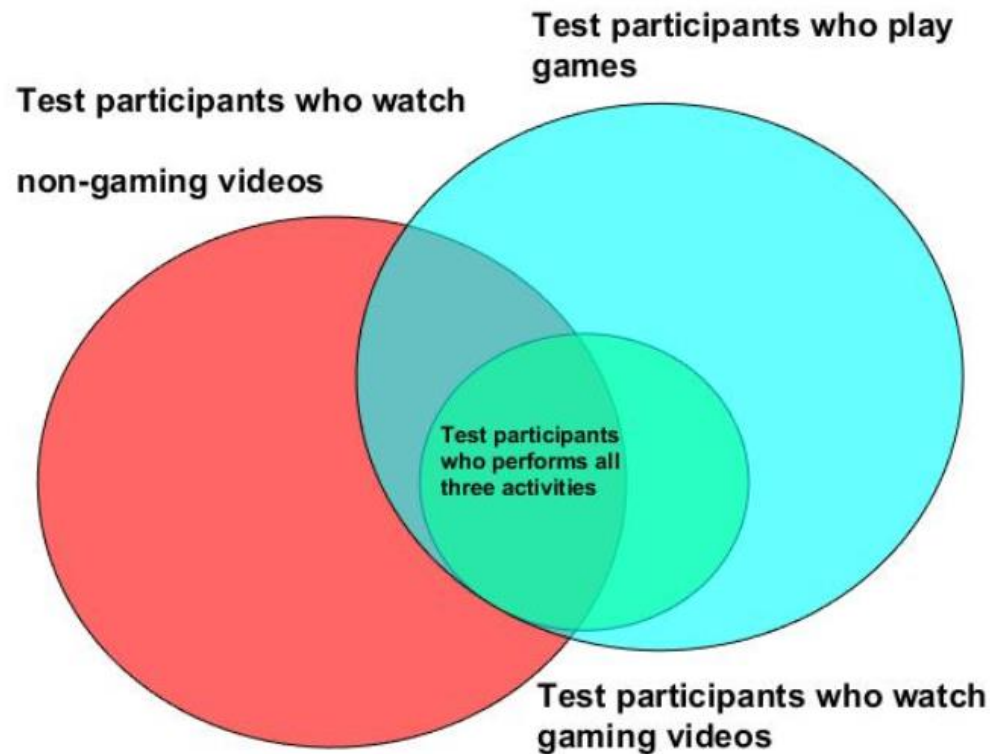
Main Observations:

- For both non-gaming and gaming videos, and in terms of both PLCC and SROCC, VMAF outperforms traditional objective metrics such as PSNR, SSIM and VIFP.
- For a given objective metric, the PLCC and SROCC values are lower for the gaming videos when compared to those of non-gaming videos. One of the possible reasons for the same could be subject bias.
- In terms of PLCC values, out of all four objective metrics, SSIM performs the worst for gaming videos compared to non-gaming videos.
- In terms of SROCC values, while PSNR exhibits the worst performance for non-gaming videos, SSIM has the worst performance for gaming videos.

The heatmap of the PPSNR values (dB) for a sample gaming video sequence, Dota 2



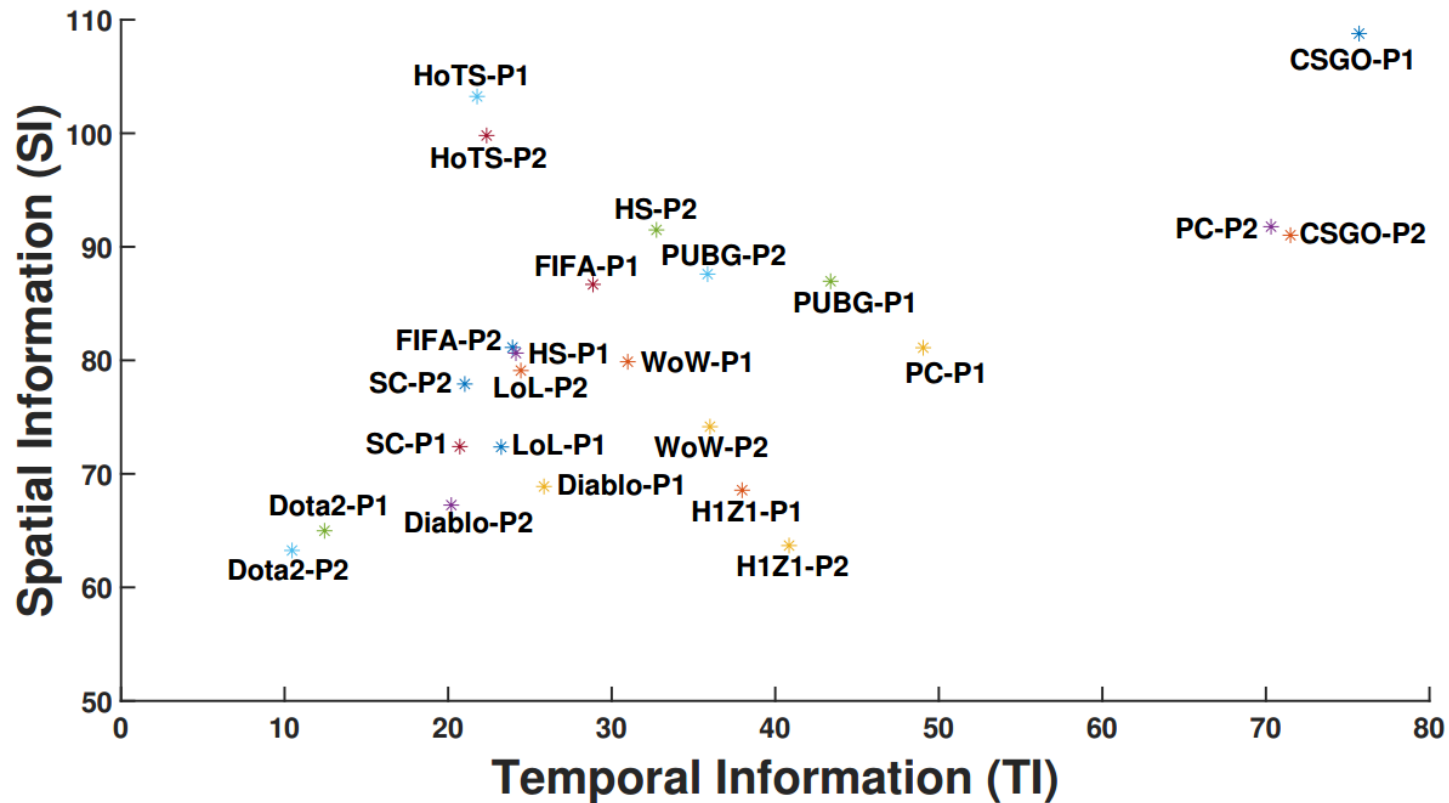
Venn Diagram illustrating the percentage of test participants interested in the three activities.



An Evaluation of Video Quality Assessment Metrics for Gaming Video Streaming

N. Barman, S. Schmidt, S. Zadtootaghaj, M. G. Martini, S. Möller

Dataset: 24 videos, 2 each from 12 different games



Video encoding settings

Parameter	Value
Duration	30 sec
Resolution	1080p, 720p, 480p
Frame Rate	30
Encoder	FFmpeg
Encoding Mode	CBR
Video Compression Standard	H.264, Main 4.0
Preset	veryfast

Resolution-bitrate pairs used to encode the reference videos. The bitrates in bold text refer to the bitrates used in the subjective quality assessment.

Resolution	Bitrate (kbps)
1080p	600, 750 , 1000, 1200 , 1500, 2000 , 3000, 4000
720p	500, 600 , 750, 900, 1200 , 1600, 2000 , 2500, 4000
480p	300, 400, 600 , 900, 1200, 2000, 4000

Comparison of the performance of the VQA metrics scores with MOS ratings in terms of PLCC and SROCC values. “All Data” refers to the combined data of all three resolution-bitrate pairs. The best performing metric is shown in bold.

Metrics		480p		720p		1080p		All Data	
		PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
FR Metrics	PSNR	0.67	0.64	0.80	0.78	0.86	0.87	0.74	0.74
	SSIM	0.57	0.43	0.81	0.78	0.86	0.90	0.80	0.80
	VMAF	0.81	0.74	0.95	0.94	0.97	0.96	0.87	0.87
RR Metrics	ST-RREDOpt	-0.61	-0.51	-0.82	-0.85	-0.79	-0.92	-0.71	-0.74
	SpEEDQA	-0.63	-0.52	-0.83	-0.87	-0.77	-0.93	-0.71	-0.75
NR Metrics	BRISQUE	-0.57	-0.48	-0.83	-0.89	-0.88	-0.91	-0.49	-0.51
	BIQI	-0.53	-0.51	-0.73	-0.72	-0.81	-0.80	-0.43	-0.46
	NIQE	-0.73	-0.74	-0.85	-0.81	-0.89	-0.90	-0.77	-0.76

Comparison of the performance of the VQA metric scores with VMAF scores in terms of PLCC and SROCC values. “All Data” refers to the combined data of all three resolution-bitrate pairs. The best performing metric is shown in bold.

Metrics		480p		720p		1080p		All Data	
		PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
FR Metrics	PSNR	0.62	0.60	0.79	0.77	0.91	0.92	0.87	0.87
	SSIM	0.56	0.56	0.68	0.70	0.80	0.83	0.70	0.74
RR Metrics	ST-RREDOpt	-0.66	-0.85	-0.74	-0.89	-0.77	-0.91	-0.53	-0.61
	SpEEDQA	-0.68	-0.88	-0.76	-0.92	-0.77	-0.93	-0.55	-0.63
NR Metrics	BRISQUE	-0.68	-0.68	-0.79	-0.79	-0.77	-0.78	-0.14	-0.14
	BIQI	-0.57	-0.54	-0.70	-0.71	-0.67	-0.68	-0.05	-0.05
	NIQE	-0.75	-0.77	-0.81	-0.81	-0.78	-0.76	-0.42	-0.42

Conclusion:

- VMAF results in the highest correlation with subjective scores followed by SSIM and NIQE.
- Many metrics failed to capture the MOS variation at lower resolutions, hence resulting in lower correlation values.
- The performance of the NR metrics decreases when considering different resolution-bitrate pairs together. Possible reasons can be the lack of proper training, gaming video content, etc.

References:

- N. Barman and M. G. Martini. H.264/MPEG-AVC, H.265/MPEG-HEVC and VP9 codec comparison for live gaming video streaming. In Ninth International Conference on Quality of Multimedia Experience (QoMEX), pages 1–6, (Erfurt, Germany) May 2017.
- N. Barman, S. Zadtootaghaj, M. G. Martini, S. Moller, and S. Lee. A Comparative Quality Assessment Study for Gaming and Non-Gaming Videos. In Tenth International Conference on Quality of Multimedia Experience (QoMEX), May 2018. Accepted.
- N. Barman, S. Schmidt, S. Zadtootaghaj, M. G. Martini, and S. Moller. An Evaluation of Video Quality Assessment Metrics for Gaming Video Streaming. Submitted

Thank you!



This presentation is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 643072.