Optimized Viewport Dependent Streaming of Stereoscopic Omnidirectional Video

D. Naik, Igor D.D. Curcio, H. Toukomaa

Bell Labs Distinguished Member of Technical Staff

Media Technology Research, Nokia Technologies, Tampere (Finland)

Email: igor.curcio@nokia.com

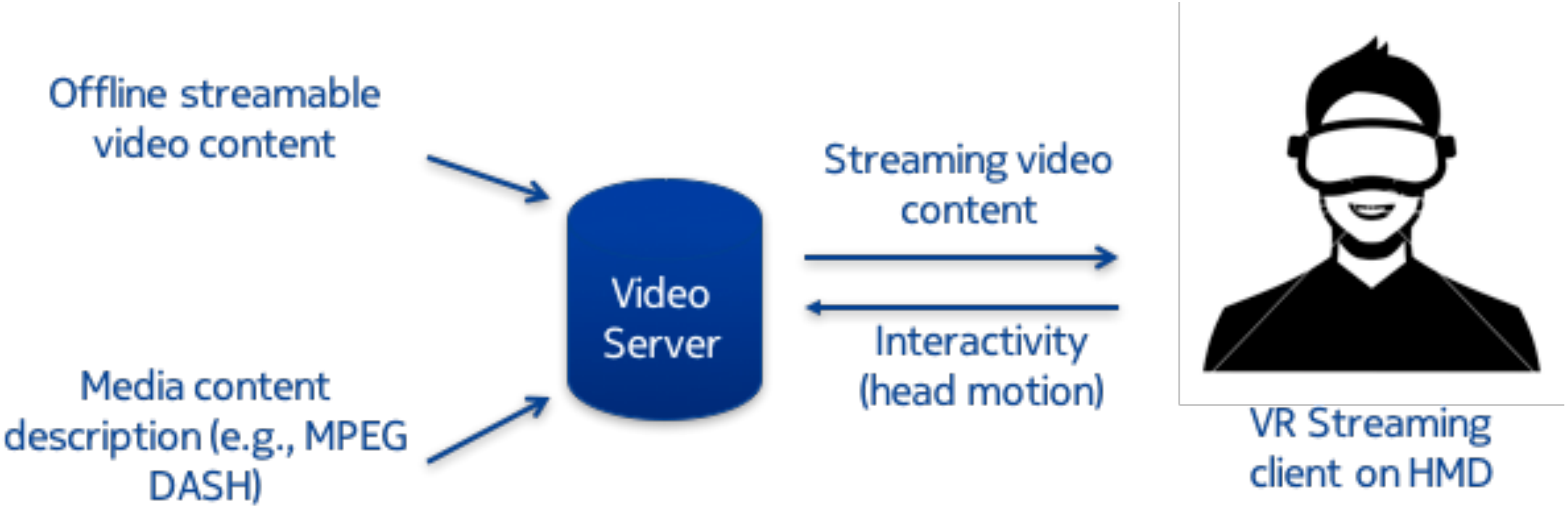
NOKIA

Packet Video Workshop (within the ACM Multimedia Systems Conference), Amsterdam, Netherlands, 12 June '18

Outline

- Viewport Dependent Streaming
- Stereoscopic Omnidirectional Video
- Two subjective quality experiments
  - Streaming rate reduction with asymmetric video quality in foreground view
  - Streaming rate reduction with asymmetric video quality in background view.

E2E system for streaming omnidirectional video



Problem

- There are new (?) challenges for 360-Degree video, compared to traditional 2D video:
  - The **bandwidth** required for streaming 360-Degree video is huge
  - ...but luckily only a ”viewport” of the whole omnidirectional space can be viewed by users at any point of time.

Viewport-Dependent Streaming

- Stream at **high quality** only what (i.e., the viewport) can be seen by the Head Mounted Display (HMD) Field of View (FoV). Stream everything else at **lower quality**.
  - Suffers from increased latency.

- This strategy allows a great bandwidth saving, compared to *viewport-independent* streaming, where the 360-degree content is streamed at high quality, with theoretical no latency at the cost of a greater bandwidth requirement.

Viewport-Dependent Streaming

- Two key parameters:
  - **Viewport size:** too large size requires more bandwidth; too small size requires too many **viewport switches** in case of head motion, and may reduce the perceived video quality
  - **Motion-to-high-quality delay**: the elapsed time between the head motion to an area outside of the viewport, and the subsequent system reaction to display a refreshed high-quality viewport on the HMD. Must the the shortest possible.

NOKIA

Viewport-Dependent Streaming

- This technology has been **standardized** in
  - MPEG: Omnidirectional Media Format (OMAF)
  - 3GPP: TS 26.118 ”3GPP Virtual Reality Profiles for Streaming Applications”
  - VR-IF: Guidelines

Stereoscopic Omnidirectional Video

- A pair of views consume a large bandwidth, compared to monoscopic video.
- It is possible to take advantage of the properties of the Human Visual System (HVS)
- **Suppression and fusion theories**: if the two eyes is given **asymmetric quality** (e.g., quantization-based or spatial resolution-based), the HVS is capable of perceiving a fused quality which is closer to the higher-quality image, partially suppressing the lower quality one.

Stereoscopic Omnidirectional Video

- For saving streaming bandwidth it is possible to apply this concept to omnidirectional video that makes use of VDS
- There are new interesting **combinations of asymmetric video quality** for
  - Left, right views in VR
  - Foreground, background views.

NOKIA

Subjective Quality Assessment Procedure

- Pre-screening of test subjects is important.
- For assessing video quality we used Absolute Category Rating with Hidden Reference (ACR-HR)
  - Sequences shown in random order
  - One of the sequences is used as hidden reference
  - Fractional scale 1..5.
  - Differential Mean Opinion Score (DMOS) for data analysis.

Subjective Quality Assessment Procedure

- Each test was not exceeding 1 hour (30 min. max wearing the HMD)
- Typical session structure:
  1. Instruction
  2. Pre- visual fatigue questionnaire
  3. Training session
  4. Viewing session(s)
  5. Post- visual fatigue questionnaire

NOKIA

Two experiments

1. Aim to reduce the streaming bit rate in the **foreground view** (visible tile(s)) by using asymmetric video quality
   - Eye dominance invariance was also assessed

2. Aim to further reduce the streaming bit rate in the **background view** (non visible tile(s)) by using asymmetric video quality

NOKIA

Settings (both experiments)

- Client with Samsung Galaxy S8 + Gear VR 2016 HMD
- E2E system was MPEG-DASH compliant with the needed VR extensions
- Streaming over WLAN 802.11 ac tri-band AP
- Equirectangular panorama video with 4K resolution (3840x1920 pixels) encoded with HEVC at 30 fps.

First experiment

- Adventure sequence (Bear) and music video clip (Kids) both 20 sec. duration. 9 test subjects.
- We wanted a score only in the foreground view (we did not want VDS effects because of head turns)
  - Tile of 180 degrees horizontal and 111 degrees vertical sizes . Removed the top and bottom tiles (34.5 degrees each).
- List of QP values for FG left and right views >

QP (Left View)	QP (Right View)
23	25
23	26
23	27
23	28
23	29
23	32
23	23
25	23
26	23
27	23
28	23
29	23
32	23

NOKIA

Second experiment

- Adventure sequence (Bear) and military sequence (Jet) both 20 sec. duration.
- Foreground tile 120 degrees and 240 degrees background tile horizontally.
  - Vertical sizes were as the first experiment.
  - No top and bottom tiles removal.
  - VDS active with 1s DASH segments.
  - No asymmetric quality for the foreground tile.
- List of QP values for foreground and background left and right views:

Foreground QP (L and R Views)	Background QP (Left View)	Background QP (Right View)
24	27	27
24	27	28
24	27	29
24	27	30

NOKIA

Results (first experiment)

- Bit rate savings for foreground tile asymmetry: 30% for 4.4 DMOS (all DMOSs are averaged considering the ”dual" case). Useful for adaptation!

QP L / R	Kids			Bear			Bit rate saving (Avg. %)	DMOS (Avg.)
	Bit rate (Mbps)	Bit rate saving %	DMOS	Bit rate (Mbps)	Bit rate saving %	DMOS		
23 / 23	5.4	0	5.0	16.9	0	5.0	0	5.0
23 / 25	4.5	17	4.5	13.1	22	4.5	20	4.5
23 / 26	4.2	22	4.4	11.9	30	4.4	26	4.4
23 / 27	4.0	26	4.5	11.2	34	4.2	30	4.4
23 / 28	3.8	30	4.2	10.8	36	4.1	33	4.2
23 / 29	3.6	33	4.1	10.4	38	4.4	36	4.3
23 / 30	3.5	35	4.0	9.9	41	4.0	38	4.0
23 / 32	3.3	39	3.5	9.6	43	3.5	41	3.5

Results (first experiment) – Eye dominance

- If the hypothesis of eye dominance were to hold good, then when right eye dominant subjects received a high quality view on their right eye, they should have consistently rated the clip with a better score, and also vice versa for left eye dominant subjects.
- Results show no clear trend, and that eye dominance was invariant for our experiment (4L and 5R eye dominant subjects). DMOS differences were not significant.

Test sequence	R eye dominant DMOS_Diff	L eye dominant DMOS_Diff
Bear	0.09	-0.13
Kids	0.07	-0.04
Average	**0.08**	**-0.09**

Results (second experiment)

- Streaming background tile(s) with asymmetric SNR video quality.
- Foreground view with symmetric quality.
- 12 subjects. Motion-to-high-quality delay approx 1 sec.
- Bit rate savings in the range 5-15%.

QP L / R	Bear			Jet			Bit rate saving (Avg. %)	DMOS (Avg.)
	Bit rate (Mbps)	Bit rate saving %	DMOS	Bit rate (Mbps)	Bit rate saving %	DMOS		
27 / 27	19.9	0	5.0	8.1	0	5.0	0	5.0
27 / 28	18.9	5	4.7	7.8	4	4.4	5	4.6
27 / 29	18.2	9	4.5	7.5	7	4.0	8	4.3
27 / 30	17.6	12	4.3	6.7	17	4.1	15	4.2

NOKIA

Conclusions

- Video SNR quality asymmetry can yield **streaming bit rate reductions of 30%** for a DMOS=4.4 in the foreground view
- ...and an **additional 5-15%** for the background view when using viewport dependent streaming of omnidirectional video.
- Such rate reductions may be used for **bandwidth adaptation** purposes.
- Their **joint usage** is advantageous for optimizing streaming bit rates (perhaps used together with other techniques)
- **Eye dominance in not relevant**, and could be neglected in practical deployments of streaming systems of stereo video.

NOKIA