Tiled-based Adaptive Streaming using MPEG-DASH

Jean Le Feuvre
LTCI, CNRS, Télécom ParisTech,
Université Paris-Saclay,
75013 Paris France
jean.lefeuvre@telecom-paristech.fr

Cyril Concolato
LTCI, CNRS, Télécom ParisTech,
Université Paris-Saclay,
75013 Paris France
cyril.concolato@telecom-paristech.fr

ABSTRACT
Interactive spatial navigation within a video is getting attention in the research community, especially in scenarios such as 360° video, video surveillance or region-of-interest viewing. With more content being delivered over the top, the complexity of spatial access to the video is increased by the need to perform bitrate adaptation to react to network variations. In this paper, we describe how spatial access can be performed in an adaptive HTTP streaming context, using tiling of the source content, MPEG-DASH and its SRD extensions. We describe a configurable implementation of these technologies, within the GPAC open-source player, allowing experimentations of different adaptation policies for tiled video content.

CCS Concepts
• Information systems~Multimedia streaming

Keywords
HTTP Adaptive Streaming; MPEG-DASH; video tiling; HEVC; spatial description.

1. INTRODUCTION
In a few years, adaptive streaming has become one of the major technologies for distributing video content to end users. The reason for this success is the ease with which a client can customize the media delivery to adjust to the network conditions (bandwidth variations), the terminal characteristics (screen size, display frequency, screen color depth, codec capabilities, speaker layout), user preferences (audio and subtitle language selection, accessibility) and user interactions with the content (full-screen view versus reduced window view, speed adjustment such as rewind or fast forward).

The MPEG Dynamic Adaptive Streaming over HTTP (DASH) standard has been developed to provide a unified way of describing these adaptations possibilities, and recent developments in this regard target the ability to describe spatial layout of synchronized videos within a presentation. These enhancements are called Spatial Relation Description (SRD) [1]. They can provide the spatial location of each video on a global grid. Videos may overlap or not, and provide a dense viewing zone or leave “holes”, depending on the content author’s needs.

While the recent focus on SRD has mainly targeted user navigation in panoramic environments, the ability to describe a video as a spatial collection of synchronized videos opens the way to a new class of adaptation algorithms, based on region of interest in the video content. A source video may be divided into N parts or tiles forming a dense grid, each of these tiled videos being encoded independently.

Interestingly enough, High Efficiency Video Coding (HEVC), the latest video standard from MPEG and ITU, offers the possibility to encode each picture as a set of independent tiles [2]. Within the existing profiles, constraints may be applied at the encoding time in such a way that each tile can be decoded independently from the other tiles for any frame of the video bitstream. The resulting set of tiles would still only require a single video decoder, which may be useful on constrained devices [3].

When regular coding algorithms are used, for instance using Advanced Video Coding (AVC|H.264), decoding a set of tiles requires running several decoders in parallel (one per tile). Additionally, in multi-tile scenarios, the resolution of the tiles typically is not changed very often, since this requires re-configuring the decoders. This is also true for HEVC constrained tiling. The result is that the client has a static set of tiles composing a video and may decide to choose the quality of each tile based on any criteria. One could for example envision that some meta-data is sent along with the videos to indicate which tile is currently of interest; or a client could adapt the quality based on where the user is looking (eye-tracking or head tracking), zooming, panning and/or clicking, which can be a common scenario in 360° videos.

In this demonstration, we will show an open-source player capable of HTTP adaptive streaming of tile-based video content using MPEG-DASH and its SRD extensions. We will demonstrate various adaptation modes, using both HEVC constrained tiles and independent AVC tile encoding. This demonstration will also show how the tools from the GPAC project [4] can be used to prepare such MPEG-DASH content with SRD information.

2. System Architecture
2.1 Overview
Most adaptive streaming deployments focus on the delivery of a single video together with one audio and/or one subtitle stream. In this specific setup, it is assumed that most of the user bandwidth is consumed by the video. Audio and subtitle streams are not considered to be competing with the available bandwidth, given their very low bitrates.

In tiled video streaming, such an approach cannot be kept. The client has to manage a set of videos, usually quite large (25 videos for a 5x5 grid). If care is not taken, these videos could compete with each other on bandwidth usage. It is well known that if multiple adaptive streaming clients compete for the same bandwidth, this can lead to oscillations [5]. In the case of tile streaming, a single client is used and its rate adaptation algorithm is aware of the global throughput and buffer levels of each video before taking any decision, thus avoiding such oscillations.

*MMSys'16, May 10-13, 2016, Klagenfurt, Austria*ACM 978-1-4503-4297-1/16/05.
http://dx.doi.org/10.1145/2910017.2910641

In the DASH client implementation of GPAC, the rate adaptation logic has been divided into three steps:

- the first step, called *tile priority setup,* follows a set of pre-defined rules for attributing quality ranking to each tile in the session;
- the second step, called *rate allocation,* gathers all network throughput information and tiles bitrates, and using each tile quality ranking, attempts at maximizing the quality of the video;
- the third step, called *rate adaptation*, performs usual rate adaptation algorithm(s) based on playback speed, number of quality switches, buffer levels, and so on.

This separation in three steps has the advantage of keeping the rate adaptation algorithm untouched when using a regular non-tiled DASH session. We describe hereafter the first two steps.

2.2 Tile Priority Setup

When loading a Media Presentation Description (MPD) with SRD information, the DASH client engine first identifies each tile sets using SRD descriptors. It then identifies any coding dependencies between the tiles, in particular when HEVC tiles are used. Finally, it requests to the media renderer one video object for each independently coded tile, informing it of the SRD information. The renderer then performs the final layout of the video objects after adjusting the SRD information based on desired display size.

Once the tile sets are identified, the client allocates priorities per tile. This stage is invoked every time rate adaptation has to be performed in order to allow dynamic changes of the tile priorities. The current implementation however only supports static description of the tile priorities, through a configuration file. Priorities rank from 0 (highest priority) to maximum number of priorities achievable with the desired scheme. The priorities can currently be assigned uniformly to all videos, or with column-based priority, row-based priority, center-based or edge-based. Examples of such priority assignments for a 5x5 grid are given in Figure 1.

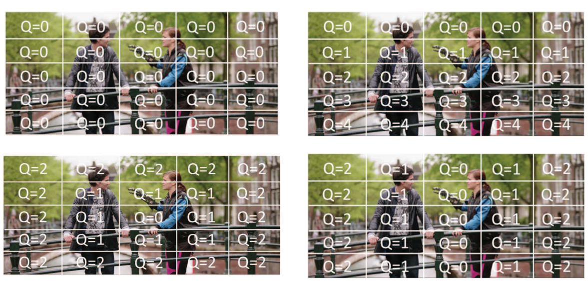

Figure 1 – Examples of priority assignment (from left to right and top to bottom): uniform, top-bottom row-based, center-based priority assignment, middle-column based priority assignment

2.3 Rate Allocation

The rate allocation step is in charge of distributing the available download bandwidth across the different tiles. It first needs to estimate the available bandwidth. This estimation has to be performed differently in tiled scenarios that in non-tiled one. In particular, during the duration of a segment, instead of downloading one video segment, the client has to download several tile video segments. These downloads could be made in parallel, but in practice servers do not support many parallel TCP connections. As a result, in our player, downloads are sequential. This could be reevaluated in future work in the context of HTTP/2 using priorities. Given this sequential approach, the bandwidth can be estimated either based on a single tile download or on the average of all tiles.

Once the available bandwidth is estimated, the rate allocation classifies all tiles based on their priority, starting with all tiles at their lowest bitrate; it then increases the quality of tiles with priority 0 so that the cumulated rate fits in the available bandwidth; if there is still bandwidth available after that step, the same step is performed with the tiles of priority 1, then 2 and so on. This ensures that at least all the highest priority tiles always get higher bitrate, but does not enforce all tiles with the same priority to have the same bitrate. Once the bitrate for each tile is selected, the regular rate adaptation algorithm is called for each tile with a target bandwidth equal to the selected bitrate.

3. Experimental Results

The following shows a 3x3 tiling using independent HEVC tiles, each tile being encoded with bitrates of (average) 100, 200, 300 to 600 and 1600 to 3000 kbps using the BQTerrace sequence. A single decoder is used to reconstruct the bitrate. The bitstream was encoded with a modified version of the HM HEVC reference software, and the packaging and DASH creation was done using MP4Box from GPAC. The tile adaptation modes shown here are uniform priority (Figure 2), center-based priority (Figure 3) and middle-row priority (Figure 4), tile being sorted in their display positions starting from top-left in raster scan order.

The duration of each DASH segment is 10 seconds, for a sequence duration of 60s, and the tables show the average bandwidth and selected rate of each tile in the video at each segment. Note that the average bitrate may vary across tiles, depending on the encoding complexity of the tile content. Access network was done over Wi-Fi in an ADSL home network, the content being hosted on GPAC download site; as can be seen in the figures below, the bandwidth is not really stable, varying between 6 and 12 mbps. The first three segments correspond to the initial buffering time, the player being configured to start with low quality to speed up the startup.

time (sec)	rate (kbps)	#1	#2	#3	#4	#5	#6	#7	#8	#9
0	10697	33	56	50	63	87	42	93	77	109
10	11104	62	129	116	133	207	102	197	181	277
20	7574	167	437	368	417	671	305	606	650	761
30	10761	1150	437	1895	417	671	305	606	650	761
40	10620	1150	2656	1895	417	671	1735	606	650	761
50	10348	1150	2656	368	417	671	1735	606	650	761
60	6035	1150	2656	1895	417	671	305	606	650	761

Figure 2 - Download statistics with uniform priority

time (sec)	rate (kbps)	#1	#2	#3	#4	#5	#6	#7	#8	#9
0	8507	33	56	50	63	87	42	93	77	109
10	6358	62	129	116	133	207	102	197	181	277
20	8559	167	437	368	417	671	305	606	181	277
30	10944	1150	437	368	417	3181	305	606	650	761
40	9879	1150	2656	368	417	3181	305	606	650	761
50	9356	1150	437	1895	417	3181	305	606	650	761
60	11117	1150	437	368	417	3181	1735	606	650	761

Figure 3 - Download statistics with center based priority

time (sec)	rate (kbps)	#1	#2	#3	#4	#5	#6	#7	#8	#9
0	5994	33	56	50	63	87	42	93	77	109
10	17496	62	129	116	133	207	102	197	181	109
20	9887	167	437	368	417	671	305	606	650	277
30	8856	167	437	368	2611	3181	1735	606	181	277
40	10763	167	129	116	2611	3181	1735	197	181	277
50	11433	167	437	368	2611	3181	1735	606	650	761
60	10139	167	437	368	2611	3181	1735	606	650	761

Figure 4 - Download statistics with middle row based priority

4. Discussion and future work

As can be seen from the selected rates, in uniform bandwidth the first tiles get higher priorities, as they are the first to be tested in the tile rate allocation step. In center mode, the middle tile (#5) gets priority over the other and plays at its maximum quality while the other tiles may be elected to higher quality depending in the bandwidth left. Similarly, for the middle-row based approach, the tiles #4, #5 and #6 play in their highest quality as requested.

As with any initial work, there is a lot of room for further improvements of the algorithm, and our future work will concentrate on interactive selection of the tiles for 360° videos, together with an improved bandwidth sharing for tiles with the same priority. We will also investigate to which extend this technique is acceptable to the end user, especially because when mixing low and high bitrates for tiles, artifacts become more visible at tile edges. Limiting those artifacts may need to be a new parameter in the rate adaptation logic. As said previously, we will also investigate the benefit of using HTTP/2 for such scenario, since we believe parallel download of the tiles may lead to a better bandwidth usage.

We encourage the adaptive streaming research community to use this work for evaluation and algorithms improvements, and are happy to make the source code of the tile player (which we believe is the only one publicly available) and test sequences available on GPAC web site at http://gpac.io/2016/04/04/srd/.

5. Conclusion

In this article, we demonstrate a tile-based video adaptation player using the open-source GPAC framework, with configurable adaptation policies. The tiles can either be independent videos, requiring multiple decoders, or motion-constrained HEVC tiles needing only a single decoder. We hope this public work can be helpful to the research community and will in the future further develop this technique for 360° video viewing.

6. Acknowledgments

This work is supported by the Live360TV project funded by the French government.

7. REFERENCES

[1] Niamut, O., Thomas, E., Gomez, D., Concolato, C., Denoual, F. and Lim, S.Y, 2016. MPEG DASH SRD - Spatial Relationship Description. In *Proceedings of the ACM MMSys* (Klagenfurt, Austria, May 10-13, 2016).

[2] Misra, K., Segall, A., Horowitz, M., Shilin, Xu, Fuldseth, A. and Minhua, Zhou, 2013. An Overview of Tiles in HEVC. *IEEE Journal of Selected Topics in Signal Processing*, 7, 6 (Dec. 2013), 969-977. DOI=10.1109/JSTSP.2013.2271451

[3] Sanchez, Y., Globisch, R., Schierl, T. and Wiegand, T. 2014. Low complexity cloud-video-mixing using HEVC. In *Proceedings of the Consumer Communications and Networking Conference* (CCNC), (Jan. 10-13, 2014), DOI=10.1109/CCNC.2014.6866573

[4] Le Feuvre, J., Concolato, C., Dufourd, J.-C., Bouqueau R. and Moissinac, J.-C. 2011. Experimenting with multimedia advances using GPAC. In *Proceedings of the 19th ACM international conference on Multimedia* (MM '11). DOI= 10.1145/2072298.2072427

[5] Li, Z., Zhu X., Gahm J., Pan R., Hu H., Begen A. and Oran D. Probe and Adapt: Rate Adaptation for HTTP Video Streaming At Scale. 2013