**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 N15819**

**October 2015, Geneva, Switzerland**

|  |  |
| --- | --- |
| **Source** | **TNO, Canon Research Centre France** |
| **Status** | **Approved** |
| **Title** | **White paper on Spatial Relationship Description in MPEG DASH part 1 AMD2** |
| **Author** | M. Oskar van Deventer, Emmanuel Thomas (TNO), Frédéric Mazé, Franck Denoual (Canon Research Centre France) |

# Introduction

The advent of ultra-high resolution video (4K, 8K and beyond) in combination with an increasing heterogeneity of display devices (UHDTV, tablet device, smartphone, smartwatch) introduces opportunities for new usages in video streaming; for instance interactive pan and zoom features. However, streaming of high resolution video over today’s networks raises problems due to bandwidth restrictions in the access and home networks. Also the video decoders of display devices may be unable to handle ultra-high resolutions, given their often limited hardware capabilities. A solution to both issues is found in streaming only spatial sub-parts of a video to the display device, in combination with the multi-rate streaming that is intrinsically supported by MPEG DASH. This way bandwidth requirement are lower which makes streaming parts of the content in native resolution possible.

This document provides a brief overview of the Spatial Representation Description (SRD) feature of the second amendment of part 1 of the MPEG DASH standard, 23009-1:2014. The feature extends the Media Presentation Description (MPD) of MPEG DASH by describing spatial relationships between associated pieces of video content. This enables the DASH client to select and retrieve only those video streams at those resolutions that are relevant to the user experience.

# Scenarios

A simple scenario is a zoom feature in a sport event, see Figure 1. In this scenario two videos are provided. The first one is a full HD panorama of the scene. The other represents a zoomed-in version of the central part of the scene encoded in a full HD resolution too, see Figure 1. In the top picture, the grid represents the coordinate system that allows to describe the spatial relation between the two videos.

An advanced scenario is spatial resolution enhancement, see Figure 2. This scenario enables a content author to describe a program as depicted in Figure 2. With SRD descriptors, the client has the choice between different representations for the same spatial part of the full panorama. The spatial parts are available in two different granularities for a given resolution level.

The SRD feature offers large flexibility. Different spatial representations may have different resolutions, frame rates and codecs. Spatial representations may partly or fully overlap and they may have arbitrary different sizes. More scenarios and worked-out SRD examples can be found in an upcoming version of MPEG DASH Part 3 [2]. In addition, several companies and organisations have been showcasing this concept of zoomable videos also called tiled streaming. A non-exhaustive list of such demonstrations include the FP7 FascinatE project [3], the EXPERIMEDIA FP7 project, the TNO-BBC R&C co-experiment at the 2014 Commonwealth Games [4].

|  |  |
| --- | --- |
| Full panorama of 1920x1080 resolution |  |
| Video of 1920x1080 resolution  (Zoom from above video) |

Figure 1. Full panorama with zoom feature

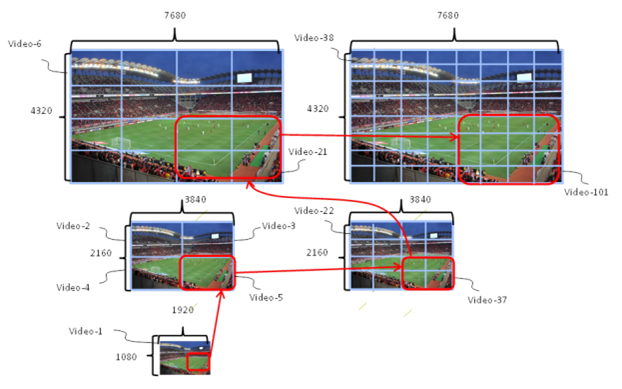


Figure 2. Panning and zooming at multiple resolution.

# Spatial relationship between videos

The SRD feature may express spatial relationship such that one video stream represents a spatial part of another full-frame video. That spatial part may be a tile or a region of interest described in terms of position and size. This description provides additional choices for the client in terms of adaptation. An SRD-aware DASH client can use the SRD annotations to select a full-frame representation or a spatial part of this full-frame that better fits its needs. This can save bandwidth and client-side computations, e.g. avoiding full-frame fetching, decoding and cropping. Alternatively, this can increase the perceived quality of a spatial part of this full-frame after a zoom, e.g. by sending spatial part with higher quality at same given bitrate.

A spatial relationship is signalled via the DASH generic descriptors in the MPD in terms of the x,y coordinates of the top-left corner of the spatial representation (e.g. the zoomed video), its width and height, and the width and height of the reference space (e.g. that of the full video) that contains the spatial representation. Details of this scheme is described in 23009-1:2014 Amendment 2, Clause 5.8.5.7 [1].

# Conclusion

Spatial Representation Description (SRD) is a new feature of MPEG DASH. It enables flexible panning and zooming in video streams, while retaining the streaming bandwidth and display device requirements limited.

References

1. MPEG DASH, “Part 1: Media presentation description and segment formats / Amendment 2: Spatial Relationship Description, Generalized URL parameters and other extensions”, draft specification ISO-IEC\_23009-1\_A2\_(E), 2015.
2. MPEG DASH, “Part 3: Implementation Guidelines / AMD1: Alignment with ISO/IEC 23009-1:2014 and extensibility”, ISO-IEC\_TR\_23009-3\_A1\_(E), 2015.
3. FP7 FascinatE project, videos, <http://www.fascinate-project.eu/index.php/gallery/videos/>
4. EXPERIMEDIA FP7 project, <http://www.experimedia.eu/2014/02/20/icacot/>
5. Judith A. Redi, Lucia D’Acunto, Omar Niamut, “Interactive UHDTV at the Commonwealth Games – An Explorative Evaluation”, TVX 2015, Brussels, http://tvx2015.com/program/full-and-short-papers/#S22