ACM MHV 2023

VVC for Immersive Video Streaming

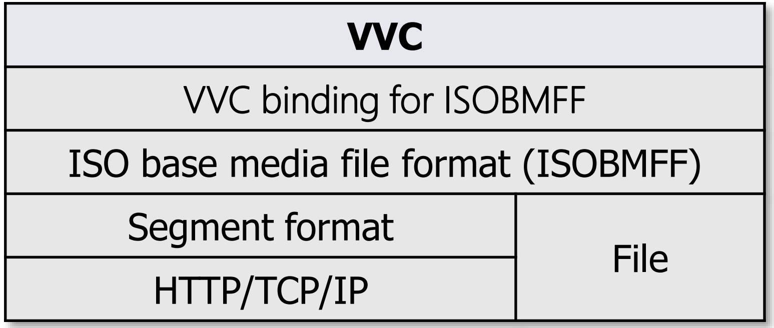
Miska Hannuksela (Nokia)
Sachin Deshpande (Sharp)

Agenda

- Introduction
- VVC Features for Immersive Video
  - Wrap-around motion compensation
  - Virtual boundaries
  - Subpictures
  - Multilayer coding
- Storage of VVC in ISOBMFF
  - VVC Tracks
- Streaming of 360° VVC video
  - Viewport-independent streaming
  - Viewport-dependent streaming
- Conclusion

Introduction

- Immersive video includes:
  - 360° video providing three degrees of freedom
  - Volumetric video where the user can move within the video and take any viewing direction
- Versatile Video Coding (VVC) standard was built considering 360° video use cases
- The omnidirectional media format (OMAF) specifies 360° video metadata for ISOBMFF storage and streaming using the Dynamic Adaptive Streaming over HTTP (DASH) and MMT standard
- Our paper provides overview of ISOBMFF Storage of VVC and its use for immersive video streaming including VVC-based OMAF video profiles

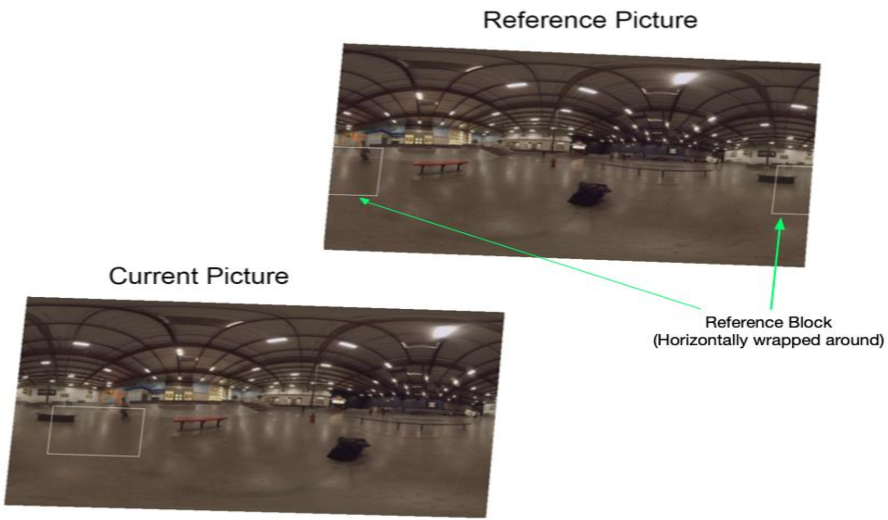


VVC Features for Immersive Video

- Wrap-around motion compensation
- Virtual boundaries
- Subpictures
- Multilayer coding

VVC : Wrap-around Motion Compensation

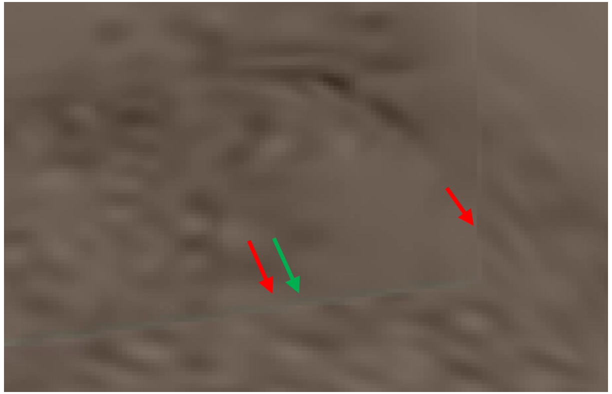
- Horizontal wrap-around MC : helpful for the equirectangular projection of 360° video
- The out-of-picture samples for the left picture boundary can be obtained from the corresponding spherical neighboring samples which are located inside the right picture boundary, and vice-a-versa.
- Benefits:
  - Better coding efficiency (~ 0.22% BDBR PSNR improvement)
  - Better subjective quality: reduced seam artifacts



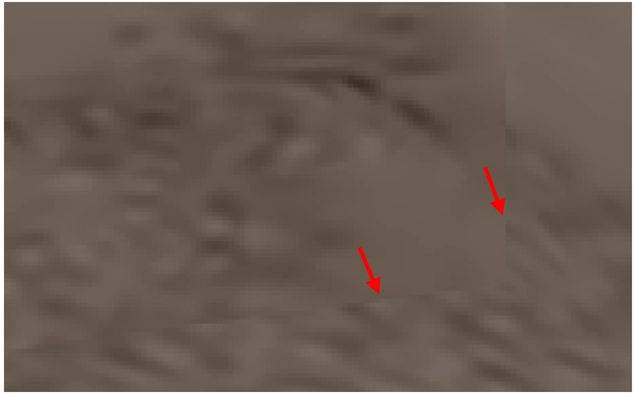
VVC : Virtual Boundaries

- Discontinuities: cubemap projection, stereoscopic frame packing
- In-loop filtering typically results in leaking artifacts at the face discontinuities
- Solution: Disabling in-loop filtering along signaled virtual boundaries
  - Position of virtual boundaries can be selected at a finer spatial granularity than the boundaries of subpictures and slices.
  - Intra prediction and context adaptation in entropy decoding operates across virtual boundaries but not across subpicture and slice boundaries
- Up to 3 vertical and 3 horizontal virtual boundaries

without VB: seam and leaking artifact



with VB: seam artifact, leaking artifact is removed



Leaking Artifact: connected areas with irrelevant content (↘)

Seam Artifact: disconnected areas with relevant content (↘)

VVC : Subpictures (1/3)

- Motion Constrained Tiles Sets (MCTS) – predecessors of VVC subpictures
- Subpictures improve MCTSs:
  - The boundaries of independent subpictures are treated like picture boundaries (~ 4 to 6.5% BDBR PSNR gain)
  - Complex encoder constraints like those for MCTS are avoided (as boundaries treated like picture boundaries)
  - Merging subpictures into a conforming VVC bitstream does not require rewriting of coded video data, such as slice header (only rewriting of parameter sets)
  - Subpictures from different picture types can be merged into a single bitstream, which enables refreshing some parts of a picture with intra random access point (IRAP) subpictures

VVC : Subpictures (2/3)

	HEVC: motion-constrained tile set (MCTS)	VVC: subpicture
Description	Encoder avoids prediction outside MCTS	Subpicture = 1..N slices. Encoder/decoder treats subpicture boundaries like picture boundaries.
Operations for "tile" merging	Parameter set and slice header rewriting	Parameter set rewriting
Response to viewport change	Full-picture refresh	Subpicture-wise refresh

Bitrate reduction of subpictures vs. MCTSs (%-point of Bjontegaard delta bitrate, luma PSNR) [JVET-N0032]	
tile size 768x768	4.13
tile size 384x384	6.46

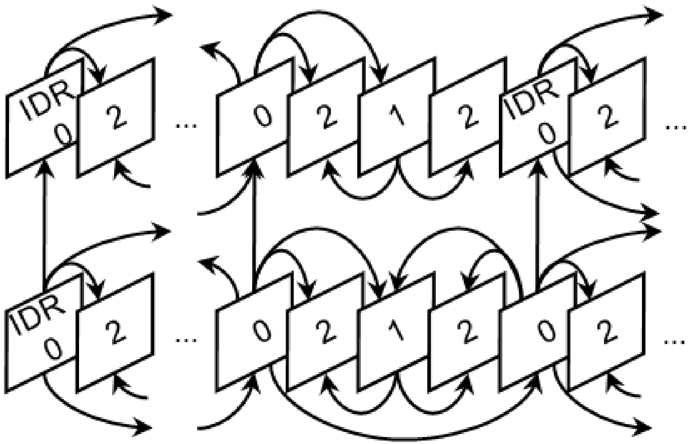
VVC : Subpictures (3/3)

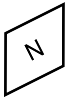
- Subpicture Use Cases:
  - Stereoscopic side-by-side or top-bottom frame packing, with each view its own independent subpicture
  - Set of continuous projection planes, such as a row of cube faces in a typical 3×2 cubemap arrangement
  - Viewport-dependent streaming
  - Texture and geometry atlases of MPEG immersive video (MIV) as separate independent subpictures
  - Subpictures for personalization and accessibility in broadcast services, with user capability to select only one view of a mosaic video to be decoded and displayed or to create a mosaic video bitstream from several views of the same broadcast.
  - Picture-in-picture : sign language video to replace main video region

VVC : Multilayer Coding (1/2)

- **No additional coding tools**, since inter-layer prediction uses reference picture resampling
- **Only selected pictures used for inter-layer prediction**
  - Intra random access point pictures or certain temporal sublayers only
  - Decoder only receives the required reference layer pictures
- **Reduced complexity and memory usage**
- Can be used for **any type of scalability**, e.g.:
  - Spatial and quality scalability
  - Stereoscopic and multiview 3D video coding

Multilayer bitstream example





picture at temporal sublayer N

VVC : Multilayer Coding (2/2)

- Multilayer coding for immersive video:
  - Each view of stereoscopic video can be coded as a layer. A dependent view can be predicted from the independent view to improve compression efficiency.
  - An independent view represents 360° video at a basic resolution and quality. Additionally, several dependent layers each enclosing a different viewport can be encoded at an enhanced resolution and/or quality.
  - Texture and depth can be coded as separate layers in a single VVC bitstream, which can be decoded with a single decoder instance, storage and/or delivery without systems level synchronization between texture and depth, and with improved compression (due to motion vector prediction from texture to depth).

Storage of VVC in ISOBMFF

- VVC Tracks

ISOBMFF: VVC Tracks

- **A VVC subpicture track** is typically used to carry a sequence of subpictures.
  - When several VVC subpicture tracks of different bitrates are made available for streaming, a player can select which ones are requested for transmission, for example based on the viewing orientation for 360° video.
- **A substitute subpicture track** is a VVC subpicture track that contains content not intended for displaying but can be used if the subpictures would not otherwise fill in the picture area to make the merged VVC bitstream compliant.
  - A box header flag in the decoder configuration box in the sample entry of a VVC subpicture track is used to identify a substitute subpicture track.
- **A VVC merge base track** guides how subpictures from several VVC subpicture tracks are merged into a VVC bitstream.
  - VVC merge base tracks were designed to avoid any awareness of VVC bitstream syntax in subpicture merging
- **A VVC extraction base track** references a VVC track and defines which ones of the subpictures of the VVC track are used to form a VVC bitstream.
  - A VVC extraction base track can be used is to extract one view of a mosaic video without parsing the VVC bitstream.
- **A VVC non-video-coding-layer (non-VCL) track** carries parameter sets and other NAL units that do not contain coded video data.
  - Can carry adaptation parameter sets that apply to some but not all VVC subpicture tracks.

OMAF

- Streaming of 360° VVC video
- Viewport-independent Streaming
- Viewport-dependent Streaming

OMAF: Streaming of 360° VVC video

- Viewport-independent streaming
- Viewport-dependent streaming

Table 1: Summary of the VVC-based OMAF video profiles

OMAF video profile	VVC-based viewport-independent	VVC-based simple tiling
Codec profile	VVC Main 10	VVC Main 10
Bit depth	≤10 bits	≤10 bits
Projection formats	Equirectangular	Equirectangular and cubemap
Viewport-dependent streaming	Not supported	Viewport-specific streams and subpicture-based
DASH segment formats	Conventional Media Segments	Conventional Media Segments, or Tile Index and Tile Data Segments

OMAF : Viewport-independent Streaming

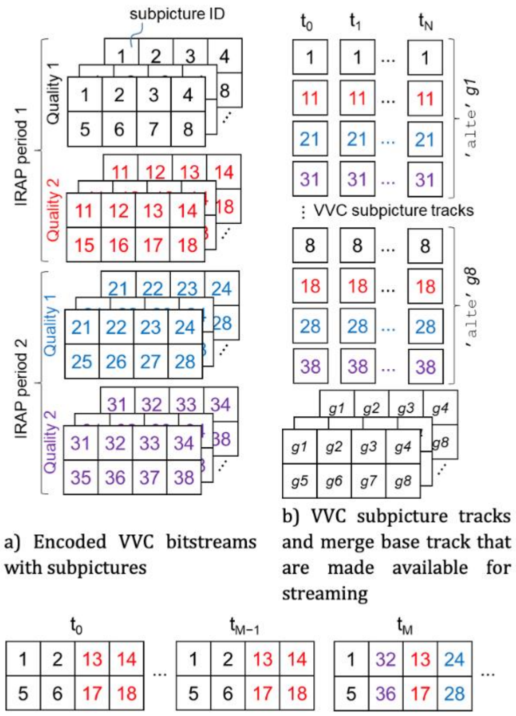
- The video pictures are encoded with similar picture quality over all spatial parts as a single-layer bitstream using temporal inter prediction
- The entire coded bitstream (possibly different bitrate versions) is stored on a server
- A player selects a suitable version for its prevailing network conditions and display resolution
- Selected version is typically completely transmitted to the player, fully decoded by the decoder, and the area of the decoded picture corresponding to the current viewport is rendered to the user.
- The VVC-based viewport-independent OMAF video profile defined in OMAF 3rd edition
  - Supports both monoscopic and stereoscopic 360° video
  - Suitable when using VVC encoders, decoders and file parsers in a system which does not include viewport dependent delivery or viewport dependent decoding
- OMAF 3rd edition also defines a Common Media Application Format (CMAF) media profile for the VVC-based viewport independent OMAF video profile.
  - Adds constraints related to frame packing format and coverage information for CMAF switching sets

OMAF : Viewport-dependent Streaming (1/2)

- Viewport-dependent delivery (VDD) methods aim at representing the content being currently displayed on the viewport at a better picture quality and/or resolution than the remaining parts of the omnidirectional video content.
- Viewport Dependent streaming:
  - Viewport-specific (VS)
  - Tile-based (tile = independently decodable region, e.g. VVC subpicture)
- VVC-based simple tiling OMAF video profile in the 3rd Edition of OMAF supports both viewport-specific VVC streams and subpicture-based VDD
- Viewport Specific:
  - Viewport-specific VVC bitstreams are encoded and encapsulated in its own VVC track.
  - Region-wise quality ranking (RWQR) metadata is included in each track and/or associated with the respective representation in a DASH MPD.
  - RWQR metadata enables players to determine the intended viewport of the viewport-specific stream and hence select a proper stream for the currently prevailing viewport.
- Subpicture-based:
  - At least one VVC merge base track and the respective VVC subpicture tracks are required to be available
  - The 'alte' track group, the 'subp' track reference, and the subpicture order sample group is typically used

OMAF : Viewport-dependent Streaming (2/2)

- Encoding with two different IRAP periods and two qualities:
  - Shorter IRAP: React to viewport changes
  - Longer IRAP: better compression efficiency
- Each sequence of subpictures encapsulated in separate VVC subpicture track
- Collocated subpicture tracks gathered in same '`alte`' track group (g1 to g8)
- VVC merge base track formed including ‘`subp`’ track reference to each '`alte`' track group and a subpicture order sample group which references each '`alte`' track group in spatial order
- To merge subpictures, a player follows the VVC merge base track, which references an '`alte`' track group for each subpicture location.
- At time tM, tracks 1, 5, 13, and 17 from the longer IRAP period, tracks 32, 36, 24, and 28 are received from the shorter IRAP period version.
- The picture merged at `tM` contains mixed IRAP, non-IRAP subpictures
- Later when an IRAP picture appears for the longer IRAP period, the player switches back to using the longer IRAP period versions for all subpictures.



a) Encoded VVC bitstreams with subpictures

b) VVC subpicture tracks and merge base track that are made available for streaming

c) Example of received subpictures merged to a VVC bitstream

Figure 2: Example of subpicture-based mixed-quality viewport-dependent streaming

Conclusion

- Provided an overview of Versatile Video Coding (VVC) standard from the perspective of immersive video.
- Described the ISOBMFF storage of VVC exemplified for 360° video streaming.
- Summarized the VVC-based profiles for the Omnidirectional Media Format (OMAF) standard

SHARP®