
ACM Multimedia Systems Conference – Amsterdam, 12 June 2018
Rob Koenen

- Principal Consultant, TNO
- President, VR Industry Forum
- Chair, MPEG Roadmap AHG
- Co-Founder and Chief Business Officer, Tiledmedia
VR360: STATE OF THE INDUSTRY

VR360: STATE OF THE STANDARDS

VR360: STATE OF THE ART
VR360: STATE OF THE INDUSTRY
Source: HypeVR
Sources:
- BT Sport
- Road to VR
- Sky UK
Where are we headed?

- Full immersion - “six degrees of freedom”
- Real and computer-generated – and indistinguishable
- Immersive story-telling
- Enjoy an event as if you were there
- Enjoy it with friends
Where are we today?

Gartner **Hype Cycle** for Emerging Technologies, 2017

**Virtual Reality**

Slowly Climbing Out!

gartner.com/SmarterWithGartner

Source: Gartner (July 2017)
© 2017 Gartner, Inc. and/or its affiliates. All rights reserved.

As of July 2017
Where are we today?

- First steps: VR360
- Video: Mono or Stereo
- Audio: Stereo or Spatial
- Very low resolution
- Limited (head) motion
- Large HMDs
VR Adoption is promising
Some Predictions

By 2022, VR will grow to $74.8B global marketplace, including all hardware, content, software and services

Superdata (Stephanie Llamas, at VRX Europe2018): Consumer Revenue

Greenlight Insights (Alexis Macklin, CES 2018, at VRIF Masterclass): Total Revenue
Ecosystem is developing
Lots of new products

- Stand-alone devices, no strings attached
- Tracking built in 3DoF or 6 DoF

- But: most of VR consumption is still on flat devices!
  (multiple sources, e.g. Sky at NAB 2018)
What is required

• Attractive user experience
• Great content
• Easy to use
• No side-effects
• Affordable
  – for consumers
  – for providers
• Interoperable
VR360 in Practice

- Distribution: 4k x 2k if you’re lucky
  - 1k x 1k per eye for the viewport (even less than 1280 x 1440 available on e.g. Samsung S7)
- Only 4k x 1k per eye if it’s stereoscopic
- Consensus: can use up to 8k x 8k per eye
  - But: only required in fovea!
- Audio in stereo
- Better headsets coming
- Capture at increasingly high resolutions
  - 8k, 12k, even 16k
- Audio increasingly in Ambisonics
  - First order, higher order; binaural rendering works great
VR needs better **quality**
VR needs to be more **interactive**
VR needs to be more **social**
... and it’s all coming!
VR360: STATE OF THE STANDARDS
Standards for VR360 (just a few)

• MPEG’s Omnidirectional MediA Format (OMAF)
  – Coding, packaging, metadata, delivery
• Khronos
  – Interfaces to renderer
• DVB
  – Commercial Requirements under development
• 3GPP
  – Profiles of MPEG Coding Tools for VR360 distribution
• DECE
  – Glossary (adopted and maintained by VRIF now)
• W3C
  – WebVR & WebXR – geared towards CGI-type content
• VRIF
  – Guidelines; Promotion & Adoption of VR standards
MPEG’s 5 year Roadmap Focuses on Immersive Media
Internet Video Coding

Genome Compression

Descriptors for Video Analysis (CDVA)

6 DoF Audio

Point Cloud Compression

Versatile Video Coding

Dense Representation of Light Fields

Scene Description for Immersive Media

Network-Based Media Processing

6 DoF Application Format

OMAF v1

OMAF v2

Media Orchestration

MIAF

Web Resource Tracks

IoMT

IoMT
PCC Extensions?

OMAF v2

2018

2020

2021

2022

Jan 2017

2018

2019

2020

Internet Video Coding

Genome Compression

Descriptors for Video Analysis (CDVA)

2019

2021

2023

Jan 2023

Coding

Systems

and Tools

Immersive Media with 6 Degrees of Freedom

Combining Natural and Synthetic content

6 DoF Audio

Point Cloud Compression

Versatile Video Coding

Scene Description for Immersive Media

Coding

Network-Based Media Processing

6 DoF Application Format

VR360, on-demand and live (3 DoF)

OMAF v1

Media Orchestration

VR360, on-demand

Web Resource Tracks

IoMT

MIAF

Scene Description for Immersive Media
MPEG-I(immersive)

Most Recent MPEG project: ISO/IEC 23090

*Coded Representation of Immersive Media*

8 parts are underway:
1. Architectures for Immersive Media
2. Omnidirectional MediA Format
3. Versatile Video Coding
4. New & Immersive Audio Coding (name t.b.d.)
5. Point Cloud Coding
6. Metadata for Immersive Services and Applications
7. Metrics for Immersive Services and Applications
8. Network-Based Media Processing

Talk by Phil Chou tomorrow 9:15!
MPEG OMAF

- **Interoperable exchange of VR360 is a significant challenge**
OMAF - Projections and Mappings

- Equirectangular
- Cubemap

... and other ways of doing “region-wise packing”
OMAF - Coordinate Systems

- Surprisingly hard to get consistent across all subsystems
- X, Y, Z,
- azimuth ($\phi$) and elevation ($\theta$)
- yaw / pitch / roll

Source Pictures: OMAF specification
OMAF - Encoding and Metadata

• Coding (Profiles)
  – HEVC and AVC for Video
  – Audio: MPEG-4 AAC and MPEG-H Audio; (spatial and “2D”)
  – Pictures: HEVC; JPEG

• Metadata
  – Initial viewport, recommended viewing direction, director’s cut ...
  – Timed, and needs to be in sync with media data
OMAF - Delivery

• Encapsulation in ISO Base Media File Format
  – Adding timed text

• Transport using DASH and MMT

• Viewport-independent (or -agnostic) streaming
  – just send everything, no matter where the viewer looks

• Viewport-dependent streaming
  – Send viewport with better quality
  – Several ways to do this - we’ll get back to this
VR Industry Forum
Mission VR Industry Forum

To further the widespread availability of high quality audiovisual VR experiences, for the benefit of consumers

- Non-profit organisation established during CES 2017, after a year of informal meetings
Who’s in?
Goal – Creating a Market for 360VR

For consumers:
• Make 360VR a high-quality, immersive, cross-platform experience

For content producers & service providers:
• Broaden reach and reduce cost caused by format proliferation (cost of production, distribution, etc.)

For device makers:
• Ensure a wealthy, premium quality content pipeline

For advertisers:
• Drive the creation of a broad, unique & innovative sales channel
Guidelines

- Published Guidelines at CES 2018
  - Production
  - Distribution
  - Security
  - Creation of Interoperable points

- Lexicon for common terminology available at www.vr-if.org
Human Factors

- Human Factors that impact the VR experience
  - Physiological (eye/human visual system, ear/human auditory system)
  - Physio-cognitive (motion sickness, sensory conflicts)
  - Psycho-cognitive (presence, realism of immersion, interaction)
  - Psycho-social (violence, addictions, etc.)

Source Pictures: Wikipedia/Wikimedia
Production

- How to produce immersive quality content
- Started from SKY’s “Technical Guidelines”
- Technical recommendations (capture, recording, resolution, immersive audio, storage and exchange formats, frame rates ...)
- Incorporating results of human factor studies (cuts, motion, etc.)
- Content Exchange Metadata
Distribution

- Based on “OTT Download and Streaming” cases
- Guidance and recommendations to implement VR video and audio profiles from MPEG OMAF (“Omnidirectional MediA Format”)
  - Viewport Independent media profile
  - Viewport Dependent media profile
  - 3D Audio media profile
- Configuration of packing, projection and supporting metadata
- Use of Adaptation Sets for MPEG DASH based streaming
- Now working on Live VR Services, and will soon address HDR
Security

- High quality VR productions very expensive; need to be able to be monetized
- This means content protection
- Starts from MovieLabs’ Enhanced Content Protection;
- Challenge: use Common Encryption with tiled streaming

- Now working on watermarking for VR content
VRIF: Next up …

• Tools and Content to help the ecosystem
VR360: STATE OF THE ART
Take a VR video
Users see small part of sphere
Viewport-adaptive Streaming

- Foveated rendering (*rendering*)
- Predict where people look & encode that better
- Facebook’s Pyramid approach
- Pixvana’s FOVAS (Field of View Adaptive Streaming)
- Tiled Streaming

*Source: Facebook*

*Source: Pixvana*
What Matters in Viewport-Adaptive?

- Video Quality and Required Bitrate
- Motion-to-High-Resolution Latency
Tiled Streaming Principles

- Cut image up in tiles
- Some tiles are high-resolution, some low
- Use high-resolution tiles for viewport
- Low-resolution tiles displayed briefly when viewport changes, until high-resolution tiles available
- Use one single 4k decoder to display 6k or even 8k ERPs

Two approaches:
- Early Binding \(\rightarrow\) prepare possible tiling configurations in advance
- Late Binding \(\rightarrow\) let client determine what to retrieve at which resolution.
Late Binding:
Brief Explanation & Experimental Results
Late Binding:
Brief Explanation & Experimental Results
Take a VR video
Users see small part of sphere
Cut the video into “tiles”
Only Stream Tiles in View ...
Only Stream Tiles in View
Send Low-Resolution Background
Only Switch Part of Viewport

Tiles in viewport

- Cancelled tiles
- Newly requested tiles
Tiling works better with a cubemap
Layered for Adaptive Bitrate Streaming
### Comparing the Two Approaches

#### “Early Tile Binding”
- Use pre-determined configurations using “extractor tracks”
- Low processing overhead
- Need separate configs for different clients and viewports
- Which config to retrieve depends on viewing direction and some adaptive bitrate logic
- Switch at random access points in Dash segments
- Easier to implement and make interoperable
- More efficient than “legacy”

#### “Late Tile Binding”
- Determine what to retrieve and decode in real-time
- Bitstream rewriting on the Client
- Accommodates different clients and viewports
- Smart clients take intelligent, last millisecond decisions; Client decides which quality tiles to retrieve
- Switch on any tile of any frame to rapidly display high quality content
- Implementing late binding requires a bit of advanced client logic
- Much more efficient than “legacy”
Resulting Bitrates

- 4k x 2k Mono ERP: ~ 5 Mbit/s
- 4k x 2k Stereo ERP: ~10 Mbit/s
- 6k x 3k Mono ERP: ~10 Mbit/s
- 8k x 4k Mono ERP: ~15 Mbit/s

- Resolution of ERP; actual distribution uses cubemaps
- 70 - 80% bitrate reduction over “legacy”
- Rates depend on content complexity and viewport (head) motion
- Using actual networks (Akamai, CloudFront), not just local tests
Experimental Results: Cache Hints

TILE SWITCHING LATENCY (TILEDMEDIA V2.9, OVER WIFI, FROM AKAMAI)

- Cold cache
- Cold with cache-hints
- Warm cache

% OF TILE SWITCHES COMPLETED

# OF FRAMES AFTER TILE SWITCH

R. van Brandenburg, R. Koenen (Tiledmedia), D. Sztykman (Akamai), CDN optimization for VR streaming, IBC 2017
Experimental Results: QUIC

**Time-to-first NAL**

- Very good network
- Average network (Akamai State of the Internet)
- Bad network

- TCP
- Default QUIC
- Tuned QUIC

**Average percentage low-resolution in viewport**

- Very good network
- Average network (Akamai State of the Internet)
- Bad network

- TCP
- Default QUIC
- Tuned QUIC

---

R. van Brandenburg, R. Koenen (Tiledmedia), D. Sztykman (Akamai), CDN optimization for VR streaming, IBC 2017
Video coding and Networking need to be addressed **together** for the best performance.
Tip of the Day ...

Video coding and Networking need to be addressed together for the best performance (and I have a demo :)}
Video coding and Networking need to be addressed together for the best performance

(and I have a demo :)
(oh - and we are hiring !!)