VRIF's View on State of the Immersive Media Industry

Ozgur Oyman (Intel), Rob Koenen (Tiledmedia), Paul Higgs (Huawei), Chris Johns (Sky), Richard Mills (Sky), Mick O'Doherty (Irdeto)

1- Introduction

Founded at the Consumer Electronics Show in 2017, the Virtual Reality Industry Forum (VRIF) (www.vr-if.org) is chartered with the mission to further the widespread availability of high quality audiovisual Virtual Reality (VR) experiences. Its charter includes delivering guidelines and advocating industry consensus around common technical standards for end-to-end VR interoperability from content production through distribution to consumption. VRIF delivered the first release of its guidelines in early 2018 detailing an initial set of best practices and recommendations for realizing high quality, interoperable 360 degree video experiences. It is currently working toward releasing Guidelines version 2.0 with the inclusion of new features such as live VR services, HDR support, use of text and fonts, and watermarking for theatrical VR. New features on VRIF's 2019 roadmap include guidelines around volumetric content production and distribution, cloud VR and edge computing, augmented reality (AR) support and enabling social VR experiences.

This paper summarizes VRIF's views on the current state of the immersive media industry. It explains how the adoption of VR has been slower than predicted a few years ago, and the main technical and business reasons behind this. It will show how these reasons are being addressed, and that the outlook is very healthy. Next, we will look at what's being done to make immersive experiences even more attractive, and how networks are evolving to support these more advanced services that VRIF believes will pave the way for mass adoption of immersive media products and services in the market.

2- Current State of the Immersive Media Industry

VR/AR use cases continue to generate significant interest in the industry, for both consumer and enterprise, and many people continue to see immersive media as one of the key disruptive trends to change the future of how people live and work. Just to mention a few: Live entertainment events have huge potential for immersive media usage, for instance changing the way people can enjoy sports games and music concerts. Social VR and immersive

communication / telepresence could fundamentally change how people collaborate and interact. On the enterprise side, immersive VR/AR technologies could disrupt many sectors including education, health care, retail, tourism, marketing, training, public services and factory production.

In the meantime, the growth of the VR market has been considerably slower than initially expected when the VRIF was launched in January 2017. No immersive media application including VR360 has yet become mainstream. VR hardware, software, and content simply were not ready for mass adoption and expectations outpaced reality. In 2019, there is still no killer app for immersive media and the VR user base is still considered to be too small for an attractive and sustainable developer ecosystem.

Some of the reasons for the slow growth of the VR market can be described as follows:

1. Headsets: A clear limitation of premium VR360 experiences today is in the HMD specifications. That includes low display resolutions and the need for tethering to a powerful personal computer. In order to be able to match the quality of on-screen TV experiences, the pixel-per-degree count for Head Mounted Displays (HMDs) needs to be a lot higher as the content is displayed very close to the user's eyes. Going forward, HMD resolutions will need to increase by several orders of magnitude from today's roughly 1k per eye in order to deliver photorealistic quality that has a similar level of detail as 4K televisions. Even as hardware improves to support better resolutions, rendering interactive content at higher resolutions takes increased amounts of local computing power, which leads to a more expensive computing platform for the end user. Low visual quality in a VR experience reduces the lack of realism leading to less overall usage with shortened durations. Furthermore, many HMDs today either need to be tethered to a personal computer (e.g., Oculus Rift, Sony Playstation VR, etc.) with a powerful CPU and GPU (e.g., Intel Core CPU and Nvidia GTX GPU), limiting the flexibility and freedom of user movement or need to be standalone (e.g., Oculus Quest / Go), which means that premium high-end VR experiences may not be supported due to potential limitations on the local compute capabilities of the HMD. For example, frame rates for standalone HMDs remain at 60-72 fps, while 90-100 fps is achievable with PC-connected HMDs delivering smoother immersive experiences.

2. Content: Immersive content production is still in experimentation, without well established production methodologies and tools. Storytelling in the 360 environment is fundamentally very different from traditional content production. In comparison to 2D video production, immersive video production requires significantly more manual processes to create and publish content, often requiring significant capital investment. Furthermore, much of the production requires specialized stitching software (still in its infancy) to make the final product. Outside of expensive volumetric video studios, such as Intel Studios, that can create more fluid VR content without the need to stitch, creating high-quality VR360 video is still hard, complex, and expensive.

3. Networks: High bandwidth and low latency requirements in order to meet the high quality and interactivity needs of immersive media experiences may not always be supported in today's networks, even for VR360 content distribution which requires at least 4K - 6K resolution for reasonable quality with state-of-the art codecs such as HEVC and AV1. The bandwidth demands get even stronger for higher end VR experiences involving higher resolutions (e.g., 8K, 16K, etc.) and a greater degree of immersion (e.g., with volumetric content). The negative consequence of lack of sufficient QoS support (e.g., low bandwidth, high latency) from the network can mean poor user experience resulting in low video quality.

4. Standards and interoperability: Many immersive media standards have been developed over the last few years with strong potential for broad industry adoption, but work remains on new codecs, formats, protocols, APIs, etc. to support the essential end-to-end interoperability needs of immersive media production, distribution and consumption, and prevent fragmentation.

5. Cost: Monetization remains a challenge for immersive media products and services. More affordable devices, content and services are necessary to enable mainstream adoption. On the other hand, with the currently limited number of consumers and revenue opportunities, it remains harder to justify the major investment required to tackle challenging technical issues such as the need for specialized hardware and software necessary for immersive professional content production.

6. Human Factors: Users get frustrated with their VR experiences due to the following: (1) VR is isolationary and not yet social in its nature. (2) Reasonable expertise is necessary to install and operate VR headsets which is beyond a "retail shopping experience". (3) Wearing VR headsets is not comfortable for long periods (weight, heat, ventilation).

While the list of hurdles above may sound quite bleak, the reality gives us reasons to be more optimistic based on the recent developments in the market, including the following:

1. Headsets: HMDs with better resolution are coming to market (2k x 2k per eye) driven by advances in LCD/LED tech, which provide some example products (e.g., HP Reverb, Pimax 8K, etc.). Moreover, standalone HMDs with integrated system-on-chip (SoC) processor capabilities (e.g., those based on Qualcomm Snapdragon processor) help in resolving the problem of tethering and overall complexity, although tethering to a powerful personal computer may still be necessary for premium high-end VR experiences.

2. Content: Lots of interesting content is being developed and made available through app stores (e.g.,Google's VR Creator Lab with award winning YouTube VR content, etc.).

3. Networks: When using advanced viewport-adaptive streaming technologies, content can be distributed over today's networks with Equi-Rectangular Projection (ERP) resolutions

as high as 8k or 6k stereoscopic. This viewport-dependent approach allows different areas/regions of the VR360 video to be delivered with different quality or resolution, realizing the best quality-bandwidth tradeoff.

4. Standards and interoperability: The de facto standards and guidelines to deploy VR360 are already present today, including relevant specifications from MPEG, 3GPP, Khronos, W3C, etc. and guidelines and best practices from VRIF.

5. Cost: Cost of headsets is coming down and new cameras are appearing that are more affordable.

6. Human Factors: Expanding VR360 with AR and MR on mobile and tablet devices is growing in popularity with users due to the simplicity of pick up and try experiences.

3- Recent Advances on Immersive Media

Looking a bit further out, significant investments from big tech companies on immersive media products and services continue, indicating their sustained belief in growth of VR/AR market over the long term. Here are a few trends, usages and technology advancements, which VRIF believes will help to accelerate the long-term trend toward realization of mainstream adoption of VR/AR by removing the inherent limitations of 3DoF VR and enabling further network capabilities to better support the high quality and interactivity requirements of immersive media experiences.

3.1 - Beyond VR360: 3DoF+ and 6DoF Experiences

Initial VR360 support was limited to 3 degrees of freedom (3DoF), which means that the viewing pose is only alterable through rotations on the x, y and z axes, represented as roll, pitch and yaw respectively, and purely translational movement does not result in different media being rendered. As such, VR360 delivered an overall flat experience since it positions the viewer in a static location with limited freedom of movement and low levels of interactivity. This was a limitation in the sense that fully immersive experiences were not possible thereby hurting the user experience and sense of realism. Emerging VR standards and products will provide support for 3DoF+ and 6 degrees of freedom (6DoF) enhancing the level of immersion and user experience. While 3DoF+ restricts modifications of the viewing position by limiting translational movements of the user's head around the original viewpoint, 6DoF supports both rotational and translational movements allowing the user to change not only orientation but also position to move around in the observed scene. As part of its "Coded Representation of Immersive Media" (MPEG-I) project, MPEG is currently developing the codecs, storage and distribution formats, and rendering metadata necessary for delivering interoperable and standards-based immersive 3DoF+ and 6DoF experiences [1].

3.2 - Volumetric Content and Point Clouds

Volumetric video has been recently gaining significant traction in delivering 6DoF experiences. Volumetric video contains spatial data and enables viewers to walk around and interact with people and objects, and hence it is far more immersive than 360 video footage because it captures the movements of real people in three dimensions. Users can view these movements from any angle by using positional tracking. Point clouds are a volumetric representation for describing 3D objects or scenes. A point cloud comprises a set of unordered data points in a 3D space, each of which is specified by its spatial (x, y, z) position possibly along with other associated attributes, e.g., RGB color, surface normal, and reflectance. This is essentially the 3D equivalent of well-known pixels for representing 2D videos. These data points collectively describe the 3D geometry and texture of the scene or object. Such a volumetric representation lends itself to immersive forms of interaction and presentation with 6DoF. Figure 1 depicts a volumetric video capture system for a soccer game using Intel's True View (already established in NBA, NFL and La Liga) with 38 ultra high-definition cameras placed uniformly across the stadium to generate 3D images. Since such point cloud representations require a large amount of data, development of efficient compression techniques is desirable in order to reach consumers using typical broadband access systems. As part of its "Coded Representation of Immersive Media" (MPEG-I) project, MPEG is currently developing the "Point Cloud Coding" (PCC) standard to compress point clouds using any existing or future 2D video codec, including legacy video codecs, e.g. HEVC, AV1, etc [2].

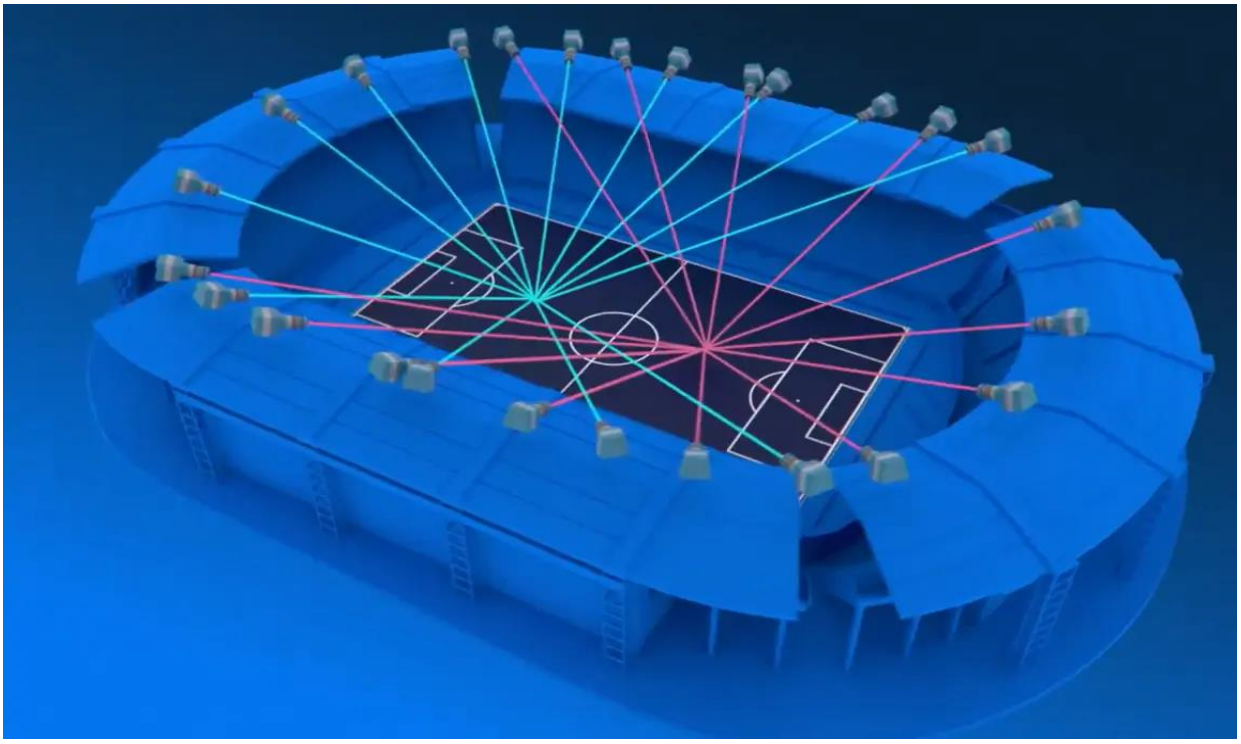


Figure 1 - Example volumetric video capture system using Intel's True View.

3.3 - Advances in Audio Representation for VR

The addition of dimensionality to audio in 3DoF and 6DoF VR experiences has significantly enhanced immersion. For 3DoF VR360 video experiences, Ambisonic audio has been widely deployed while object-based and channel-based audio are also available. Phase-shifting stereo tracks within an audio mix enables creators to add specific x,y,z positioning of audio effects within an experience. Timeline Audio editing packages are available to enable the dynamic tracking of spot audio effects to visual objects in the scene. The use of realistic scene-related Equalization, Presence and Reverberation also increases the sense of immersion. For 6DoF Game Engine-based experiences, sophisticated toolsets are available to manage object-based audio effects and realistically map the audio characteristics of spaces in the experiences. As audio is part of it's "Coded Representation of Immersive Media" (MPEG-I) project, MPEG is currently developing 6DoF audio distribution formats and rendering models necessary for delivering interoperable and standards-based immersive 3DoF+ and 6DoF experiences.

3.4 - Cloud VR/AR

The ability to leverage cloud computing and edge computing can be instrumental in reducing computational cost and complexity for the client devices when it comes to processing requirements associated with immersive media, including workloads such as decoding, rendering, graphics, stitching, encoding, transcoding, caching, etc., which may all be performed over the cloud and/or edge, as depicted in Figure 2 for a few example applications. Low latency, high throughput networks such as 5G could allow instantaneous access to remotely stored data while offering a local computing experience similar to a data center based system. Such a high capacity network with low latency characteristics also enables responsive interactive feedback, real-time cloud based perception, rendering, and real-time delivery of the display content. With a cloud-based approach, it is sufficient to use low-cost thin client devices with minimal built-in functions. For VR and AR, these include the display screen, speakers for output, vision positioning and hand-controller sensors for input. The thin client simply uses the network to pass inputs to processing functions at the cloud or edge, and receives the necessary images to display. Cloud VR/AR brings together the significant advances in cloud computing and networks with a high degree of interactivity to provide high quality experiences to those who were previously priced out of immersive technologies. This approach may thus help significantly increase the number of VR client devices sold, and create major revenue opportunities from increased consumption VR/AR services for content providers and operators. A currently ongoing GSMA Cloud VR/AR initiative [4] aims to address this opportunity from the perspective of mobile operators, who aim at leveraging their infrastructure to monetize VR/AR/XR services, cooperating with content providers.

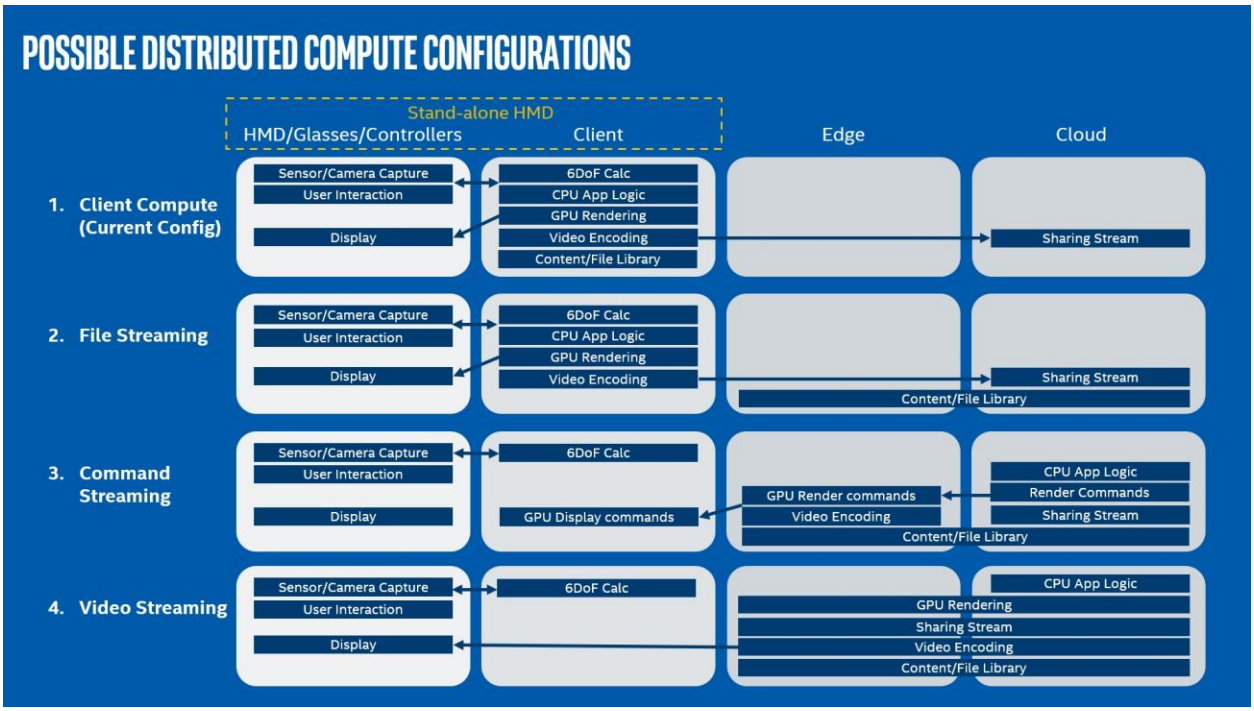


Figure 2 - Possible cloud VR configurations with media processing performed at the cloud and/or edge.

3.5 - 5G

Higher bandwidths, lower latencies and support for edge computing enabled by 5G connectivity can provide the desirable means to meet the high quality and interactivity needs of immersive media experiences. 5G can support a wider range of QoS requirements addressing high bandwidth low latency needs of interactive VR/AR/XR applications through a New Radio (NR) air interface as well as flexible QoS enabled via 5G core network architecture and network slicing. Moreover, the ability of the 5G system to leverage edge computing is essential for meeting the performance requirements of immersive media, not only for better delivery performance via edge caching but also to offload some of the complex VR/AR/XR processing to the edge to perform various operations such as decoding, rendering, graphics, stitching, encoding, transcoding, etc., thereby lowering the computational burden on the client devices. A relevant potential solution for offloading compute-intensive media processing to the edge is based on MPEG's Network-Based Media Processing (NBMP) specification ISO/IEC 23090-8, which aims to specify metadata formats and APIs for intelligent edge media processing. For instance, such a capability can be relevant to 3GPP's Framework for Live Uplink Streaming (FLUS) service in TS 26.238, in which videos captured by one or multiple omnidirectional cameras (without built-in stitching) may be sent separately to the cloud or edge via an uplink connection, where they may be stitched to create 360 videos and then encoded and encapsulated for live distribution. Edge enhancements enabled by 5G also help in improving viewport-dependent VR360 delivery [3], where high quality viewport-specific video data (e.g., tiles) corresponding to portions of the content for different fields of view (FoVs) at various quality levels may be cached at the edge and delivered to the client device with very low latency based

on the user's FOV information. Benefits of other technologies for DASH streaming enhancements at the edge are also applicable for immersive media, such as network assistance capabilities based on 3GPP Server and Network Assisted DASH (SAND) in TS 26.247. On the content production side, there is also an interest from mobile operators and content providers to leverage 5G networks for not just distribution, but also production of professional audio-visual (AV) immersive media content and services, including studio-based production and remote AV production. 3GPP is currently conducting a feasibility study on this possibility.

3.6 - Continued Hardware Innovations

It is expected that immersive media services will benefit from continuous hardware innovations over the next few years striving for even more advanced compute capabilities, resulting in higher performance CPU/GPU etc. as well as better resolution displays, that can deliver 4K resolution per eye and process immersive content at higher resolutions, e.g., 8K, 16K, 32K, etc. Similarly, on the infrastructure side, cloud and edge solutions with optimized immersive media processing and delivery capabilities for performing operations such as decode, rendering, stitching, encode, etc. as well as those for efficient transport with minimal latency are also expected.

3.7 - Security

Balancing VR and AR security concerns and the content processing drivers that they bring, while minimizing processing overhead and power use, both at the headend and in the play back client, requires careful engineering. Full end to end immersive media content security includes encryption, the ability to forensically watermark content, to verify the source of content, to verify that a target device is who or what it says it is and to detect, report on and take down pirated content on the Internet and on broadcast networks. Valuable VR and AR content, like other forms of high value content, will need to leverage content security technologies in each of these domains and each technology will provide different challenges and solutions for this content. Immersive content distribution and rendering brings its own security challenges, particularly in the area of a device's secure media path and in tiled or viewport dependent delivery and display. For existing high value content, many devices now support a secure media path, meaning encrypted media is decrypted into secure memory and displayed from there, without allowing any access from other parts of the system or from other apps. As immersive content may require extra manipulation of the media before display, for example to create the view to display from a given 360 projection format, devices need to provide a mechanism to allow this media manipulation securely. Secure mechanisms are also necessary to support any work required to construct the user view from an immersive media tiled delivery solution. Forensic watermarking similarly needs special techniques to allow for tiled or viewport dependent media delivery, and for partial media capture and reproduction. Copying just the viewport visible portion of the media may in some cases be a piracy risk itself. While addressing these challenges to develop solutions that enable operators and service providers to roll out immersive media offerings as simply as possible, it is key to leverage industry forums and standards initiatives to ensure that interoperability and scalability is maintained.

4 - Conclusion

This paper summarized the views of the VRIF on the current state of the immersive media industry. Despite the various challenges faced, significant investments from big tech companies on immersive media products and services continue consistently and the outlook is very healthy based on the recent developments in the VR/AR market signaling solid prospects of long-term growth. Moreover, VR/AR use cases continue to generate significant interest in the industry, for both consumer and enterprise. Many people continue to see immersive media as one of the key disruptive trends to change the future of how people live and work. VRIF also believes in the imminent mainstream market adoption of immersive media products and services over the long term, but also observes that it could take another 4-5 years until the technology is mature, consumers are happy from both cost and experience perspectives, and business models become more attractive for service providers, operators, content providers and device / infrastructure vendors.

5 - References

[1] M. Wien, J. M. Boyce, T. Stockhammer, W.-H. Peng, "Standardization Status of Immersive Video Coding", IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol:9, no:1, pp. 5-17, March 2019.

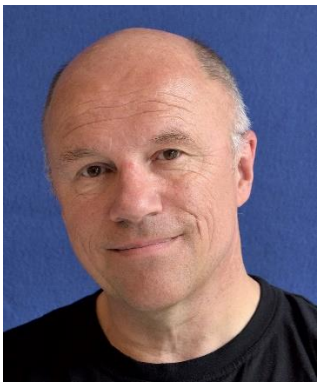
[2] S. Schwarz et. al., "Emerging MPEG Standards on Point Cloud Compression", IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol:9, no:1, pp. 133-148, March 2019.

[3] VR Industry Forum Guidelines, available at: https://www.vr-if.org/guidelines/

[4] GSMA Cloud AR/VR Whitepaper, available at: https://www.gsma.com/futurenetworks/resources-2/gsma-online-document-cloud-ar-vr-whitepaper/

6 – Author Bios

Ozgur Oyman is a Principal Engineer at Intel's Next Generation & Standards (NGS) organization, leading various mobile multimedia related research and standardization initiatives for Intel in 3GPP SA4, DASH-IF and VR-IF, addressing areas such as VR/AR/XR, 5G media distribution, IMS/VoLTE/ViLTE services and edge computing. In VR-IF, he currently serves on VR-IF Board of Directors as a board member and treasurer and also chairs VR-IF's Liaison Working Group. He also serves as head of delegation for Intel in 3GPP SA4 working group that specializes on mobile multimedia services, and related codecs, protocol stacks and file formats. He's held numerous editorship, rapporteurship and chairmanship roles for 3GPP SA4, MPEG and DASH-IF. He is an established innovator with 80+ granted patents in the USA (and many more globally) on mobile communications and multimedia networking. He holds Ph.D. and M.S. degrees from Stanford University and a B.S. degree from Cornell University (all in EE).



Rob Koenen is a co-Founder and the Chief Business Officer of Tiledmedia, the leading tiled VR streaming company. Rob is a co-founder of the VR Industry Forum, and served as its first President.

Mr. Koenen has held many leading roles in multimedia industry initiatives and standardization groups. He chaired MPEG's Requirements Group for almost 10 years, and played a key role in the development of the MPEG-4 standard. He initiated the MPEG Industry Forum in 1999, and served as its President for the first five years, successfully bringing MPEG-4 technology from a paper specification to a dominant market force.

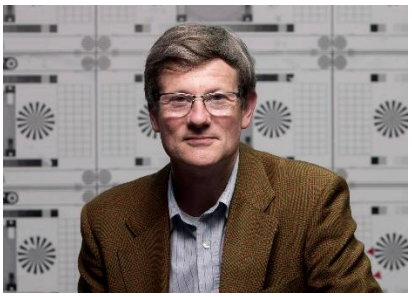
Paul Higgs leads the video strategy at Huawei where new concepts of immersive and traditional media are conceived and developed for deployment in their entertainment solutions. Since December 2018 he is the President of the VR Industry Forum after serving as a founding Board Member and leader of the Guidelines Working Group, while in the overall standards area he is also the co-chair of the technical group developing the Internet delivery aspects for the DVB Project.

Prior to Huawei, Paul held various positions at Ericsson where he was involved in moving communications and entertainment to IP platforms through VoIP and IPTV.

Chris Johns has been with Sky since its inception in 1989 having started his career in broadcast with the BBC. Chris formed part of the initial technical team tasked with launching Sky's multi-channel analogue satellite offering.

As chief engineer, Chris has been at the forefront of delivering broadcast functionality to the platform such as the multichannel digital playout facilities, Dolby Digital audio, server-based solutions and compression systems. Having played a key role in Sky's HD launch and design of its HD infrastructure, he continues to evolve many new experiences such as UHD, Virtual Reality and enhanced audio as well as the associated technologies to deliver better images and sound to the consumer.

Sitting on and chairing many broadcast groups and societies, Chris is also a SMPTE Fellow.



Richard Mills, Technical Director, Sky VR Studios, has over 35 years' experience of operations, management and development in both Film and Broadcasting industries. He has expertise in Studio operations and management, Post Production and Transmission (MOLINARE), Board-level technical leadership, design and management in facilities and Outside Broadcast (NEP VISIONS). Richard is a SMPTE UK Board Member.

For the past ten years he has provided award-winning imaging and workflow solutions to film, documentary and drama.

Richard now provides production solutions and consultancy in complex imaging, Scene, Motion and Object capture and Virtual Reality.

Richard is a strong believer in education and training. He has worked extensively with industry bodies including BSC, BKSTS, SMPTE, UK Screen, VRIF and many Universities to educate and encourage new talent, both technical and creative.

Mick O'Doherty works for Irdeto as a Technical Solutions Manager, focusing on partners in the video content security domain.

Mick is originally from Dublin, Ireland and has worked in a number of countries over the last 20 years, mainly the UK, Turkey and Canada. His background is in telecoms audio and video software and systems development, but he has worked on the operator side of the table also as well as with a number of web and mobile start-ups, mainly in the West of Ireland.

Mick has previously worked in standards in the Java JAIN area and supported standards efforts in the SIP domain. He is active in the VR Industry Forum, focusing on immersive media security.