



*MOVING THE EUROPEAN
XR TECH INDUSTRY FORWARD*

Revised Landscape Report 2020

www.xr4all.eu



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 825545.

XR4ALL - eXtended Reality for All

Main authors		
Name	Organisation	Country
Oliver Schreer	Fraunhofer HHI	Germany
Ivanka Pelivan	Fraunhofer HHI	Germany
Peter Kauff	Fraunhofer HHI	Germany
Ralf Schäfer	Fraunhofer HHI	Germany
Anna Hilsmann	Fraunhofer HHI	Germany
Paul Chojecki	Fraunhofer HHI	Germany
Thomas Koch	Fraunhofer HHI	Germany
Serhan Gül	Fraunhofer HHI	Germany
Aurela Shehu	Fraunhofer HHI	Germany
Weiwen Hu	Fraunhofer HHI	Germany
Youssef Sabbah	Europe Unlimited S.A.	Belgium
Jérôme Royan	b<>com	France
Muriel Deschanel	b<>com	France
Albert Murienne	b<>com	France
Laurent Launay	b<>com	France
Jacques Verly	Image & 3D Europe	Belgium
Alain Gallez	Image & 3D Europe	Belgium
Sylvain Grain	Image & 3D Europe	Belgium
Alexandra Gérard	Image & 3D Europe	Belgium
Leen Segers	LucidWeb	Belgium
Mealle Quevillard	LucidWeb	Belgium
Gauthier Lafruit	Université Libre de Bruxelles	Belgium
Donna Schipper	Leiden University, Centre for Innovation	The Netherlands
Mitchell Bosch	Leiden University, Centre for Innovation	The Netherlands
Xiaoqing Jiu	Leiden University, Centre for Innovation	The Netherlands
Anastasia Pash	Globetrotter VR	Cyprus
Alan Chalmers	University of Warwick	United Kingdom
Luciana Gaspar	University of Warwick	United Kingdom



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 825545.

LEGAL NOTICE

The information and views set out in this report are those of the authors and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

It is important to notice that the official review process of this report is not being formally finalised yet by the European Commission.

© XR4ALL Consortium, 2020

Reproduction is authorised provided the source is acknowledged.



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 825545.

Table of Contents

ACRONYMS.....	7
1 INTRODUCTION	8
2 THE SCOPE OF EXTENDED REALITY	9
2.1 References.....	11
3 XR MARKET WATCH	12
3.1 Market development and forecast	12
3.2 Areas of application	16
3.3 Investments.....	18
3.4 Shipment of devices	19
3.5 Main players.....	22
3.6 International, European and regional associations in XR	25
3.7 Patents	29
3.8 References.....	32
4 XR TECHNOLOGIES.....	35
4.1 Video capture for XR	35
4.2 3D sound capture	38
4.3 Scene analysis and computer vision	41
4.4 3D sound processing algorithms	52
4.5 Interactive technologies for virtual flavour	55
4.6 Input and output devices	58
4.7 Render engines and authoring tools	66
4.8 Cloud services	66
4.9 Conclusion.....	69
5 XR APPLICATIONS	70
5.1 Advertising and commerce	70
5.2 Cultural Heritage	73
5.3 Education and Research.....	77
5.4 Industry 4.0	87
5.5 Health and medicine	101
5.6 Security and Sensing	110
5.7 Journalism & weather	114
5.8 Social VR.....	115
5.9 Travel and Tourism.....	121
5.10 Conclusion.....	126
6 STANDARDS	127
6.1 XR specific standards.....	127
6.2 XR related standards	130
7 REVIEW OF CURRENT EC RESEARCH	134

7.1	References.....	136
8	CONCLUSION.....	141

List of figures

<i>Figure 1: Extended reality scheme.</i>	9
<i>Figure 2: Major components of an eXtended Reality system.</i>	10
<i>Figure 3: VR/AR market forecast by Gartner and Credit Suisse [7].</i>	12
<i>Figure 4: Market growth rates by worldwide regions [4].</i>	13
<i>Figure 5: AR/VR regional revenue between 2017 and 2022 [15].</i>	14
<i>Figure 6: Global XR boost to gross domestic product by 2030 [18].</i>	14
<i>Figure 7: Major EU countries XR boost to gross domestic product from 2019 to 2030 [18].</i>	15
<i>Figure 8: Job enhancement by 2030 through XR [18].</i>	15
<i>Figure 9: Distribution of VR/AR companies analysed in survey by Capgemini Research [19].</i>	16
<i>Figure 10: Separated AR and VR sector revenue from 2017 to 2022 [8]. Note: This diagram represents just relations of revenue between different sectors, the scale for AR and VR are not the same.</i>	17
<i>Figure 11: XR use cases boost to gross domestic product from 2019 to 2030 [18].</i>	18
<i>Figure 12: XR4ALL analysis of European investors for start-ups [22].</i>	19
<i>Figure 13: VR unit shipments in the last three years [24].</i>	20
<i>Figure 14: Forecast of AR unit shipments from 2016 to 2022 [26].</i>	21
<i>Figure 15: Forecast of VR and AR shipments [14].</i>	21
<i>Figure 16: Forecast of VR and AR shipments with/without COVID-19 impact.</i>	22
<i>Figure 17: AR Industry Landscape by Venture Reality Fund [28].</i>	23
<i>Figure 18: VR Industry Landscape by Venture Reality Fund [28].</i>	24
<i>Figure 19: FlaVR at the British Science Festival in September 2019.</i>	56
<i>Figure 20: Comparison chart of VR headset resolutions [193].</i>	60
<i>Figure 21: AR applications in furniture domain: IKEA app (left), Homestyler (right).</i>	70
<i>Figure 22: AR applications in retail and food domain: real estate app by Onirix (left), food preview app by Jarit (right).</i>	71
<i>Figure 23: Sample views of ZREALITY virtual show room.</i>	71
<i>Figure 24: The fashion eco-system.</i>	72
<i>Figure 25: Mobile AR exposition.</i>	73
<i>Figure 26: VR Book: Researchers' Night participants using the digital book.</i>	73
<i>Figure 27: Holographic Stand: participants looking at the holographic display.</i>	74
<i>Figure 28: Bow Simulator: Students trying the haptic bow.</i>	74
<i>Figure 29: 3D digitisation of cultural heritage artefacts as developed by CultLab3D.</i>	75

Figure 30: Luxembourg in 1867 experience as developed by URBAN TIMETRAVEL.	76
Figure 31: AR can help medical students understand complex anatomy.	78
Figure 32: Students increase their media literacy by creating their own VR experiences.	79
Figure 33: Cellverse uses VR in a blended learning setting to teach students about cell functions.	80
Figure 34: Exploring Mars: VR can give access to normally unreachable places.	80
Figure 35: A virtual chemistry lab (left) and teaching practical lab skills in AR (right).	81
Figure 36: Industry 4.0.	87
Figure 37: AR application for assembly in industrial environment.	88
Figure 38: AR application for quality control in industrial environment.	90
Figure 39: AR application for maintenance in industrial environment.	91
Figure 40: AR application for planning in industrial environment.	92
Figure 41: AR application for logistics.	94
Figure 42: Warehouse concept by European Project SafeLog.	95
Figure 43: SafeLog concept exhibition at Logimat trade fair in Stuttgart, Germany (2019).	96
Figure 44: Classification of Attention Guiding Techniques [233].	96
Figure 45: A holographic augmented reality display as developed by WayRay [304].	98
Figure 46: Last-meter navigation prototype Error! Reference source not found..	98
Figure 47: AR application for training in industrial environment.	99
Figure 48: Example of training and learning use of AR in the medical domain.	102
Figure 49: Example of pre-operative use of AR/VR.	103
Figure 50: Example of intra-operative use of AR/VR.	105
Figure 51: Example of intra-operative use of AR.	106
Figure 52: Mindmotion VR by MindMaze (left) and Nirvana by BTS Bioengineering (right).	108
Figure 53: Example for post-operative use of VR/AR.	109
Figure 54: Security and Privacy Approaches to Mixed Reality.	111
Figure 55: Use of AR to enhance security and privacy.	111
Figure 56: Experimental MR Systems to evaluate the risk-taking behaviour of construction workers.	112
Figure 57: Smoke simulation to increase awareness and understanding of disaster risk.	113
Figure 58: Typhoon helmet developed by BAE systems.	113
Figure 59: Illustration of interaction in a virtual space, here based upon the vTime platform.	116
Figure 60: Illustration of interaction in a virtual space based upon the Rec Room platform.	117
Figure 61: Marriot Teleporter.	122
Figure 62: Globetrotter VR Live Virtual Tour.	123
Figure 63: Memoria: Stories of La Garma.	124

Acronyms

3D	Three-dimensional
AR	Augmented Reality
AV	Augmented Virtuality
BIM	Building Information Modeling
CAGR	Compound Annual Growth Rate
CNN	Convolutional Neural Networks
DoF	Degrees of Freedom
EC	European Commission
FOA	First-Order Ambisonics
GAN	Generative Adversarial Networks
GPS	Global Positioning System
GPU	Graphics Processing Unit
HEVC	High Efficiency Video Coding
HMD	Head-Mounted Display
HOA	Higher-Order Ambisonics
HRTF	Head-Related Transfer Function
ILD	Interaural Level Difference
ITD	Interaural Time Difference
IoT	Internet of Things
MPEG	Moving Pictures Expert Group
MR	Mixed Reality
OST	Optical See-Through
PCC	Point Cloud Compression
PST	Projective See-Through
QoE	Quality of Experience
ROM	Range Of Motion
S3D	Stereoscopic 3D
UK	United Kingdom
US	United States
V3C	Visual Volumetric Video-based Coding
V-PCC	Video Point Cloud Compression
VR	Virtual Reality
VST	Video See-Through
VVC	Versatile Video Coding
XR	Extended Reality

1 Introduction

This report provides a thorough analysis of the landscape of immersive interactive XR technologies carried out in the time period July 2019 until November 2020 by the members of the XR4ALL consortium. It is based on the expertise and contribution by a large number of researchers from Fraunhofer HHI, B&W and Image & 3D Europe. For some sections, additional experts outside the consortium were invited to contribute.

The document is organised as follows. In Sec. 2, the scope of eXtended Reality (XR) is defined setting clear definitions of fundamental terms in this domain. A detailed market analysis is presented in Sec. 3. It consists of the development and forecast of XR technologies based on an in-depth analysis of most recent surveys and reports from various market analysts and consulting firms. The major application domains are derived from these reports. Furthermore, the investments and expected shipment of devices are reported. Based on the latest analysis by the Venture Reality Fund, the main players and sectors in VR & AR are laid out. The Venture Reality fund is an investment company looking at technology domains ranging from artificial intelligence, augmented reality, to virtual reality to power the future of computing. A complete overview of international, European and regional associations in XR and a most recent patent overview concludes this section.

In Sec. 4, a complete and detailed overview is given on all the relevant technologies that are necessary for the successful development of future immersive and interactive technologies. Latest research results and the current state-of-the-art are described with a comprehensive list of references.

The major application domains in XR are presented in Sec. 5. Several up-to-date examples are given in order to demonstrate the capabilities of this technology.

In Sec. 6, the relevant standards and the current state is described. Finally, in Sec. 7, a detailed overview of EC projects is given that were or are still active in the domain of XR technologies. The projects are clustered in different application domains, which demonstrate the widespread applicability of immersive and interactive technologies.

2 The scope of eXtended Reality

Paul Milgram defined the well-known Reality-Virtuality Continuum in 1994 [1]. It explains the transition between reality on the one hand, and a complete digital or computer-generated environment on the other hand. However, from a technology point of view, a new umbrella term has been introduced, named eXtended Reality (XR). It is the umbrella term used for Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), as well as all future realities such technologies might bring. XR covers the full spectrum of real and virtual environments. In Figure 1, the Reality-Virtuality Continuum is extended by the new umbrella term. As seen in the figure, a less-known term is presented, called Augmented Virtuality. This term relates to an approach, where the reality, e.g. the user's hand, appears in the virtual world, which is usually referred to as Mixed Reality.

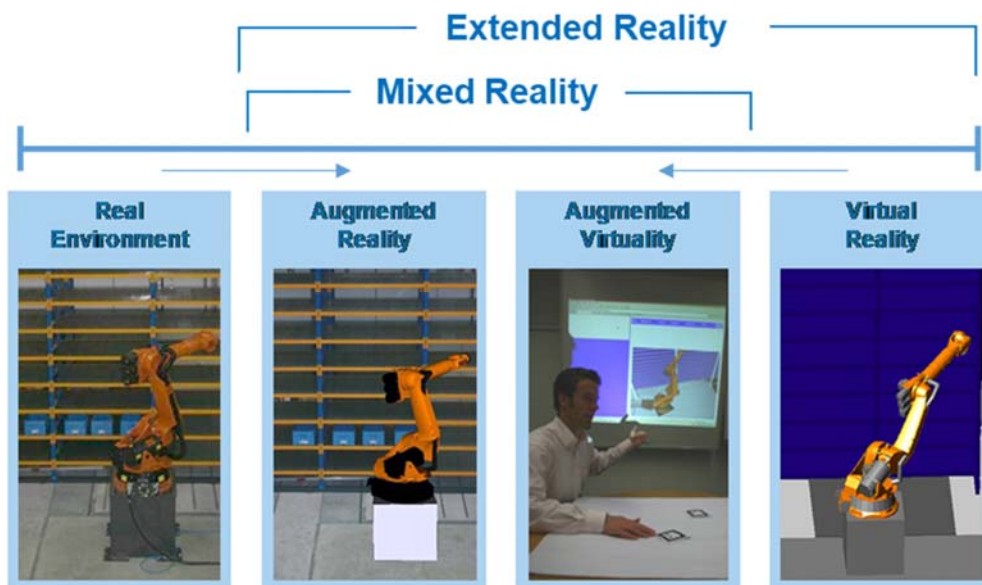


Figure 1: Extended reality scheme.

Following the most common terminology, the three major scenarios of extended reality are defined as follows.

Starting from left-to-right, Augmented Reality (AR) consists in augmenting the perception of the real environment with virtual elements by mixing in real-time spatially-registered digital content with the real world [2]. Pokémon Go and Snapchat filters are commonplace examples of this kind of technology used with smartphones or tablets. AR is also widely used in the industry sector, where workers can wear AR glasses to get support during maintenance, or for training.

Augmented Virtuality (AV) consists in augmenting the perception of a virtual environment with real elements. These elements of the real world are generally captured in real-time and injected into the virtual environment. The capture of the user's body that is injected into the virtual environment is a well-known example of AV aimed at improving the feeling of embodiment.

Virtual Reality (VR) applications use headsets to fully immerse users in a computer-simulated reality. These headsets generate realistic images and sounds, engaging two senses to create an interactive virtual world.

Mixed Reality (MR) includes both AR and AV. It blends real and virtual worlds to create complex environments, where physical and digital elements can interact in real-time. It is defined as a continuum between the real and the virtual environments but excludes both of them.

An important question to answer is how broad the term eXtended Reality (XR) spans across technologies and application domains. XR could be considered as a fusion of AR, AV, and VR technologies, but in fact it involves many more technology domains. The necessary domains range from sensing the world (such as image, video, sound, haptic), processing the data and rendering. Besides, hardware is involved to sense, capture, track, register, display, and to do many more things.

In Figure 2, a simplified schematic diagram of an eXtended Reality system is presented. On the left hand side, the user is performing a task by using an XR application. In section 5, a complete overview of all the relevant domains is given covering advertisement, cultural heritage, education and training, industry 4.0, health and medicine, security, journalism, social VR and tourism. The user interacts with the scene and his interaction is captured with a range of input devices and sensors, which can be visual, audio, motion, and many more (see sec 4.1 and 4.2.). The acquired data serves as input for the XR hardware where further necessary processing in the render engine is performed (see sec. 4.7). For example, the correct view point is rendered or the desired interaction with the scene is triggered. In sec. 4.3 and 4.4, an overview of the major algorithms and approaches is given. However, not only captured data is used in the render engine, but also additional data that comes from other sources such as edge cloud servers (see sec. 4.8) or 3D data available on the device itself. The rendered scene is then fed back to the user to allow him sensing the scene. This is achieved by various means such as XR headsets or other types of displays and other sensorial stimuli.

The complete set of technologies and applications will be described in the following chapters.

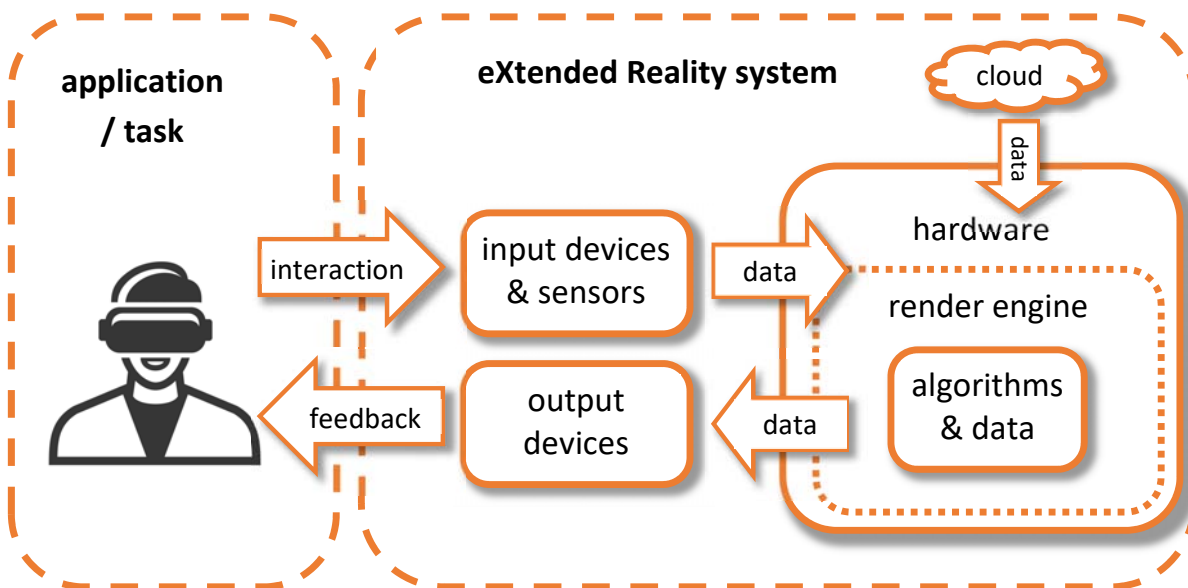


Figure 2: Major components of an eXtended Reality system.

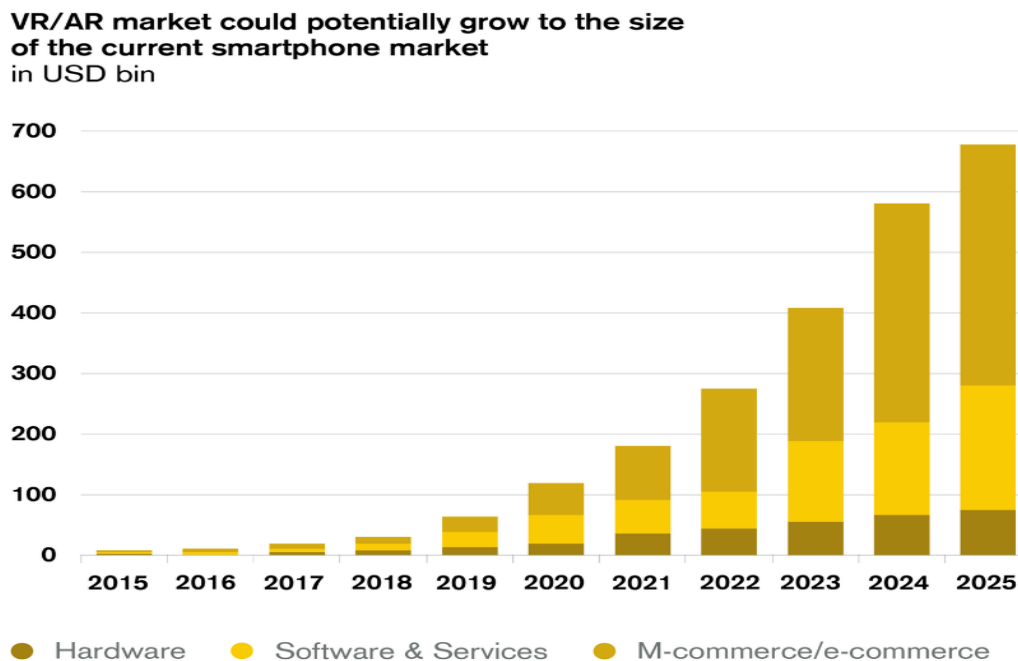
2.1 References

- [1] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented Reality: A class of displays on the reality-virtuality continuum", *Proc. SPIE vol. 2351, Telemanipulator and Telepresence Technologies*, pp. 2351–34, 1994.
- [2] Ronald T. Azuma, "A Survey of Augmented Reality", *Presence: Teleoperators and Virtual Environments*, vol. 6, issue 4, pp. 355-385, 1997.

3 XR market watch

3.1 Market development and forecast

Market research experts all agree on the tremendous growth potential for the XR market. The global AR and VR market by device, offering, application, and vertical, was valued at around USD 26.7 billion in 2018 by Zion Market Research. According to the report issued in February 2019, the global market is expected to reach approximately USD 814.7 billion by 2025, at a compound annual growth rate (CAGR) of 63.01% between 2019 and 2025 [3]. With over 65% in a forecast period from 2019 to 2024, similar annual growth rates are expected by Mordorintelligence [4]. It is assumed that the convergence of smartphones, mobile VR headsets, and AR glasses into a single XR wearable could replace all the other screens, ranging from mobile devices to smart TV screens. Mobile XR has the potential to become one of the world's most ubiquitous and disruptive computing platforms. Forecasts by MarketsandMarkets [5][6] individually expect the AR and VR markets by offering, device type, application, and geography, to reach USD 72.7 billion by 2024 (AR, valued at USD 10.7 billion in 2019) and USD 20.9 billion (VR, valued at USD 6.1 billion in 2020) by 2025. Gartner and Credit Suisse [7][8] predict significant market growth for VR & AR hardware and software due to promising opportunities across sectors up to 600-700 billion USD in 2025 (see Figure 3). With 762 million users owning an AR-compatible smartphone in July 2018, the AR consumer segment is expected to grow substantially, also fostered by AR development platforms such as ARKit (Apple) and ARCore (Google).



Source: Gartner, Credit Suisse

Figure 3: VR/AR market forecast by Gartner and Credit Suisse [7].

Several recent market studies including [6][9] have factored in the COVID-19 impact - yet to fully manifest itself - identifying growth drivers and barriers. Technavio forecasts a CAGR of over 35% and a market

growth of \$ 125.19 billion during 2020-2024 [10]. Growth driving factors are identified as the increasing demand for AR/VR technology, e.g. an increasing demand for VR/AR HMDs in the healthcare sector [9][11], or in general for remote work [12] and socializing [13]. Barriers on the other hand are associated with the prevailing potential high cost of XR app development [10], and COVID-19 adversely impacting the supply chain of the markets [6] [13][14], among others.

Regionally, the annual growth rate will be particularly high in Asia, moderate in North America and Europe, and low in other regions of the world [4][9] (see Figure 4). MarketsandMarkets finds Asia to lead the VR market by 2024 [6], and to lead the AR market by 2025 [5], whereas the US is still dominating the XR market with the large number of global players during the forecast period.

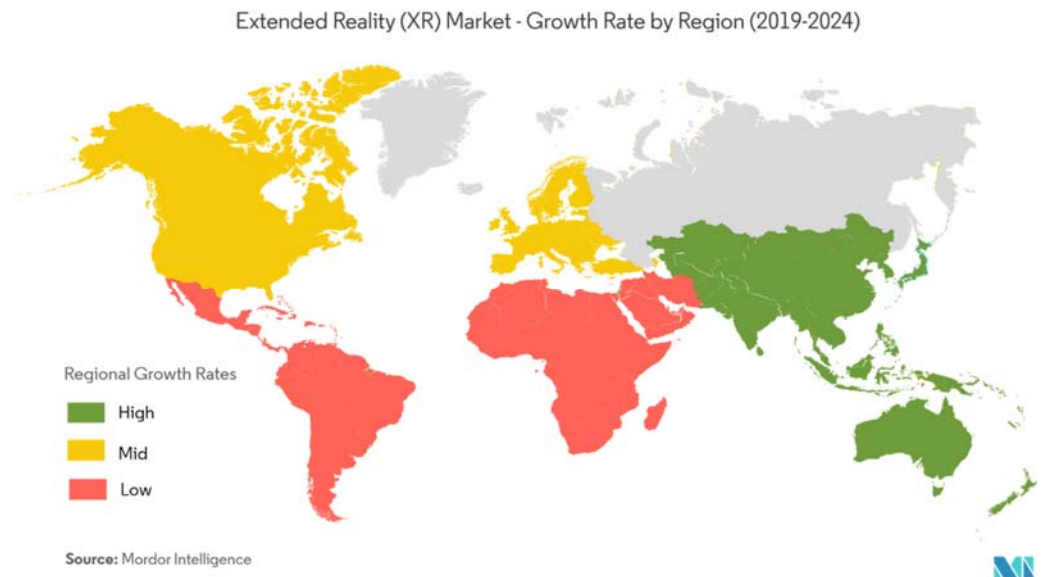
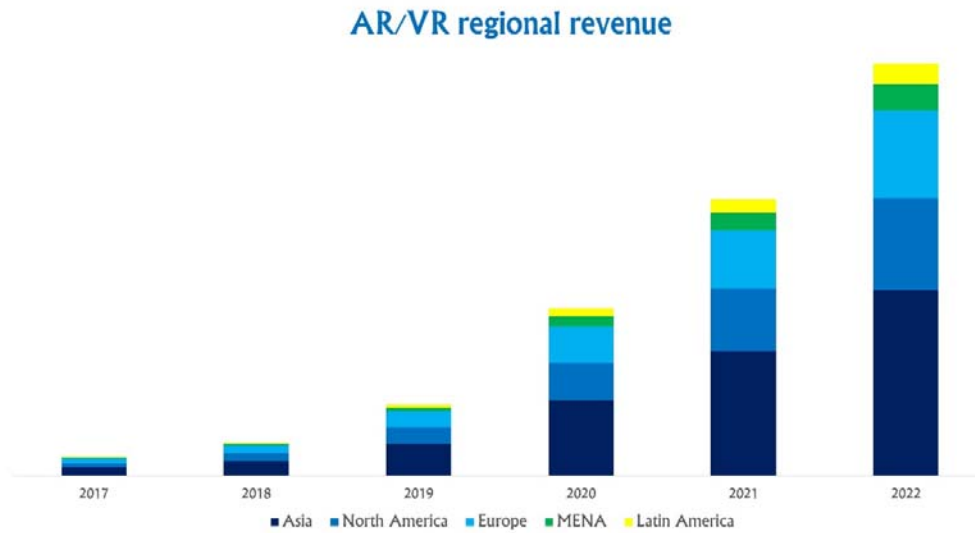


Figure 4: Market growth rates by worldwide regions [4].

With the XR market growing exponentially, Europe accounts for about one fifth of the market in 2022 [15], with Asia as the leading region (mainly China, Japan, and South Korea) followed by North America and Europe at almost the same level (see Figure 5). The enquiry in [16] sees Europe in 2023 even at second position of worldwide revenue regions (25%) after Asia (51%) followed by North America (17%). In a study about the VR and AR ecosystem in Europe in 2016/2017 [17], Ecorys identified the potential for Europe when playing out its strengths, namely building on its creativity, skills, and cultural diversity. Leading countries in VR development include France, the UK, Germany, The Netherlands, Sweden, Spain, and Switzerland. A lot of potential is seen for Finland, Denmark, Italy, Greece as well as Central and Eastern Europe. In 2017, more than half of the European companies had suppliers and customers from around the world.



© 2018 Digi-Capital. All rights reserved. No publication, adaptation, modification, reproduction or compilation without written permission from Digi-Capital

Figure 5: AR/VR regional revenue between 2017 and 2022 [15].

PwC released a study about the impact of AR and VR on the global economy by 2030 [18] highlighting the development in several countries. Globally, AR has a higher contribution to gross domestic product (GDP) than VR. The USA is expected to have the highest boost to GDP by 2030, followed by China and Japan (Figure 6).

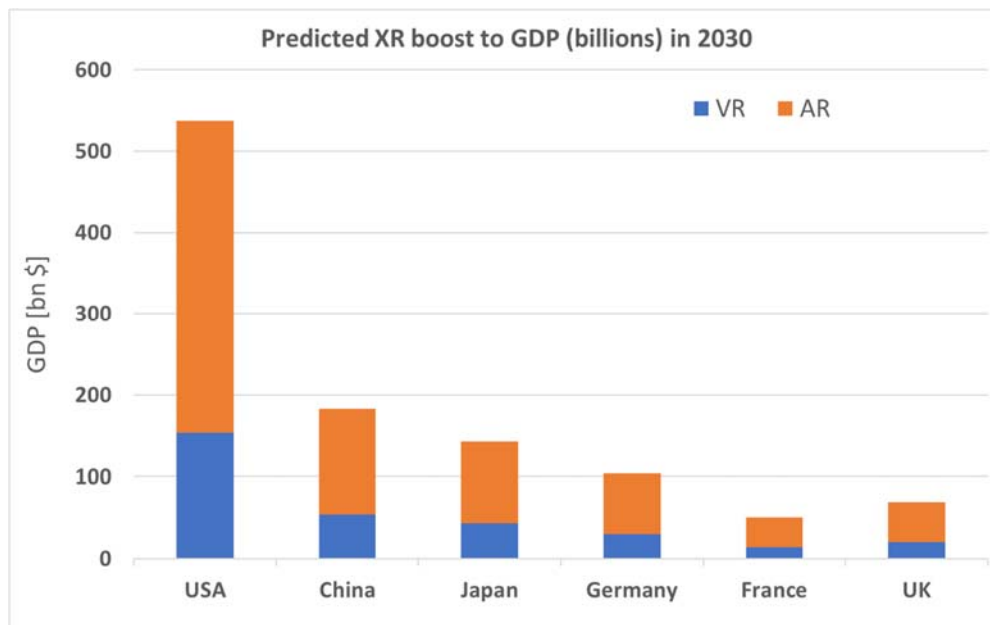


Figure 6: Global XR boost to gross domestic product by 2030 [18].

Concerning the major European countries Germany, France and UK, the major XR boost is expected for Germany followed by France and UK (see Figure 7).

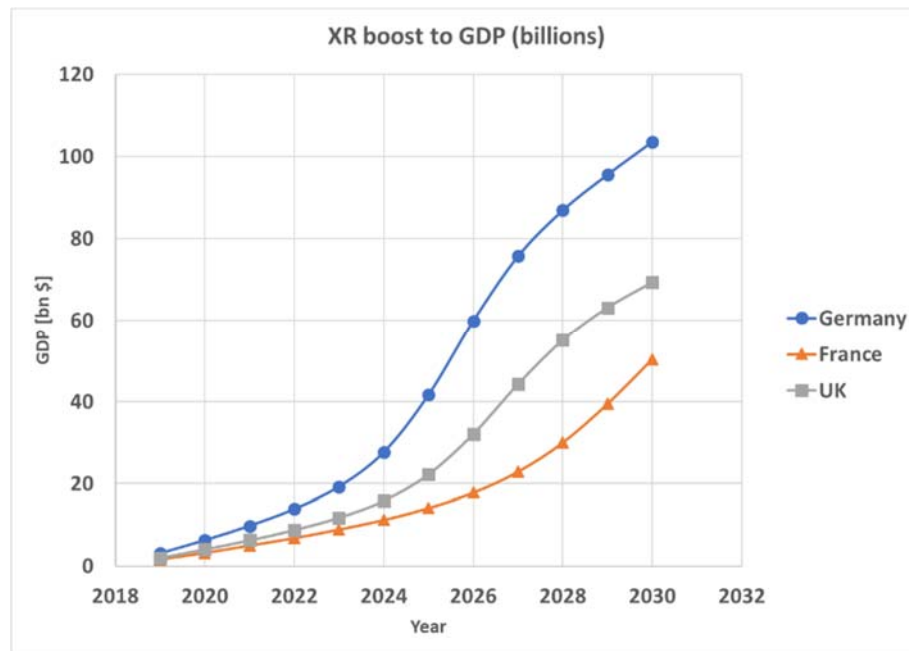


Figure 7: Major EU countries XR boost to gross domestic product from 2019 to 2030 [18].

The impact on employment through XR technology adoption will result in a major growth worldwide considering job enhancement (see Figure 8). From nearly 825 000 jobs enhanced in 2019 a rise is expected to more than 23 billion in 2030 worldwide [18]. China outnumbers all other countries by total numbers. Considering the share of jobs enhanced, the USA, UK and Germany are among the countries expected to experience the largest boost.

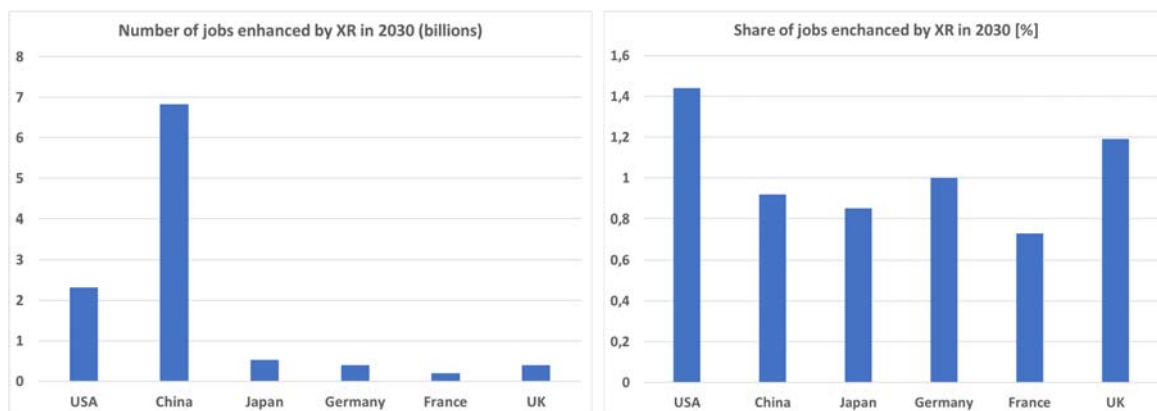
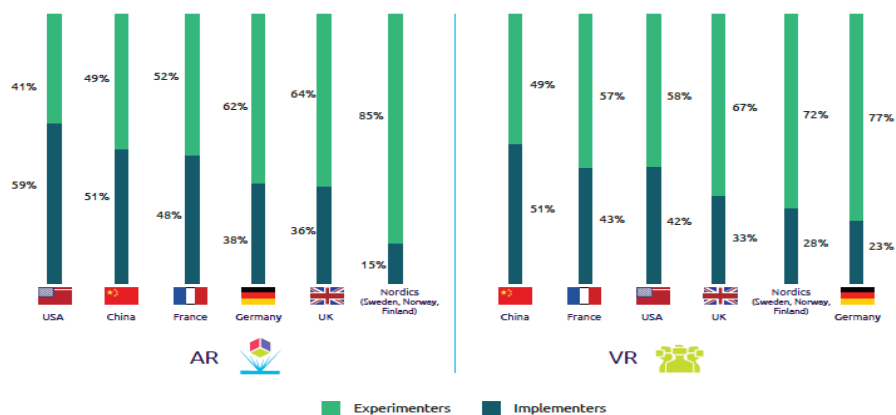


Figure 8: Job enhancement by 2030 through XR [18].

3.2 Areas of application

Within the field of business operations and field services, AR/VR implementations are found to be prevalent in four areas, where repair and maintenance have the strongest focus, closely followed by design and assembly. Other popular areas of implementation cover immersive training, and inspection and quality assurance [19]. Benefits from implementing AR/VR technologies include substantial increases in efficiency, safety, productivity, and reduction in complexity.

In a survey conducted in 2018 [19], Capgemini Research Institute focused on the use of AR/VR in business operations and field services in the automotive, manufacturing, and utilities sectors; companies considered were located in the US (30%), Germany, UK, France, China (each 15%) and the Nordics (Sweden, Norway, Finland). They found that, among 600+ companies with AR/VR initiatives (experimenting or implementing AR/VR), about half of them expects that AR/VR will become mainstream in their organisation within the next three years, the other half predominantly expects that AR/VR will become mainstream in less than five years. AR hereby is seen as more applicable than VR; consequently, more organisations are implementing AR (45%) than VR (36%). Companies in the US, China, and France are currently leading in implementing AR and VR technologies (see Figure 9). All European countries have less or equal implementers in AR and VR compared to US and China. A diagram relating US and China vs. Europe is not available.



Source: Capgemini Research Institute, Augmented and Virtual Reality Survey, May-June 2018, N=603 organizations that are exploring and implementing Augmented Reality and Virtual Reality. Implementers: companies with small or large-scale implementations; Experimenters: companies with proof of concepts or pilots.

Figure 9: Distribution of VR/AR companies analysed in survey by Capgemini Research [19].

The early adopters of XR technologies in Europe are in the automotive, aviation, and machinery sectors, but the medical sector plays also an important role. R&D focuses on health-care, industrial use and general advancements of this technology [19]. Highly specialised research hubs support the European market growth in advancing VR technology and applications and also generate a highly-skilled workforce, bringing non-European companies to Europe for R&D. Content-wise, the US market is focused on entertainment while Asia is active in content production for the local markets. Europe benefits from its cultural diversity and a tradition of collaboration, in part fostered by European funding policies, leading to very creative content production.

It is also interesting to compare VR and AR with respect to the field of applications (see Figure 10). Due to a smaller installed base, lower mobility and exclusive immersion, VR will be more focussed on entertainment use cases and revenue streams such as in games, location-based entertainment, video, and related hardware, whereas AR will be more based on e-commerce, advertisement, enterprise applications, and related hardware [8].

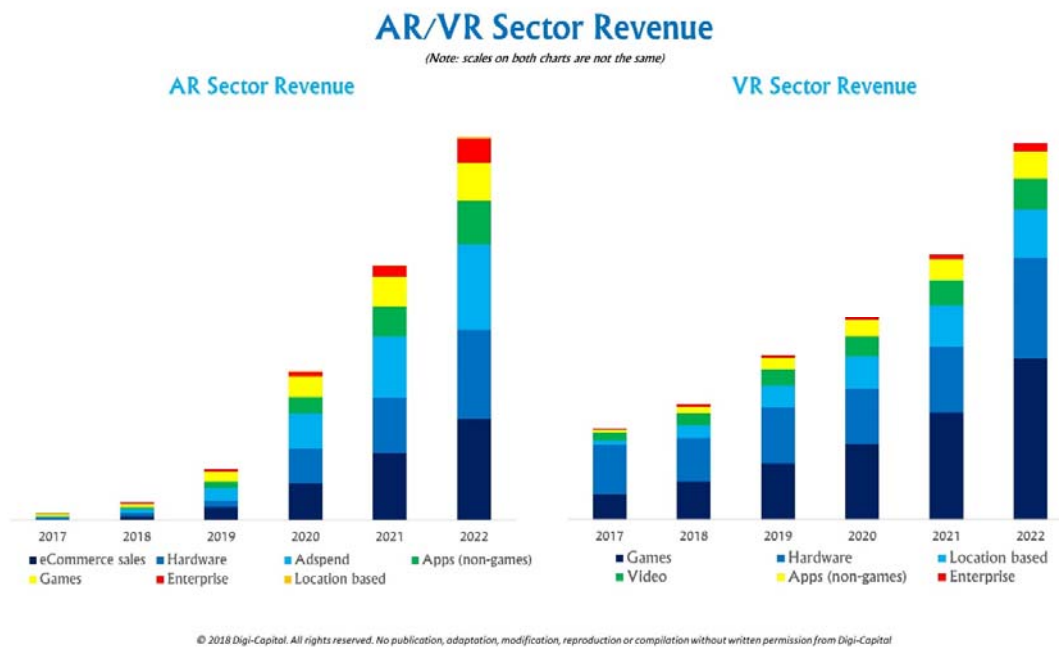


Figure 10: Separated AR and VR sector revenue from 2017 to 2022 [8].

Note: This diagram represents just relations of revenue between different sectors, the scale for AR and VR are not the same.

A PwC analysis [18] groups major use cases into five categories:

- 1) Product and service development;
- 2) Healthcare;
- 3) Development and training;
- 4) Process improvements;
- 5) Retail and consumer.

Among those, XR technologies for product and service development as well as healthcare are expected to have the highest impact with a potential boost to GDP of over \$350 billion by 2030 (see Figure 11).

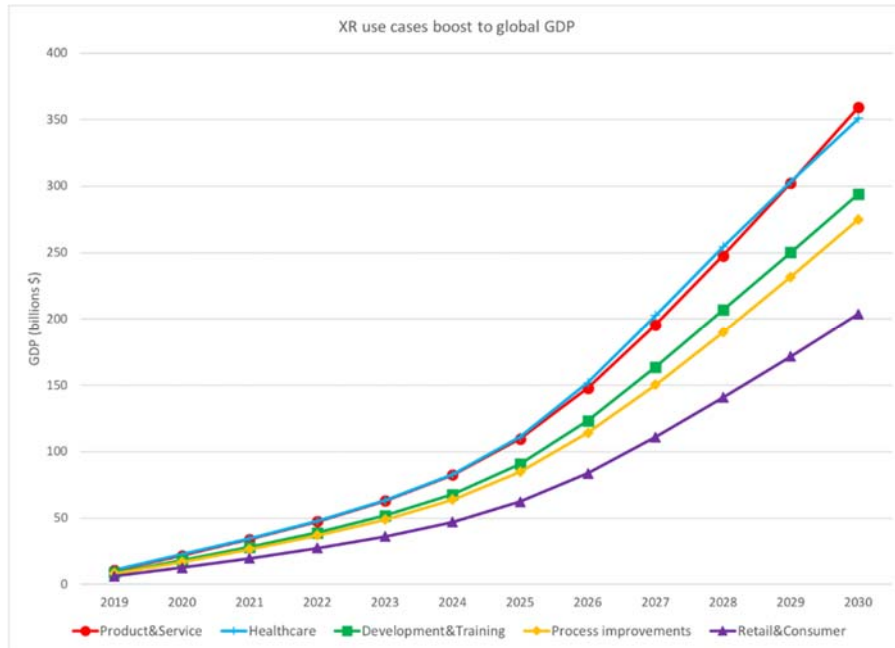


Figure 11: XR use cases boost to gross domestic product from 2019 to 2030 [18].

3.3 Investments

While XR industries are characterised by global value chains, it is important to be aware of the different types of investments available and of the cultural settings present. Favourable conditions for AR/VR start-ups are given in the US through the availability of venture capital towards early technology development. The Asian market growth is driven through concerted government efforts. Digi-Capital has tracked over \$5.4 billion XR investments in the 12 months from Q3 2018 to Q2 2019 showing that Chinese companies have invested by a factor of 2.5 more than their North American counterparts during this period [20]. Investment considerably dropped worldwide over the last 12 months to Q1 2020 [21] with the US and China continuing to dominate XR investment, followed by Israel, the UK and Canada. In Europe, the availability of research funding fostered a tradition in XR research and the creation of niche and high-precision technologies. The XR4ALL Consortium has compiled a list of over 455 investors investing in XR start-ups in Europe [22]. The investments range from 2008-2019. A preliminary analysis shows that the verticals attracting the greatest numbers of investors are: Enterprise, User Input, Devices/Hardware, and 3D Reality Capture (see Figure 12).

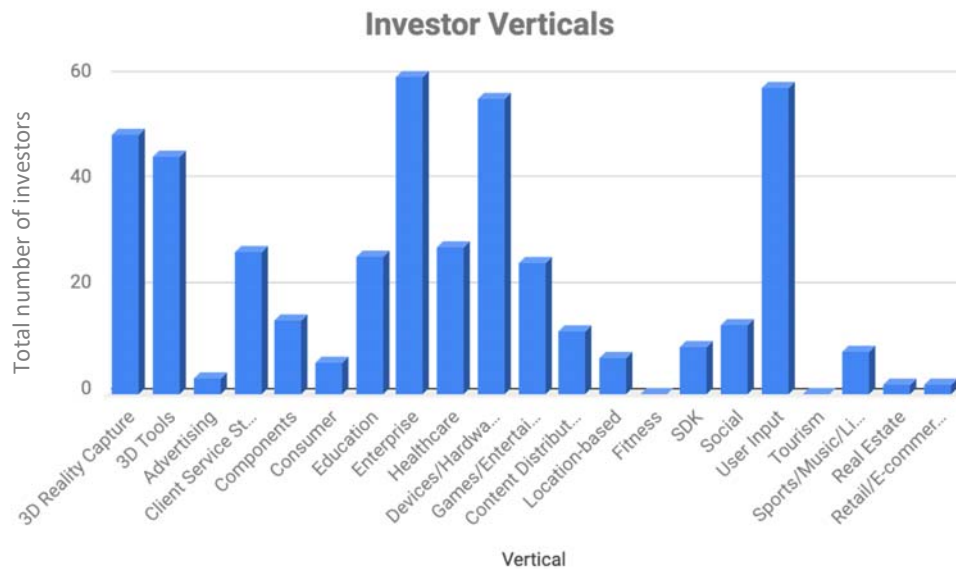


Figure 12: XR4ALL analysis of European investors for start-ups [22].

The use cases that are forecasted by IDC to receive the largest investment in 2023 are education/training (\$8.5 billion), industrial maintenance (\$4.3 billion), and retail showcasing (\$3.9 billion) [23]. A total of \$20.8 billion is expected to be invested in VR gaming, VR video/feature viewing, and AR gaming. The fastest spending growth is expected for the following: AR for lab and field education, AR for public infrastructure maintenance, and AR for anatomy diagnostic in the medical domain.

3.4 Shipment of devices

The shipment of VR headset has steadily been growing for several years and has reached a number of 4 million devices in 2018 [24]. It raised up to around 6 million in 2019 and is mainly dominated by North American companies (e.g. Facebook Oculus) and major Asian manufacturers (e.g. Sony, Samsung, and HTC Vive) (see Figure 13). The growth on the application side is even higher. For instance, at the gaming platform Steam, the yearly growing rate of monthly-connected headsets is up 80% since 2017 [25].

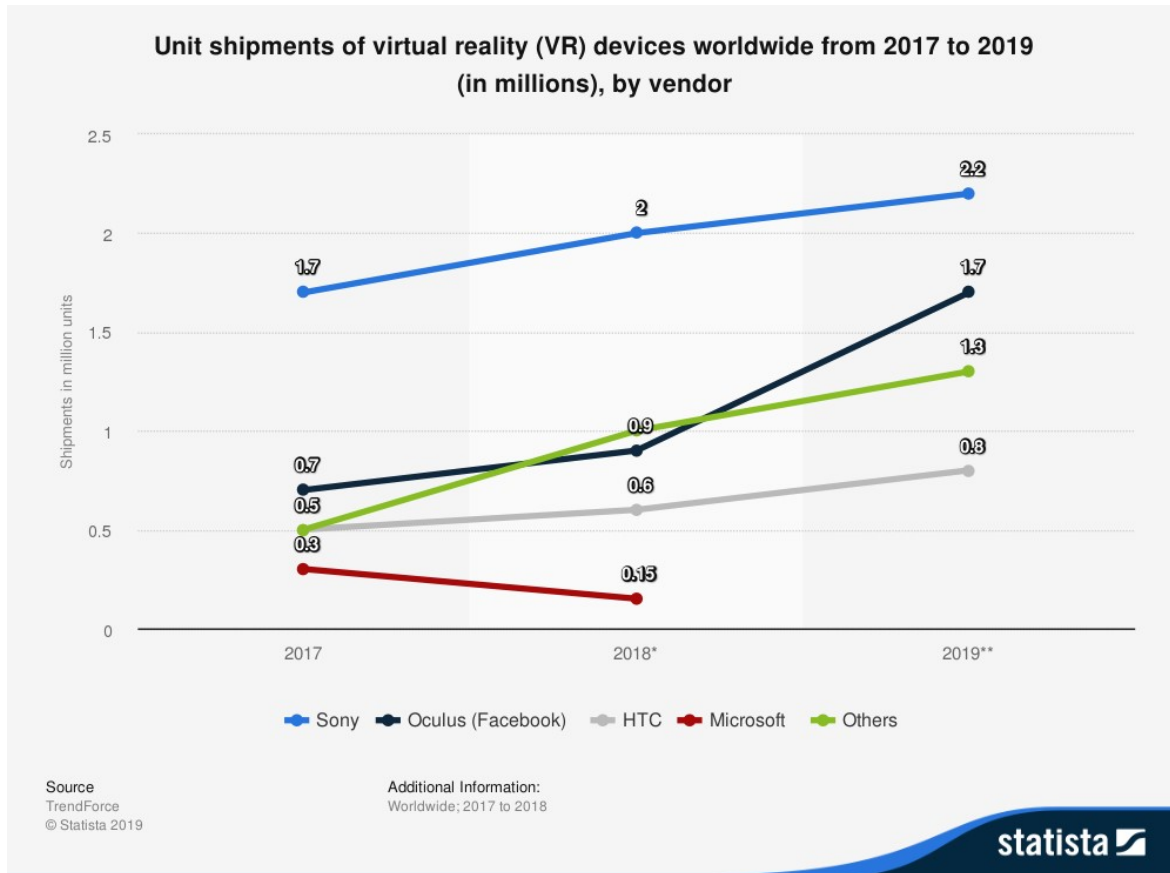


Figure 13: VR unit shipments in the last three years [24].

The situation is completely different for AR headsets. Compared to VR, the shipments of AR headsets in 2017 were much lower (less than 0.4 million), but the actual growing rate is much higher than for VR headsets [26] (see Figure 14). In 2019, the number of unit shipments was almost at the same level for AR and VR headsets (about 6 million), and, beyond 2019, it will be much higher for AR. This is certainly, due to the fact that there is a wider range of applications for AR than for VR (see also section 3.2).

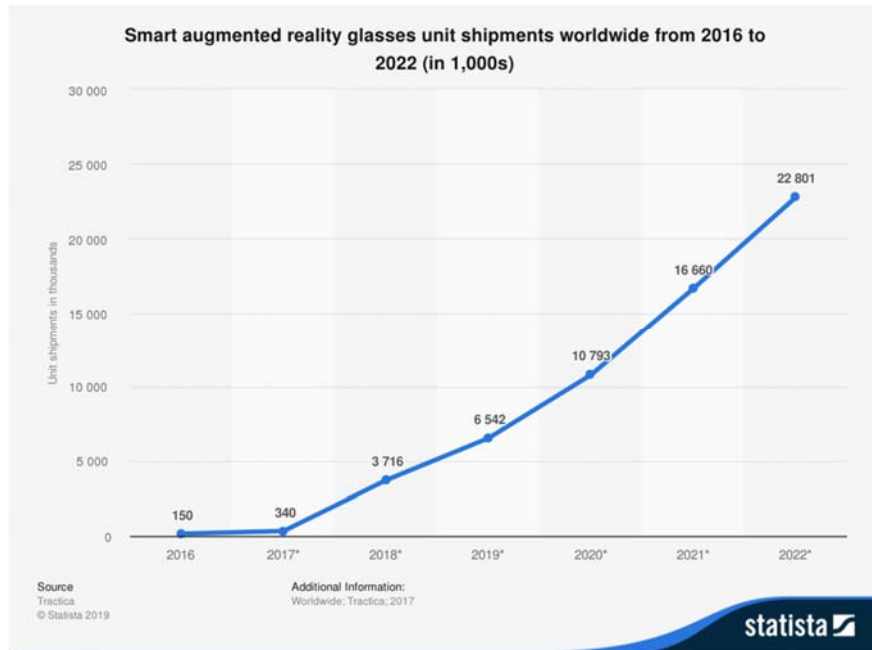


Figure 14: Forecast of AR unit shipments from 2016 to 2022 [26].

Shipment of VR and AR devices are expected to grow considerably from below 9 million devices in 2020 to more than 50 million devices by 2024 [12][14] (see Figure 15). However, the shipments of smartphone shell VR will decrease and only the shipments of AR, standalone VR and tethered VR devices will increase substantially. Especially, the growth of standalone VR devices seems to be predominant, since the first systems appeared on the market in 2018 and the global players like Oculus and HTC launched their solutions in 2019. ABC Research predicts that over 70% of VR shipments in 2024 will be standalone devices [12].

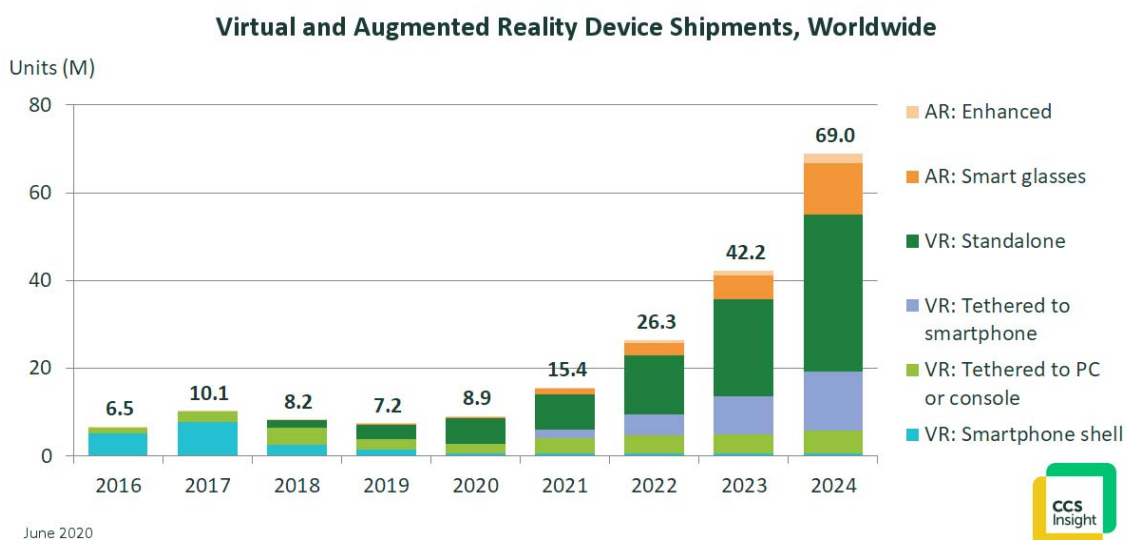


Figure 15: Forecast of VR and AR shipments [14].

Taking into account COVID-19 impacts, pre-COVID expectations for AR and VR shipments will be reached in 2024 [12] (see Figure 16).

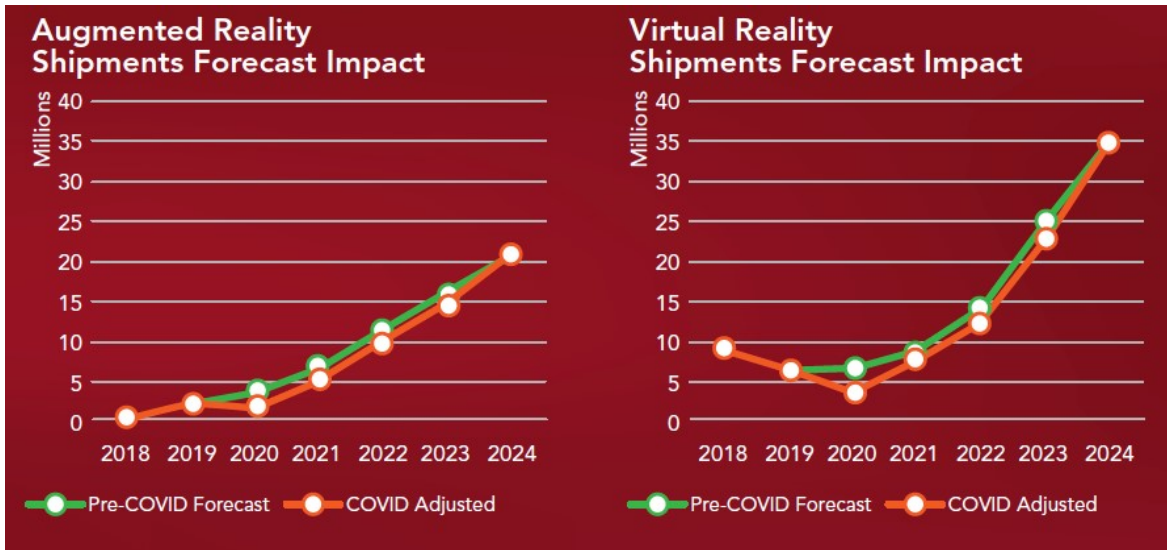


Figure 16: Forecast of VR and AR shipments with/without COVID-19 impact.

3.5 Main players

With a multitude of players from start-ups and SMEs to very large enterprises, the VR/AR market is fragmented [27], and dominated by US internet giants such as Google, Apple, Facebook, Amazon, and Microsoft. By contrast, European innovation in AR and VR is largely driven by SMEs and start-ups [17].

Main XR players [8][17] are from (1) the US (e.g., Google, Microsoft, Oculus, Eon Reality, Vuzix, CyberGlove Systems, Leap Motion, Sensics, Sixsense Enterprises, WorldViz, Firsthand Technologies, Virtuix, Merge Labs, SpaceVR), and (2) Asian Pacific region (e.g., Japan: Sony, Nintendo; South Korea: Samsung Electronics; Taiwan: HTC). Besides the main players, there are plenty of SMEs and smaller companies worldwide. Figure 17 gives a good overview of the AR industry landscape, while in Figure 18, the current VR industry landscape is depicted.

THE VR FUND 2019 AR INDUSTRY LANDSCAPE

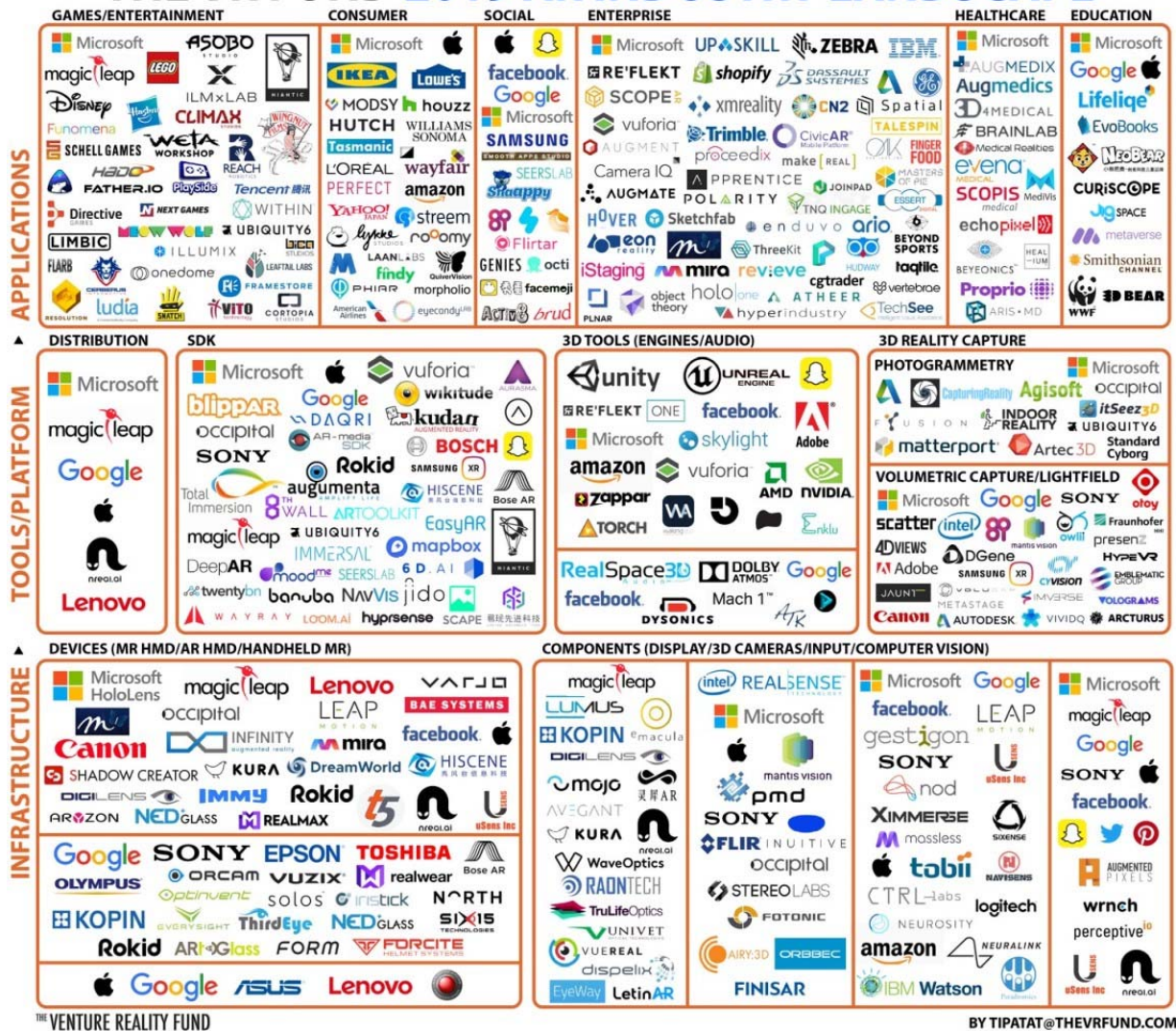


Figure 17: AR Industry Landscape by Venture Reality Fund [28].

THE VR FUND 2019 VR INDUSTRY LANDSCAPE

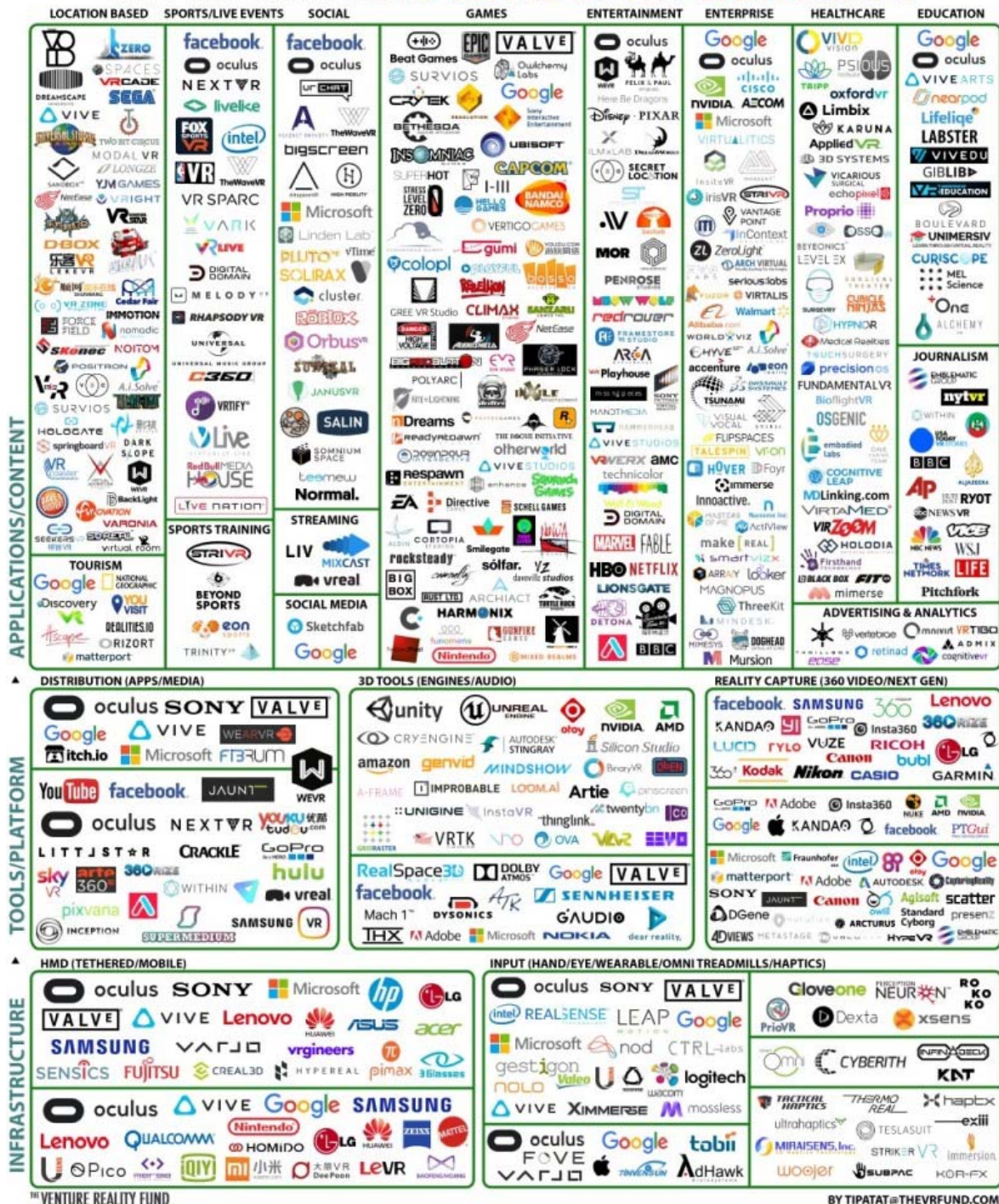


Figure 18: VR Industry Landscape by Venture Reality Fund [28].

In addition to the above corporate activities, Europe also has a long-standing tradition in research [17]. Fundamental questions are generally pursued by European universities such as ParisTech (FR), Technical University of Munich (DE), and King's College (UK) and by non-university research institutes like B&W (FR), Fraunhofer Society (DE), and INRIA (FR). Applied research is also relevant, and this is also true for the creative sector. An important part is also played by associations, think tanks, associations and institutions such as EuroXR, Realities Centre (UK), VRBase (NL/DE) and Station F (FR) that connect stakeholders, provide support, and enable knowledge transfer. Research activities tend to concentrate in France, the UK, and Germany, while business activities tend to concentrate in France, Germany, the UK, and The Netherlands.

The VR Fund published the VR/AR industry landscapes [28] providing a good overview of industry players. Besides some of the companies already mentioned, one finds other well-known European XR companies such as: Ultrahaptics (UK), Improbable (UK), Varjo (FI), Meero (FR), CCP Games (IS), Immersive Rehab (UK), and Pupil Labs (DE). Others are Jungle VR, Light & Shadows, Lumiscaphe, Thales, Techviz, Immersion, Haption, Backlight, ac3 studio, ARTE, Diota, TF1, Allegorithmic, Saint-Gobain, Diakse, Wonda, Art of Corner, Incarna, Okio studios, Novelab, Timescope, Adok, Hypersuit, Realtime Robotics, Wepulsit, Holostoria, Artify, VR-bnb, Hololamp (France), and many more.

3.6 International, European and regional associations in XR

There are several associations worldwide, in Europe, but also on regional level, that aim to foster the development of XR technology. The major associations are shortly described below.

3.6.1 International

XR Association (XRA)

The XRA's mission is to promote responsible development and adoption of virtual and augmented reality globally with best practices, dialogue across stakeholders, and research [29]. The XRA is a resource for industry, consumers, and policymakers interested in virtual and augmented reality. XRA is an evolution of the Global Virtual Reality Association (GVRA). This association is very much industry-driven due to the memberships of Google, Microsoft, Facebook (Oculus), Sony Interactive Entertainment (PlayStation VR) and HTC (Vive).

VR/AR Association (VRARA)

The VR/AR Association is an international organisation designed to foster collaboration between innovative companies and people in the VR and AR ecosystem that accelerates growth, fosters research and education, helps develop industry standards, connects member organisations and promotes the services of member companies [30]. The association states over 400 organisations registered as members. VRARA has regional chapters in many countries around the globe.

VR Industry Forum (VRIF)

The Virtual Reality Industry Forum [31] is composed of a broad range of participants from sectors including, but not limited to, movies, television, broadcast, mobile, and interactive gaming ecosystems, comprising content creators, content distributors, consumer electronics manufacturers, professional equipment manufacturers and technology companies. Membership in the VR Industry Forum is open to all parties that support the purposes of the VR Industry Forum. The VR Industry Forum is not a standards

development organisation, but will rely on, and liaise with, standards development organisations for the development of standards in support of VR services and devices. Adoption of any of the work products of the VR Industry Forum is voluntary; none of the work products of the VR Industry Forum shall be binding on Members or third parties.

THE AREA

The Augmented Reality for Enterprise Alliance (AREA) presents itself as the only global non-profit, member-driven organisation focused on reducing barriers to and accelerating the smooth introduction and widespread adoption of Augmented Reality by and for professionals [32]. The mission of the AREA is to help companies in all parts of the ecosystem to achieve greater operational efficiency through the smooth introduction and widespread adoption of interoperable AR-assisted enterprise systems.

International Virtual Reality Professionals Association (IVRPA)

The IVRPA mission is to promote the success of Professional VR Photographers and Videographers [33]. We strive to develop and support the professional and artistic uses of 360° panoramas, image-based VR and related technologies worldwide through education, networking opportunities, manufacturer alliances, marketing assistance, and technical support of our member's work. The association currently consists of more than 500 members, either individuals or companies spread among the whole world.

The Academy of International Extended Reality (AIXR)

The AIXR is an international network with strong support by leading small and big companies in the immersive media domain [34]. The aim is connecting people, projects, and knowledge together and enable growth, nurture talent, and develop standards, bringing wider public awareness and understanding to the international VR & AR industry. A number of advisory groups in different application and technology domains perform focused discussion to foster the progress on their topic.

MedVR

MedVR is an international network dedicated to the healthcare sector [35]. The aim is to bring together clinicians, scientists, developers, designers, and other experts into interdisciplinary teams to lead the future of augmented and virtual reality (AR & VR) in healthcare. The goal is to educate, stimulate discussion, identify novel applications, and build cutting-edge prototypes.

Open AR Cloud Association (OARC)

The "Open AR Cloud Association" (OARC) is a global non-profit organization registered in Delaware, USA [36]. Its mission is to drive the development of open and interoperable spatial computing technology, data and standards to connect the physical and digital worlds for the benefit of all.

3.6.2 European

EuroXR

EuroXR is an International non-profit Association [37], which provides a network for all those interested in Virtual Reality (VR) and Augmented Reality (AR) to meet, discuss and promote all topics related to VR/AR technologies. EuroXR (EuroVR) was founded in 2010 as a continuation of the work in the FP6 Network of Excellence INTUITION (2004 – 2008). The main activity is the organisation of the EuroXR annual event. This series was initiated in 2004 by the INTUITION Network Excellence in Virtual and Augmented Reality, supported by the European Commission until 2008, and incorporated within the Joint Virtual

Reality Conferences (JVRC) from 2009 to 2013. Beside individual membership, several organisational members are part of EuroXR such as AVRLab, Barco, List CEA Tech, AFVR, GoTouchVR Haption, catapult, Laval Virtual, VTT, Fraunhofer FIT and Fraunhofer IAO and some European universities.

Extended Reality for Education and Research in Academia (XR ERA)

XR ERA was recently founded in 2020 by Leiden University, Centre for Innovation [38]. The aim is to bring people from education, research and industry together, both online and offline to enhance education and research in academia by making use of what XR has to offer.

Women in Immersive Technologies Europe (WiiT Europe)

WiiT Europe is a European non-profit organization that aims to empower women by promoting diversity, equality and inclusion in VR, AR, MR and other future immersive technologies [39]. Started in 2016 as a Facebook group, WiiT Europe is an inclusive network of talented women who are driving Europe's XR sectors.

3.6.3 National

ERSTER DEUTSCHER FACHVERBAND FÜR VIRTUAL REALITY (EDFVR)

The EDFVR is the first German business association for immersive media [40]. Start-ups and established entrepreneurs, enthusiasts and developers from Germany are joined together to foster immersive media in Germany.

Virtual Reality e.V. Berlin Brandenburg (VRBB)

VRBB is a publicly-funded association dedicated to advancing the virtual, augmented and mixed reality industries [41]. The association was founded in 2016 and its members are HighTech companies, established Media Companies, Research Institutes and Universities, Start-Ups, Freelancers and plain VR enthusiasts. The VRBB organises a yearly event named VRNowCon since 2016, which has an international reach of participants.

Virtual Dimension Center (VDC)

VDC considers itself as the largest B2B network for XR technologies in Germany [42]. It was founded in 2020 and consists of currently 90 members from industry, IT, research and higher-education. The focus is on Virtual Engineering, Virtual Reality and 3D-simulation. The VDC offers a communication platform for members, a knowledge database, networking, and support for funding acquisition.

Virtual and Augmented Reality Association Austria (VARAA)

VARAA is the independent association of professional VR/AR users and companies in Austria [43]. The aim is to promote, raise awareness and support in handling VR/AR. The association represents the interests of the industry and links professional users and developers. Through a strong network of partners and industry contacts it is the single point of contact in Austria to the international VR/AR scene and the global VR/AR Association (VRARA Global).

AFXR (France)

The AFXR was born from the merger in 2019 of two major French associations AFVR and Uni-XR [44]. It aims to bring together the community of French professionals working in immersive technologies or using XR technologies. The association is neutral, non-commercial and is not affiliated to any economic, territorial or political body. It has over 200 members.

Virtual Reality Finland

The goal of the association is to help Finland become a leading country in VR and AR technologies [45]. The association is open to everyone interested in VR and AR. The association organises events, supports VR and AR projects and shares information on the state and development of the ecosystem.

Finnish Virtual Reality Association (FIVR)

The purpose of the Finnish Virtual Reality Association is to advance virtual reality (VR) and augmented reality (AR) development and related activities in Finland [46]. The association is for professionals and hobbyists of virtual reality. FIVR is a non-profit organisation dedicated to advancing the state of Virtual, Augmented and Mixed Reality development in Finland. The goal is to make Finland a world leading environment in XR activities. This happens by establishing a multidisciplinary and tightly-knit developer community and a complete, top-quality development ecosystem, which combines the best resources, knowledge, innovation and strength of the public and private sectors.

XR Nation (Finland)

Starting in the spring of 2018 in Helsinki, Finland, XR Nation's goal has always been to bring the AR & VR communities in the Nordics and Baltic region closer together [47]. XR Nation counts 500+ members and 80+ companies.

VIRTUAL SWITZERLAND

This Swiss association has more than 60 members from academia and industry [48]. It promotes immersive technologies and simulation of virtual environments (XR), their developments and implementation. It aims to foster research-based innovation projects, dialogue and knowledge exchange between academic and industrial players across all economic sectors. It gathers minds and creates links to foster ideas via its nation-wide professional network and facilitates the genesis of projects and their applications to Innosuisse for funding opportunities.

Immerse UK

Immerse UK is the UK's leading membership organisation for immersive technologies [49]. They bring together industry, research and academic organisations, public sector and innovators to help fast-track innovation, R&D, scalability and company growth. They are the UK's only membership organisation dedicated to supporting content, applications, services and solution providers developing immersive technology solutions or companies creating content or experiences using immersive tech.

VRINN (Norway)

VRINN is a cluster of companies operating in Norway in the fields of VR, AR, and gamification [50]. The aim of the cluster is to offer its members a platform to exchange ideas, develop projects and thus jointly advance the development of future learning. VRINN also helps companies to network internationally, to market themselves and to develop further. Since 2017, VRINN organizes the VR Nordic Forum, a conference focusing on "immersive learning technologies" – the use of VR & AR in learning, training and storytelling. With 750 participants gathered during the last edition in October 2020, VR Nordic Forum is the biggest XR event in northern Europe.

3.7 Patents

The number of patents filed is a useful indicator for technology development. In a working paper by Eurofound the years 2010 for AR and 2014 for VR are identified as the starting years for increased patent activity [51]. Analysing patent data until 2017, the USA emerges as leader in patent applications, followed by China. The XR4ALL consortium recently carried out a study using the database available at the European Patent Office [52]. The database was searched for the period 2019 until today and the search was limited to the following keywords: Virtual Reality, Augmented Reality, immersive, eXtended Reality, Mixed Reality, haptic. The most relevant 50 European patents have been selected and listed in the table below. The publication dates in the rightmost column of the table below range between April 15th, 2019 and October 1st, 2020.

#	Applicant	Country	Title	Publication Date
1	Accenture Global Services Ltd.	IE	Virtual Reality Based Hotel Services Analysis and Procurement	15/04/2019
2	Accenture Global Services Ltd.	IE	Augmented Reality Based Component Replacement and Maintenance	02/05/2019
3	Accenture Global Solutions Ltd.	IE	Augmented Reality Enabled Cargo Loading Optimization	02/09/2020
4	Accenture Global Solutions Ltd.; Univ Tehnica Din Cluj Napoca	IE, RO	Method For Viewing The Path Of An Autonomous Vehicle Using Augmented Reality	29/11/2019
5	Accenture Global Solutions Ltd.	IE	Real-Time Motion Feedback For Extended Reality	06/05/2020
6	Airbus Operations S.L.U	ES	A Real Time Virtual Reality (VR) System And Related Methods	04/03/2020
7	Aldin Dynamics Ehf.	IS	Methods and Systems for Path-Based Locomotion in Virtual Reality	25/04/2019
8	Alexandra HUSSENOT DESENONGES	GB	Mixed Reality Handsfree Motion	25/12/2019
9	Arkio Ehf.	IS	Virtual/Augmented Reality Modeling Application for Architecture	16/05/2019
10	ARM IP Ltd.	GB	Image Processing For Augmented Reality	16/10/2019
11	Atos Integration	FR	System for Composing or Modifying Virtual Reality Sequences, Method of Composing and System for Reading Said Sequences	06/06/2019

12	Audi AG.	DE	Driver Assistance System And Method For A Motor Vehicle For Displaying Augmented Reality Displays	16/09/2020
13	Bavastro Frederic	MC	Augmented Reality Method and System for Design	30/05/2019
14	Bossut Christophe; Le Henaff Guy; Chapelain De La Villeguerin Yves	FR, FR, PT	System and Method for Providing Augmented Reality Interactions over Printed Media	16/05/2019
15	Curious Lab Tech Ltd.	GB	Method and System for Generating Virtual or Augmented Reality	28/06/2019
16	Eaton Intelligent Power Ltd.	IE	Lighting and Internet of Things Design Using Augmented Reality	20/06/2019
17	Here Global B.V.	NL	Method And Apparatus For Augmented Reality Based On Localization And Environmental Conditions	17/06/2020
18	Here Global B.V.	NL	Location Enabled Augmented Reality (AR) System And Method For Interoperability Of AR Applications	22/07/2020
19	Interdigital Ce Patent Holdings	FR	Sharing Virtual Content In A Mixed Reality Scene	25/12/2019
20	Interdigital Ce Patent Holdings	FR	A System For Controlling Audio-Capable Connected Devices In Mixed Reality Environments	18/03/2020
21	Interdigital Ce Patent Holdings	FR	A Method And Apparatus For Encoding And Decoding Volumetric Video	30/09/2020
22	Institut Nat Des Sciences Appliquees De Rennes; Orange	FR, FR	Virtual Reality Data-Processing Device, System And Method	24/09/2020
23	Luminous Group Ltd.	GB	Mixed Reality System	19/08/2020
24	Medical Realities Ltd.	GB	Virtual Reality System for Surgical Training	23/05/2019
25	Metatellus Oue	EE	Augmented Reality Based Social Platform	23/05/2019
26	Nokia Technologies	FI	Provision of Virtual Reality Content	09/05/2019
27	Nokia Technologies	FI	Apparatus and Associated Methods for Presentation of First and Second Virtual-or-Augmented Reality Content	13/06/2019
28	Nokia Technologies	FI	Apparatus and Associated Methods for Presentation of Augmented Reality Content	13/06/2019

29	Nokia Technologies	FI	Virtual Reality Device and a Virtual Reality Server	27/06/2019
30	Nokia Technologies	FI	Method, System And Apparatus For Collaborative Augmented Reality	23/10/2019
31	Nokia Technologies	FI	Method And Apparatus For Adding Interactive Objects To A Virtual Reality Environment	29/01/2020
32	Nokia Technologies	FI	An Apparatus And Associated Methods For Presentation Of A Virtual Reality Space	15/07/2020
33	Nokia Technologies	FI	A Method, An Apparatus And A Computer Program Product For Virtual Reality	15/07/2020
34	Przed Produkcjino Usługowe Stolgraf Pasternak Rodziewicz Spolka Jawna	PL	A System And A Method For Generating A Virtual Reality Environment For Exercises Via A Wearable Display	11/03/2020
35	Roto VR Ltd.	GB	Virtual Reality Apparatus	13/06/2019
36	Siemens AG.	DE	Display of Three-Dimensional Model Information in Virtual Reality	13/06/2019
37	Siemens AG.	DE	Postures Recognition Of Objects In Augmented Reality Applications	05/08/2020
38	Siemens AG.	DE	Direct Volume Haptic Rendering	19/08/2020
39	Siemens Healthcare GmbH	DE	Method And Device To Control A Virtual Reality Display Unit	24/06/2020
40	Signify Holding B.V.	NL	Augmented Reality- Based Acoustic Performance Analysis	01/10/2020
41	Somo Innovations Ltd.	GB	Augmented Reality with Graphics Rendering Controlled by Mobile Device Position	23/05/2019
42	Stoecker Carsten; Innogy Innovation GmbH.	DE, DE	Augmented Reality System	25/04/2019
43	Tsapakis Stylianos Georgios	GR	Virtual Reality Set	24/05/2019
44	Tobii AB	SE	Eye Tracking Application In Virtual Reality And Augmented Reality	04/12/2019
45	Thomson Licensing	FR	Stray Light Resistant Augmented Reality Device	24/06/2020

46	Unity lpr Ap.	DK	Method and System for Synchronizing a Plurality of Augmented Reality Devices to a Virtual Reality Device	27/06/2019
47	Univ Muenchen Tech	DE	Method And Control Unit For Controlling A Virtual Reality Display, Virtual Reality Display And Virtual Reality System	22/01/2020
48	Wave Optics Ltd.	GB	Device For Augmented Reality Or Virtual Reality Display	06/05/2020
49	Wave Optics Ltd.	GB	Optical Structure For Augmented Reality Display	28/08/2020
50	Wave Optics Ltd.	GB	Improved Angular Uniformity Waveguide For Augmented Or Virtual Reality	24/09/2020

3.8 References

- [3] Zion Market Research. <https://www.zionmarketresearch.com/report/augmented-and-virtual-reality-market> (accessed Nov. 11, 2020).
- [4] “Extended Reality (XR) Market - Growth, trends, and forecast.” Mordor Intelligence. <https://www.mordorintelligence.com/industry-reports/extended-reality-xr-market> (accessed Nov. 11, 2020).
- [5] “Augmented Reality Market worth \$72.7 billion by 2024.” Marketsandmarkets. <https://www.marketsandmarkets.com/PressReleases/augmented-reality.asp> (accessed Nov. 11, 2020).
- [6] “Virtual Reality Market worth \$20.9 billion by 2025.” Marketsandmarkets. <https://www.marketsandmarkets.com/PressReleases/ar-market.asp> (accessed Nov. 11, 2020).
- [7] U. Neumann. “Virtual and Augmented Reality have great growth potential.” Credit Suisse. <https://www.credit-suisse.com/ch/en/articles/private-banking/virtual-und-augmented-reality-201706.html> (accessed Nov. 11, 2020).
- [8] U. Neumann. “Increased integration of augmented and virtual reality across industries.” Credit Suisse. <https://www.credit-suisse.com/ch/en/articles/private-banking/zunehmende-einbindung-von-Virtual-und-augmented-reality-in-allen-branchen-201906.html> (accessed Nov. 11, 2020).
- [9] Research and Markets. <https://www.researchandmarkets.com/reports/4746768/virtual-reality-market-by-offering-technology> (accessed Nov. 11, 2020).
- [10] Businesswire. <https://www.businesswire.com/news/home/20200903005356/en/COVID-19-Impacts-Augmented-Reality-AR-and-Virtual-Reality-VR-Market-Will-Accelerate-at-a-CAGR-of-Over-35-Through-2020-2024-The-Increasing-Demand-for-AR-and-VR-Technology-to-Boost-Growth-Technavio> (accessed Nov. 11, 2020).
- [11] “Impact analysis of covid-19 on augmented reality (AR) in healthcare market.” Researchdive. <https://www.researchdive.com/covid-19-insights/218/global-augmented-reality-ar-in-healthcare-market> (accessed Nov. 11, 2020).

- [12] “Augmented and Virtual Reality: Visualizing Potential Across Hardware, Software, and Services.” ABIresearch. <https://www.abiresearch.com/whitepapers/> (accessed Nov. 11, 2020).
- [13] Digi-Capital. <https://www.digi-capital.com/news/2020/04/how-covid-19-change-ar-vr-future/> (accessed Nov. 11, 2020).
- [14] M. Koytcheva. “Pandemic makes Extended Reality a hot ticket.” CCS Insight. <https://my.ccsinsight.com/reportaction/D17106/Toc> (accessed Nov. 11, 2020).
- [15] T. Merel. “Ubiquitous AR to dominate focused VR by 2022.” TechCrunch. <https://techcrunch.com/2018/01/25/ubiquitous-ar-to-dominate-focused-vr-by-2022/> (accessed Nov. 11, 2020).
- [16] “European VR and AR market growth to ‘outpace’ North America by 2023.” Optics.org. <https://optics.org/news/10/10/18> (accessed Nov. 27, 2020).
- [17] ECORYS, “Virtual reality and its potential for Europe”, [Online]. Available: https://ec.europa.eu/futurium/en/system/files/ged/vr_ecosystem_eu_report_0.pdf
- [18] “Seeing is believing, How VR and AR will transform business and the economy.” PwC. <https://www.pwc.com/seeingisbelieving> (accessed Nov. 11, 2020).
- [19] “Augmented and Virtual Reality in Operations: A guide for investment.” Capgemini. <https://www.capgemini.com/research-old/augmented-and-virtual-reality-in-operations/> (accessed Nov. 11, 2020).
- [20] “AR/VR investment and M&A opportunities as startup valuations soften.” Digi-Capital. <https://www.digi-capital.com/news/2019/07/ar-vr-investment-and-ma-opportunities-as-early-stage-valuations-soften/> (accessed Nov. 11, 2020).
- [21] “VR/AR investment at pre-Facebook/Oculus levels in Q1.” Digi-Capital. <https://www.digi-capital.com/news/2020/05/vr-ar-investment-pre-facebook-oculus-levels/> (accessed Nov. 11, 2020).
- [22] L. Segers and D. Del Olmo, “Deliverable D5.1 Map of funding sources for XR technologies”, LucidWeb, XR4ALL project, 2019, [Online]. Available: http://xr4all.eu/wp-content/uploads/d5.1-map-of-funding-sources-for-xr-technologies_final-1.pdf (accessed Nov. 11, 2020).
- [23] “Commercial and public sector investments will drive worldwide AR/VR spending to \$160 billion in 2023, according to a new IDC spending guide.” IDC. <https://www.idc.com/getdoc.jsp?containerId=prUS45123819> (accessed Nov. 11, 2020).
- [24] H. Tankovska. “Unit shipments of Virtual Reality (VR) devices worldwide from 2017 to 2019 (in millions), by vendor.” Statista. <https://www.statista.com/statistics/671403/global-virtual-reality-device-shipments-by-vendor/> (accessed Nov. 11, 2020).
- [25] B.Lang. “Analysis: Monthly-connected VR Headsets on Steam Pass 1 Million Milestone.” Road to VR. <https://www.roadtovr.com/monthly-connected-vr-headsets-steam-1-million-milestone/> (accessed Nov. 11, 2020).
- [26] H. Tankovska. “Smart augmented reality glasses unit shipments worldwide from 2016 to 2022.” Statista. <https://www.statista.com/statistics/610496/smart-ar-glasses-shipments-worldwide/> (accessed Nov. 11, 2020).

- [27] “Augmented and Virtual Reality.” European Commission. <https://ec.europa.eu/growth/tools-databases/dem/monitor/category/augmented-and-virtual-reality> (accessed Nov. 11, 2020).
- [28] The Venture Reality Fund. <https://www.thevrfund.com/landscapes> (accessed Nov. 11, 2020).
- [29] XRA. <https://xra.org/> (accessed Nov. 11, 2020).
- [30] VR/AR Association. <https://www.thevrara.com/> (accessed Nov. 11, 2020).
- [31] VR-Industry forum. <https://www.vr-if.org/> (accessed Nov. 11, 2020).
- [32] Augmented Reality for Enterprise Alliance . <https://thearea.org/> (accessed Nov. 11, 2020).
- [33] IVRPA. <https://ivrpa.org/> (accessed Nov. 11, 2020).
- [34] “The Academy of International Extended Reality”. <https://aixr.org/> (accessed Nov. 20, 2020).
- [35] MedVR. <https://medvr.io/> (accessed Nov. 20, 2020).
- [36] “Open AR Cloud (OARC)” <https://www.openarcloud.org/> (accessed Nov. 28, 2020).
- [37] EuroXR. <https://www.eurovr-association.org/> (accessed Nov. 11, 2020).
- [38] “Extended Reality for Education and Research in Academia”. <https://xrera.eu/> (accessed Nov. 20, 2020).
- [39] “Women in Immersive Tech”. <https://www.wiiteurope.org/> (accessed Nov. 20, 2020).
- [40] EDFVR e.V.. <http://edfvr.org/> (accessed Nov. 11, 2020).
- [41] VRBB. <https://virtualrealitybb.org/> (accessed Nov. 11, 2020).
- [42] Virtual Dimension Center (VDC). <https://www.vdc-fellbach.de/en/> (accessed Nov. 20, 2020).
- [43] GEN Summit. <https://www.gensummit.org/sponsor/varaa/> (accessed Nov. 11, 2020).
- [44] AFXR. <https://www.afxr.org> (accessed Nov. 19, 2020).
- [45] Virtual Reality Finland ry. <https://vrfinland.fi> (accessed Nov. 11, 2020).
- [46] FIVR. <https://fivr.fi/> (accessed Nov. 11, 2020).
- [47] XRNATION. <https://www.xrnation.com/> (accessed Nov. 19, 2020).
- [48] Virtual Switzerland. <http://virtualswitzerland.org/> (accessed Nov. 11, 2020).
- [49] ImmerseUK. <https://www.immerseuk.org/> (accessed Nov. 19, 2020).
- [50] VRINN. <https://vrinn.no/> (accessed Nov. 23, 2020).
- [51] Eurofound, Game-changing technologies: Transforming production and employment in Europe, Luxembourg: Publications Office of the European Union, 2020.
- [52] Espacenet. <https://worldwide.espacenet.com> (accessed Nov. 11, 2020).

4 XR technologies

In this section, all the relevant technologies for extended reality are reviewed. The aim of this section is to describe the current state-of-the-art and to identify the technologies that European companies and institutions play a relevant role in. A list of references in each technology domain points the reader to relevant publications or web sites for further details.

4.1 Video capture for XR

The acquisition of visual footage for the creation of XR application can be organised in three different major technologies: (1) 360-degree video (3-DoF), (2) 3-degrees of freedom (3DoF) + head motion parallax (3-DoF+), and (3) 3D data from real scenes (6-DoF).

For 360-degree video, the inside-out capture of panoramic video acquisition is used. The observer stands at the centre of a scene and looks around, left & right or up & down. Hence, the interaction has just three degrees of freedom (3DoF), namely the three Euler angles.

An intermediate category is labelled as 3-DoF+. It is similar to 360-degree video with 3-DoF, but it additionally supports head motion parallax. Here too, the observer stands at the centre of the scene, but he/she can move his/her head, allowing him/her to look slightly to the sides and behind near objects. The benefit of 3-DoF+ is an advanced and more natural viewing experience, especially in case of stereoscopic 3D video panoramas.

Finally, for the creation of 3D data from real scenes, the outside-in capture approach is used. The observer can freely move through the scene while looking around. The interaction allows six degrees of freedom (6DoF), the three directions of translation plus the 3 Euler angles. In this category, several sensors fall in, such as (1) multi-view cameras including light-field cameras, depth, and range sensors, RGB-D cameras, and (2) complex multi-view volumetric capture systems. A good overview on VR technology and related capture approaches is presented in [53][54][55].

4.1.1 360-degree video (3-DoF)

Panoramic 360-degree video is certainly one of the most exciting viewing experiences when watched through VR glasses. However, today's technology still suffers from some technical restrictions.

One restriction can be explained very well by referring to the capabilities of the human vision system. It has a spatial resolution of about 60 pixels per degree. Hence, a panoramic capture system requires a resolution of more than 20,000 pixel (20K) at the full 360-degree horizon and meridian, the vertical direction. Current state-of-art commercial panoramic video cameras are far below this limit, ranging from 2,880 pixel horizontal resolution (Kodak SP360 4K Dual Pro, 360 Fly 4K) via 4,096 pixel (Insta360 4K) up to 11k pixel (Insta360 Titan). In [56], a recent overview on the top ten 360-degree video cameras is presented, which all offer monoscopic panoramic video.

Fraunhofer HHI has already developed an omni-directional 360-degree video camera with 10K resolution in 2016. This camera uses a mirror system together with 10 single HD camera along the horizon and one 4K camera for the zenith. Upgrading it completely to 4K cameras would even support the required 20K resolution at the horizon. The capture system of this camera also includes real-time stitching and online preview of the panoramic video in full resolution [57].

However, the maximum capture resolution is just one aspect. A major bottleneck concerning 360-degree video quality is the restricted display resolution of the existing VR headsets. Supposing that the required field of view is 120 degrees in the horizontal direction and 60 degrees in the vertical direction, VR headsets need two displays, one for each eye, each with a resolution of 8K by 4K. As discussed in section 4.5, this is far away from what VR headsets can achieve today.

4.1.2 Head motion parallax (3-DoF+)

A further drawback of 360-degree video is the missing capability of head motion parallax. In fact, 360-degree video with 3DoF is only sufficient for monocular video panoramas, or for stereoscopic-3D panoramic views with far objects only. In case of stereo 3D with near objects, the viewing condition is confusing, because it is different from what humans are accustomed to from real-world viewing.

Nowadays, a lot of VR headsets support local on-board head tracking (see section 4.6.2). This allows for enabling head motion parallax while viewing a 360-degree panoramic video in VR headsets. To support this option, capturing often combines photorealistic 3D scene compositions (see section 4.1.3) with segmented stereoscopic videos. For example, one or more stereoscopic videos are recorded and keyed in a green screen studio. In parallel, the photorealistic scene is generated by 3D modelling methods like photogrammetry (see sections 4.1.3 and 4.3.2). Then, the separated stereoscopic video samples are placed at different locations into the above-mentioned photorealistic 3D scene, probably in combination with additional 3D graphic objects. The whole composition is displayed as a 360-degree stereo panorama in a tracked VR headset via usual render engines. The user can slightly look behind the inserted video objects while moving the head and, hence, gets the natural impression of head motion parallax.

Such a 3-DoF+ experience was shown for the first time by Intel in cooperation with Hype VR in January 2017 at CES as a so-called walk-around VR video experience. This experience featured a stereoscopic outdoor panorama from Vietnam with a moving water buffalo and some static objects presented in stereo near to the viewer [58]. The user could look behind the near objects while moving the head. Similar and more sophisticated experiences have later been shown, e.g., by Sony, Lytro, and others. Likely the most popular one is the Experience “Tom Grennan VR” that was presented for the first time in July 2018 by Sony on PlayStation VR. Tom Grennan and his band have been recorded in stereo in a green screen studio and have then been placed in a photorealistic 3D reconstruction of a real music studio that has been scanned by Lidar technology beforehand.

4.1.3 3D capture of static objects and scenes (6-DoF)

The 3D capture of objects and scenes has reached a mature state to allow professionals and amateurs to create and manipulate large amount of 3D data such as point clouds and meshes. The capture technology can be classified into active and passive ones. On the active sensor side, laser or LIDAR (light detection and ranging), time-of-flight, and structured-light techniques can be mentioned. Photogrammetry is the passive 3D capture approach that relies on multiple images of an object or a scene captured with a camera from different viewpoints. Especially the increase in quality and resolution of cameras developed the use of photogrammetry. A recent overview can be found in [59]. The maturity of the technology led to a number of commercial 3D body scanners available on the market, ranging from 3D scanning booth and 3D scan cabins to body scanning rigs, body scanners with a rotating platform, and even home body scanners embedded in a mirror, all for single-person use [60].

4.1.4 3D capture of volumetric video (6DoF)

The techniques from section 4.1.3 are limited to static scenes and objects. For dynamic scenes, static objects can be animated by scripts or motion capture system, and a virtual camera can be navigated through the static 3D scene. However, the modelling and animation process of moving characters is time consuming and often it cannot really represent all moving details of a real human, especially facial expressions and the motion of clothes.

In contrast to these conventional methods, volumetric video is a new technique that scans humans, in particular actors, with plenty of cameras from different directions, often in combination with active depth sensors. During a complex post-production process that we describe in Section 4.3.8, this large amount of initial data is then merged to a dynamic 3D mesh representing a full free-viewpoint video. It has the naturalism of high-quality video, but it is a 3D object where a user can walk around in the virtual 3D scene.

In recent years, a number of volumetric studios have been created that are able to produce high-quality volumetric videos. Usually the subject of the volumetric video is the entire human body, but some volumetric studios provide specific solutions designed to handle explicitly the human face [61]. The volumetric video can be viewed in real-time from a continuous range of viewpoints chosen at any time during playback. Most studios focus on a capture volume that is viewed spherically in 360 degrees from the outside. A large number of cameras are placed around the scene (e.g. in studios from 8i [62], Volucap [63], 4DViews [64], Evercoast [65], HOLOOH [66], and Volograms [67]) providing input for volumetric video similar to frame-by-frame photogrammetric reconstruction of the actors, while Microsoft's Mixed reality Capture Studios [68] additionally rely on active depth sensors for geometry acquisition. In order to separate the scene from the background, all studios are equipped with green screens for chroma keying. Only Volucap [63] uses a bright backlit background to avoid green spilling effects in the texture and to provide diffuse illumination. This concept is based on a prototype system developed by Fraunhofer HHI [69].

4.1.5 References

- [53] C. Anthes, R. J. García-Hernández, M. Wiedemann and D. Kranzlmüller, "State of the art of virtual reality technology," 2016 IEEE Aerospace Conference, Big Sky, MT, 2016, pp. 1-19. doi: 10.1109/AERO.2016.7500674.
- [54] State of VR. <http://stateofvr.com/> (accessed Nov. 11, 2020).
- [55] "3DOF, 6DOF, RoomScale VR, 360 Video and Everything In Between." Packet39. <https://packet39.com/blog/2018/02/25/3dof-6dof-roomscale-vr-360-video-and-everything-in-between/> (accessed Nov. 11, 2020).
- [56] L. Brown. "Top 10 professional 360 degree cameras." Wondershare. <https://filmora.wondershare.com/virtual-reality/top-10-professional-360-degree-cameras.html> (accessed Nov. 11, 2020).
- [57] "OmniCam-360". Fraunhofer HHI. <https://www.hhi.fraunhofer.de/en/departments/vit/technologies-and-solutions/capture/panoramic-uhd-video/omnicam-360.html> (accessed Nov. 11, 2020).
- [58] "Intel demos world's first 'walk-around' VR video experience". Intel, <https://www.youtube.com/watch?v=DFobWjSYst4> (accessed Nov. 11, 2020).

- [59] F. Fadli, H. Barki, P. Boguslawski, L. Mahdjoubi, “3D Scene Capture: A Comprehensive Review of Techniques and Tools for Efficient Life Cycle Analysis (LCA) and Emergency Preparedness (EP) Applications,” *presented at International Conference on Building Information Modelling (BIM) in Design, Construction and Operations*, Bristol, UK, 2015, doi: 10.2495/BIM150081.
- [60] “The 8 best 3D body scanners in 2020.” Aniwaa. <https://www.aniwaa.com/best-3d-body-scanners/> (accessed Nov. 11, 2020).
- [61] OTOY. <https://home.otoy.com/capture/lightstage/> (accessed Nov. 11, 2020).
- [62] 8i. <http://8i.com> (accessed Nov. 11, 2020).
- [63] Volucap. <http://www.volucap.de> (accessed Nov. 11, 2020).
- [64] 4DViews. <http://www.4dviews.com> (accessed Nov. 11, 2020).
- [65] Evercoast. <https://evercoast.com/> (accessed Nov. 11, 2020).
- [66] HOLOOH. <https://www.holooh.com/> (accessed Nov. 11, 2020).
- [67] Volograms. <https://volograms.com/> (accessed Nov. 11, 2020).
- [68] Microsoft. <http://www.microsoft.com/en-us/mixed-reality/capture-studios> (accessed Nov. 11, 2020).
- [69] O. Schreer, I. Feldmann, S. Renault, M. Zepp, P. Eisert, P. Kauff, “Capture and 3D Video Processing of Volumetric Video”, *2019 IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan, Sept. 2019.

4.2 3D sound capture

There are several approaches for capturing spatial 3D sound for an immersive XR experience. Most of them are extensions of existing recording technologies, while some are specifically developed to capture a three-dimensional acoustic representation of their surroundings.

4.2.1 Human sound perception

To classify 3D sound capture techniques, it is important to understand how human sound perception works. The brain uses different stimuli when locating the direction of a sound. The most well-known is probably the interaural level difference (ILD) of a soundwave entering the left and right ears. Because low frequencies are bend around the head, the human brain can only locate a sound source through ILD, if this sound contains frequencies higher than 15,00Hz [70]. To locate sound sources containing lower frequencies, the brain uses the interaural time difference (ITD). The time difference between sound waves arriving at the left and right ears is used to determine the direction of a sound [70]. Due to the symmetric positioning of the human ears in the same horizontal plane, these differences only allow one to locate the sound in the horizontal plane but not in the vertical direction. Also with these stimuli, the human sound perception cannot distinguish between soundwaves that come from the front or from the back. For an exact further analysis of the sound direction, the Head-Related Transfer Function (HRTF) is used. This function describes the filtering effect of the human body, especially of the head and the outer ear. Incoming sound waves are reflected and absorbed at the head surface in a way that depends from their directions, therefore the filtering effect changes as a function of the direction of the sound source. The

brain learns and uses these resonance and attenuation patterns to localise sound sources in three-dimensional space. Again see [70] for a more detailed description.

4.2.2 3D microphones

Using the ILD and ITD stimuli as well as specific microphone arrangements, classical stereo microphone setups can be extended and combined to capture 360-degree sound (only in horizontal plane) or truly 3D sound. Complete microphone systems are Schoeps IRT-Cross, Schoeps ORTF Surround, Schoeps ORTF-3D, Neveon BPT, Josephson C700S, Edge Quadro. Furthermore, any custom microphone setup can be used in combination with a spatial encoder software tool. As an example, Fraunhofer upHear is a software library to encode the audio output from any microphone setup into a spatial audio format [71]. Another example is the Schoeps Double MS Plugin, which can encode specific microphone setups.

4.2.3 Binaural microphones

An easy way to capture a spatial aural representation is to use the previously mentioned HRTF (see section 4.2.1). Two microphones are placed inside the ears of a replica of the human head to simulate the HRTF. The time response and the related frequency response of the received stereo signal contain the specific HRTF information and the brain can decode it when the stereo signal is listened to over headphones. Typical systems are Neumann KU100, Davinci Head Mk2, Sennheiser MKE2002, and Kemar Head and Torso. Because every human has a very individual HRTF, this technique only works when the HRTF recorded by the binaural microphone is similar to the HRTF of the person listening to the recording. Moreover, most problematic in the context of XR applications is the fact that the recording is static, which means that the position of the listener cannot be changed afterwards. This makes binaural microphones incompatible with most XR cases. To solve this problem, binaural recordings in different directions are recorded and mixed afterwards depending on the user position in the XR environment. As this technique is complex and costly, it is not used so frequently anymore. Examples of such systems are the 3Dio Omni Binaural Microphone and the Hear360 8Ball. Even though HRTF-based recording techniques for XR are mostly outdated, the HRTF-based approach is very important in audio rendering for headsets (see section 4.4.3).

4.2.4 Ambisonic microphones

Ambisonics describes a sound field by spherical harmonic modes. Unlike the previously mentioned capture techniques, the recorded channels cannot be connected directly to a specific loudspeaker setup, like stereo or surround sound. Instead, it describes the complete sound field in terms of one monopole and several dipoles. In higher-order Ambisonics (HOA), quadrupoles and more complex polar patterns are also derived from the spherical harmonic decomposition.

In general, Ambisonics signals need a decoder in order to produce a playback-compatible loudspeaker signal in dependence of the direction and distance of the speakers. A HOA-decoder with an appropriate multichannel speaker setup can give an accurate spatial representation of the sound field. Currently, there are many First Order Ambisonics (FOA) microphones like the Soundfield SPS200, Soundfield ST450, Core Sound TetraMic, Sennheiser Ambeo, Brahma Ambisonic, Røde NT-SF1, Audeze Planar Magnetic Microphone, and Oktava MK-4012. All FOA microphones use a tetrahedral arrangement of cardioid directivity microphones and record four channels (A-Format), which is encoded into the Ambisonics-Format (B-Format) afterwards. For more technical details on Ambisonics, see [72].

4.2.5 Higher-Order Ambisonics (HOA) microphones and beamforming

Recently, HOA-microphones, which can be used to produce Ambisonics signals of second order (Brahma-8, Core Sound OctoMic), third order (Zylia ZM-1), and even fourth order (mhacoustics em32), have been launched. They allow for a much higher spatial resolution than their FOA counterparts. In order to construct the complex spatial harmonics of HOA, beamforming is used to create a virtual representation of the sound field, which can then be encoded into the HOA format [73]. For spherical (3D) or linear (2D) microphone arrays, beamforming can also be used to derive loudspeaker feeds directly from the microphone signals, e.g. by the application of Plane Wave Decomposition. Furthermore, in the European Framework 7 project FascinatE, multiple spherical microphone arrays were used to derive positional object-oriented audio data [74].

4.2.6 Limitations and applications

All the previously mentioned techniques record a stationary sound field. This creates 3 degrees of freedom (3DoF) in XR applications. For 6 degrees of freedom (6DoF), an object-oriented method capturing every sound source individually is usually required (see section 4.4.4). In practice, it is a common use to mix the above-described techniques in an appropriate manner. A 360-degree microphone or an Ambisonics microphone can be used to capture the spatial ambience of the scene, whereas classical microphones with specific spatial directivity are used to capture particular elements of the scene for the post production. Recently, Zylia released the 6DoF VR/AR Development Kit, which uses nine Zylia ZM-1 microphones at a time. In combination with a proprietary playback-system, it allows for spatial audio scenes with 6DoF representations [75].

4.2.7 References

- [70] Schnupp, J., Nelken, I., and King, A., *Auditory neuroscience: Making sense of sound*, MIT Press, 2011
- [71] Fraunhofer IIS. <https://www.iis.fraunhofer.de/en/ff/amm/consumer-electronics/uphear-microphone.html> (accessed Nov. 11, 2020).
- [72] Furness, R. K., “Ambisonics-an overview”, In *Audio Engineering Society Conference: 8th International Conference: The Sound of Audio*, 1990.
- [73] MH Acoustics , “Eigenbeam Dta Specification for Eigenbeams”, 2016, [Online]. Available: https://mhacoustics.com/sites/default/files/Eigenbeam%20Datasheet_R01A.pdf (accessed Nov. 11, 2020).
- [74] “FP7 Project Fascinate - Format-Agnostic SScript-based INterAcTive Experience”. <https://cordis.europa.eu/project/id/248138> (accessed Nov. 11, 2020).
- [75] ZYLIA. <https://www.zylia.co/zylia-6dof.html> (accessed Nov. 11, 2020).

4.3 Scene analysis and computer vision

4.3.1 Multi-camera geometry

3D scene reconstruction from images can be achieved by (1) multiple images from a single camera at different viewpoints or (2) multiple cameras at different viewpoints. The first approach is called structure from motion (SfM), while the second is called multi-view reconstruction. However, the knowledge of the camera position and orientation is required in both approaches, before successfully applying 3D scene analysis and reconstruction.

For SfM, this process is named self-calibration. Plenty of approaches have been proposed in the past and there are several commercial tools available that perform self-calibration and 3D scene reconstruction based on multiple images of a single camera such as Autodesk ReCap, Agisoft Metashape, AliceVision Meshroom, Pix4D, PhotoModeler, RealityCapture, Regard3D and many more.

If a fixed multi-camera setup is used for the capture of dynamic scenes, then standard camera calibration techniques are applied to achieve the required information for scene reconstruction. Here, calibration patterns or object with known 3D geometry are used to calibrate the cameras.

For both domains, SfM and multi-view geometry, the research task can be considered as completed. However, some very special use cases and problems are still under investigation.

4.3.2 3D Reconstruction

Sparse or semi-sparse (but not dense) 3D reconstruction of static scenes from multi-view images can already be considered as reliable and accurate. For instance, photogrammetry aims to produce multiple still images of a rigid scene or object and to deduce its 3D structure from this set of images [76]. In contrast, SLAM (Simultaneous Localisation and Mapping) takes a sequence of images from a single moving camera and reconstructs the 3D structure of a static scene progressively while capturing the sequence [77]. However, single-view and multi-view dense 3D reconstructions with high accuracy remain more challenging. Best performance has been achieved by deep-learning neural networks [78][79], but they still suffer from limited accuracy and overfitting. Recently, thanks to more and better 3D training data, 3D deep-learning methods have made a lot of progress [80], significantly outperforming previous model-based approaches [81][82].

4.3.3 Visual volumetric media compression for 6DoF

So far, we have assumed that 3D content is best represented by 3D meshes overlayed with 2D textures that are streamed over the internet to enable tele-immersive 3D media applications. Mesh codecs used for this purpose have been developed over the past 20 years, with an excellent overview given in [83]. There are, however, challenges not only in coding the position of the mesh vertices, but also their connectivity for creating triangles that span a 2D surface in space. Typically, mesh coding starts with a seed-triangle that is extended such that each new vertex coded in the bit stream yields a new triangle, eventually creating a one-dimensional triangle strip that is rolled up at the decoder like a potato peel to reconstruct the 3D object. For example, the Draco codec [84] based on EdgeBreaker [85][86] performs very well amongst state-of-the-art mesh codecs [83][84], finding the best cut and triangle strip for achieving a 15-30 bits per vertex coding cost (geometry and attributes, including colour and normals) with little quality degradation.

Unfortunately, comparing 3D mesh coding of vertices and triangles with 2D video coding for image pixels, the compression performance of the latter is far more superior with its 0.04 bits per pixel in HEVC (High Efficiency Video Coding) or 0.02 bits per pixel in the latest VVC (Versatile Video Coding) video codec [87], developed over the 30 years long history of the MPEG video coding standardisation activity. The main reasons for the mesh coding inferior performance are that (1) the vertex mesh connectivity is expensive to code, yielding a cost of already 1.5 to 6 bits per vertex without any attributes [83][86], and (2) it is difficult to exploit temporal redundancies between the successive positions of moving vertices in an animated 3D object, especially with time-varying levels of detail.

Therefore, the MPEG-I immersive media standardisation committee (where “I” refers to “Immersive”) started developing an alternative approach in 2018 with its Final Draft International Standard (FDIS) stage reached mid-2020, where instead of directly coding the 3D object, its 2D orthographic projections and associated depth maps are coded with conventional video codecs [88]. The object is surrounded by a cube with each of its faces representing the textures and depths (per-pixel distance from the face to the object) that allow the reconstruction of each point of the point cloud (no explicit connectivity is present), typically rendered with splatting [89]. It is therefore referred to as V-PCC, which stands for Video Point Cloud Coding. This concept was originally proposed in a Depth Image-Based Rendering (DIBR) scheme [90], later extended to video, well capturing all temporal redundancies for better coding performance. In practice, each object is coded independently, while a scene graph description - e.g. glTF [91] from the Khronos group (in exploration phase for extension to streaming capabilities in MPEG-I) – repositions all objects in the scene. The combination of V-PCC and glTF allows both free navigation (6DoF), as well as free object displacement (scene editing).

The above DIBR technique is remarkably similar to the MPEG Immersive Video (MIV) approach independently developed in a second subgroup of MPEG to address free navigation in real scenery without prior 3D reconstruction [92]. The scene is captured from a dozen of directions with conventional cameras, out of which a depth map per view is estimated to implicitly represent the scene geometry. The rendering of any virtual viewpoint can then be performed by image warping of existing camera views. To overcome any cracks (spurious missing pixels) in the rendered images, implicit triangles between each triplet of adjacent pixels in each warped view are fed to a conventional OpenGL pipeline. Of course, splatting as in V-PCC can also be used. In the end, MIV allows to freely navigate in the scene along a 6DoF scenario like in V-PCC, even detecting collisions if needed, but – in contrast to V-PCC and glTF – MIV does not allow to freely displace the objects in the scene. Nevertheless, bitstream format alignment with V-PCC has identified that V-PCC and MIV have 95% in common, and therefore MPEG has issued a single Visual Volumetric Video-based Coding (V3C) standard common to both. The 5% remaining differences are tackled in two annexes of the standard, one for V-PCC and another for MIV, reaching FDIS status early 2021. A first version of the OpenV3C software library for V3C coding has also been released [93].

With all these volumetric coding methods, one may ask which one to use: point clouds, meshes or immersive video? This will of course depend on the specific use case, where streaming considerations also play an important role. For instance, while [94] has developed streaming techniques for meshes requiring around 10 Mbps per object, the preliminary study in [95] suggests that point cloud streaming with V-PCC (tested under various configurations) and MPEG-DASH [96][97] achieves a better perceptual quality vs.

bitrate ratio in the 10-50 Mbps bitrate range than the Draco mesh coding presented before, when streaming with multi-object data prioritisation schemes.

A last consideration is the price to pay for more interactivity (6DoF free navigation, free object displacement) compared to conventional 2D video streaming: while UHD-TV requires a bandwidth of 10 Mbps to stream a TV channel, bitrates of 50 to 100 Mbps or even more (to stream the complete scene) – i.e. equivalent to a dozen of UHD-TV channels – are not uncommon in the V3C framework. Consequently, more research will be needed to evaluate quality and streaming scenarios of visual volumetric media in realistic working conditions before penetration into the market.

4.3.4 3D Motion analysis

The 3D reconstruction of dynamic and deformable objects is much more complicated than for static objects. Such reconstruction is mainly used for bodies and faces using model-based approaches. There has been significant progress for human dynamic geometry and kinematics capture, especially for faces, hands, and torso [98][99][100][101][102][103][104]. The best performing methods use body-markers. In a realistic markerless setting, a common approach is to fit a statistical model to the depth channel of an RGBD-sensor. However, even for these well-researched objects, a holistic approach to capture accurate and precise motion and deformations from casually-captured RGB images in an unconstrained setting is still challenging [105][106][107][108]. General-case techniques for deformation and scene capture are far less developed [109]. Deep learning has only recently been used for complex motion and deformation estimation as the problem is very complex and the availability of labelled data is limited. Generative Adversarial Networks (GAN) have been used recently to estimate the content of future frames in a video, but today's generative approaches lack physics- and geometry-awareness and results in a lack realism [110][111]. First approaches have addressed general non-rigid deformation modelling by incorporating geometric constraints into deep learning.

4.3.5 Human body modelling

When the animation of virtual humans is required, as it is the case for applications like computer games, virtual reality, and film, computer graphics models are usually used. They allow for arbitrary animation, with body motion generally being controlled by an underlying skeleton while facial expressions are described by a set of blend shapes [109]. The advantage of full control comes at the price of significant modelling effort and sometimes limited realism. Usually, the body model is adapted in shape and pose to the desired 3D performance. Given a template model, the shape and pose can be learned from the sequence of real 3D measurements, in order to align the model with the sequence [112]. Recent progress in deep learning also enables the reconstruction of highly accurate human body models even from single RGB images [102]. Similarly, Pavlakos et al. [103] estimate the shape and pose of a template model from a monocular video sequence such that the human model exactly follows the performance in the sequence. Haberman et al [104] go one step further and enable real-time capture of humans including surface deformations due to clothes.

4.3.6 Appearance analysis

Appearance encompasses characteristics such as surface orientation, albedo, reflectance, and illumination. The estimation of properties usually requires prior knowledge of Lambertian materials, point lights, and 3D-shape. While significant progress has been made on inferring materials and illumination

from images in constrained settings, progress in an unconstrained setting is very limited. Even for the constrained cases, estimating Bidirectional Reflectance Distribution Functions (BRDFs) is still out of reach. Classic appearance estimation methods, where an image is decomposed into pixel-wise products of albedo and shading, rely on prior statistics (e.g. from semi-physical models) [105] or user intervention [109]. Going beyond such simple decompositions, the emergence of Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN) offer new possibilities in appearance estimation and modelling. These two types of networks have successfully been used for image decomposition together with sparse annotation [113], to analyse the relationships between 3D-shape, reflectance and natural illumination [114], and to estimate the reflectance maps of specular materials in natural lighting conditions [115]. For specific objects, like human faces, image statistics from sets of examples can be exploited for generic appearance modelling [116], and recent approaches have achieved realistic results using deep neural networks to model human faces in still images [117][118]. GANs have been used to directly synthesise realistic images or videos from input vectors from other domains without explicitly specifying scene geometry, materials, lighting, and dynamics [119][120][121]. Very recently, deep generative networks that take multiple images of a scene from different viewpoints and construct an internal representation to estimate the appearance of that scene from unobserved viewpoints [122][123] have been introduced. However, current generative approaches lack a fundamental, global-understanding of synthesised scenes, with visual quality and diversity of scenes generated being limited. These approaches are thus far behind in terms of providing the high-resolution, high dynamic range, and high frame rate that videos require for realism.

4.3.7 Realistic character animation and rendering

Recently, more and more hybrid and example-based animation synthesis methods have been proposed that exploit captured data in order to obtain realistic appearances. One of the first example-based methods has been presented by [124] and [125], who synthesise novel video sequences of facial animations and other dynamic scenes by video resampling. Malleson et al. [126] present a method to continuously and seamlessly blend multiple facial performances of an actor by exploiting complementary properties of audio and visual cues to automatically determine robust correspondences between takes, allowing a director to generate novel performances after filming. These methods yield 2D photorealistic synthetic video sequences, but are limited to replaying captured data. This restriction is overcome by Fyffe et al. [127] and Serra et al. [128], who use a motion graph in order to interpolate between different 3D facial expressions captured and stored in a database.

For full body poses, Xu et al. [129] introduced a flexible approach to synthesise new sequences for captured data by matching the pose of a query motion to a dataset of captured poses and warping the retrieved images to query pose and viewpoint. Combining image-based rendering and kinematic animation, photo-realistic animation of clothing has been demonstrated from a set of 2D images augmented with 3D shape information in [130]. Similarly, Paier et al. [131] combine blend-shape-based animation with recomposing video-textures for the generation of facial animations.

Character animation by resampling of 4D volumetric video has been investigated by [132][133], yielding high visual quality. However, these methods are limited to replaying segments of the captured motions. In [134] Stoll et al. combine skeleton-based CG models with captured surface data to represent details of apparels on top of the body. Caras et al. [135] combined concatenation of captured 3D sequences with

view dependent texturing for real-time interactive animation. Similarly, Volino et al. [136] presented a parametric motion graph-based character animation for web applications. Only recently, Boukhayma and Boyer [137][138] proposed an animation synthesis structure for the re-composition of textured 4D video capture, accounting for geometry and appearance.

They propose a graph structure that enables interpolation and traversal between pre-captured 4D video sequences. Finally, Regateiro et al. [139] present a skeleton-driven surface registration approach to generate temporally consistent meshes from volumetric video of human subjects in order to facilitate intuitive editing and animation of volumetric video.

Purely data-driven methods have recently gained significant importance due to the progress in deep learning and the possibility to synthesise images and video. Chan et al.[140], for example, use 2D skeleton data to transfer body motion from one person to another and synthesise new videos with a Generative Adversarial Network. The skeleton motion data can also be estimated from video by neural networks [141]. Liu et al. [142] extend that approach and use a full template model as an intermediate representation that is enhanced by the GAN. Similar techniques can also be used for synthesising facial video as shown, e.g., in [143].

4.3.8 Pose estimation

Any XR application requires the collocation of real and virtual space so that when the user moves his head (in case of a headset device) or his hand (in the case of handheld device), the viewpoint on digital content is consistent with the user's viewpoint in the real environment. Thus, if the virtual camera used to render the digital content has the same intrinsic parameters and is positioned at the same location as the physical XR device, the digital content will be perceived as fixed in relation to the scene when the user moves. As a result, any XR system needs to estimate the pose (position and orientation) of the XR device to offer a coherent immersive experience to the user. Moreover, when a single object of the real environment is moving, its pose has to be estimated if the XR application requires to attach to it a digital content. Two categories of pose estimation system exist, the outside-in systems and the inside-out systems.

Outside-in systems (also called exteroceptive systems) requires external hardware not integrated to the XR device to estimate its pose. Professional optical solutions provided by ART™, Vicon™, OptiTrack™ use a system of infrared cameras to track a constellation of reflective or active markers to estimate the pose of these constellation using a triangulation approach. Other solutions use electromagnetic field to estimate the position of a sensor in the space, but they have limited range. More recently, HTC™ has developed a scanning laser system used with their Vive headset and tracker to estimate their pose. The Vive™ lighthouse sweeps horizontally and vertically the real space with a laser at a very high frequency. This laser activates a constellation of photo-sensitive receivers integrated into the Vive headset or tracker. By knowing when each receiver is activated, the Vive system can estimate the pose of the headset or tracker. All these outside-in systems require to equip the real environment with dedicated hardware, and the area where the pose of the XR device can be estimated is restricted by the range of the emitters or receivers that track the XR device.

To overcome the limitation of outside-in systems, most of current XR systems are now using inside-out systems to estimate the pose of the XR device. An inside-out system (also called interoceptive system) uses only built-in sensors to estimate the pose of the XR device. Most of these systems are inspired by the

human localisation system, and are mainly using a combination of vision sensors (RGB or depth camera) and inertial sensors (Inertial Measurement Unit). It consists of three main steps, the relocalisation, the tracking and the mapping. The relocalisation is used, when the XR device has no idea about its pose (initialisation or when tracking failed). It uses the data captured by the sensor at a specific time as well as a knowledge of the real environment (a 2D marker, a CAD model or a cloud of points) to estimate the first pose of the device without any prior knowledge of its pose at the previous frame. This task is still challenging as the knowledge about the real environment previously captured does not always correspond to what it observes at runtime with vision sensors (objects have moved, lighting conditions have changed, elements are occluding the scene, etc.). Then, the tracking estimates the occurring movement of the camera between to frames when the relocalisation task has been achieved. This task is less challenging as the real world observed by the XR device does not really change in a very short time. Finally, the XR device can create a 3D map of the real environment by triangulating points which match between two frames knowing the pose of the camera capturing them. This map can then be used to represent a knowledge of the real environment used by the relocalisation task. The loop that tracks the XR device and that maps the real environment is called SLAM (Simultaneous localisation And Mapping) [144][145]. Most existing inside-out pose estimation solution (e.g. ARKit from Apple, ARCore from Google, or Hololens and Mixed Reality SDKs from Microsoft), are based on a derivation implementation of a SLAM. Only for XR near-eye display, the motion-to-photon latency, i.e. the time elapsed between the movement of the user's head and the visual feedback of his movement, should be less than 20ms. If this latency is higher, it results in motion sickness for video see-through displays, and in floating objects for optical see-through displays. To achieve this low motion-to-photon latency, the XR systems interpolate the camera poses using inertial sensors, and reduce the computation time thanks to hardware optimisation based on vision processing unit. Recent implementation of SLAM pipelines are more and more using low-level components based on machine learning approaches [146][147][148][149]. Finally, future 5G network offering low latency and high bandwidth will allow to distribute into edge cloud and into centralised cloud efficient pipelines to improve AR devices localisation accuracy, even on low-resources AR devices, and address large scale AR applications.

4.3.9 Volumetric Video

In Section 4.1.4, different 3D capture of volumetric video studios were described. These studios enable the creation of high-quality 3D video content for free-viewpoint rendering on VR and AR devices. Next, we assume that the studio has a multiple view camera setup.

Firstly, the initial captured data from multiple cameras are used to generate an initial 3D point cloud as described for example in [150]. Usually a stereo depth estimation is performed per camera pair, then this partial point cloud estimation is fused together contributing to a rough initial 3D point cloud estimation.

Secondly, the 3D point cloud is converted to dynamic meshes. Surface reconstruction from the 3D point cloud is performed by using for example a standard technique called screened Poisson Surface Reconstruction [151]. Surface reconstruction from an oriented point cloud is quite a challenging problem due to surface complexity, if one considers the flexibility of human body parts and variation of facial expressions, but also due to the noisy, estimated point cloud. This step is very important as it provides with a surface that is quite more realistic and pleasant to look at than looking at a dynamic point cloud in a volumetric video framework. Noise reducing techniques are often applied on the noisy point cloud

before giving it to the surface reconstruction technique. The reconstruction part provides a 3D scene that consists of dynamic meshes. In order to reduce the complexity of the 3D scene, the dynamic meshes are usually simplified to a level that is a good trade-off between the scene quality and the scene complexity. The simplification process could for example iteratively contract edges based on quadric error metric until the desired simplification level is reached [152]. Note that the desired simplification level is defined in accordance to the application and computational capabilities of the VR and AR device leading to a good trade-off between the scene quality and the scene complexity.

Thirdly, as the 3D scene realism is improved when the meshes are shown textured, a related video is organised as texture atlas for later texture mapping. Both the simplification and the texture mapping part benefit when taking into account sensitive regions and handling them differently [153].

Furthermore, to improve the temporal consistency of the produced textured atlas and of the mesh topology, the simplified meshes could be registered using a keyframe-based technique as presented in [154]. These dynamic meshes can be inserted as volumetric video representation into 3D virtual scenes modelled with approaches from 4.1.3 with the result that the user can freely navigate around the volumetric video in the virtual scene.

As a further step, the volumetric video content could even be enriched by adding new, unseen performances based on the captured video content [155].

4.3.10 References

- [76] P.E. Debevec, C.J. Taylor, J. Malik, "Modeling and rendering architecture from photographs", *Proc. of the 23rd Annual Conference on Computer Graphics and Interactive Techniques – SIGGRAPH '96*, ACM Press, New York, USA, 1996, pp. 11-20.
- [77] R. Mur-Artal, J. Motiel, J. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system", *IEEE Trans. Robotics*, vol. 31, no. 5, pp. 1147-1163, 2015.
- [78] M. Poggi et al., "Learning monocular depth estimation with unsupervised trinocular assumptions", in *proc. 6th International Conference on 3D Vision (3DV)*, Verona, Italy, 2018, pp. 324-333.
- [79] H. Zhou, B. Ummenhofer, T. Brox, "DeepTAM: Deep tracking and mapping", *European Conference on Computer Vision (ECCV)*, 2018.
- [80] A. Chang, T. Funkhouser, L. Guibas, Q. Hung, Z. Li, S. Savarese, M. Savva, S. Song, J. Xiao, L. Yi, F. Yu, "ShapeNet: An information-rich 3D model repository", arXiv:1512.03012, 2015.
- [81] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, A. Geiger, "Occupancy Networks: Learning 3D reconstruction in function space", arXiv:1812.03828, 2018.
- [82] J. Park, P. Florence, J. Straub, R. Newcombe, S. Lovegrove, "DeepSDF: Learning continuous SDFs for shape representation", arXiv:1901.05103, 2019
- [83] A. Malgo, G. Lavoue, F. Dupont, C. Hudelot, "3D mesh compression: survey, comparisons and emerging trends", *ACM Computing Surveys*, Vol. 9, No. 4, Article 39, pp. 39:1-39:40, Sept. 2013.
- [84] A. Doumanoglou, P. Drakoulis, N. Zioulis, D. Zarpalas, P. Daras, "Benchmarking Open-Source Static 3D Mesh Codecs for Immersive Media Interactive Live Streaming", *Journal on Emerging and Selected Topics in Circuits and Systems*, Feb. 2019, doi: 10.1109/JETCAS.2019.2898768.

- [85] J. Rossignac, A. Safonova, A. Szymczak, “3D Compression Made Simple: Edgebreaker with ZipandWrap on a corner-table”, *SMI 2001 International Conference on Shape Modeling and Applications*, 2001.
- [86] T. Lewiner, H. Lopes, J. Rossignac, A. W. Vieira, “Efficient Edgebreaker for surfaces of arbitrary topology”, *Proceedings. 17th Brazilian Symposium on Computer Graphics and Image Processing*, Curitiba, Brazil, pp. 218-225, 2004, doi: 10.1109/SIBGRA.2004.1352964..
- [87] O. Stankiewicz, G. Lafruit, M. Domanski, “Multiview Video: Acquisition, Processing, Compression and Virtual View Rendering”, in *Academic Press Library in Signal Processing: Image and Video Processing and Analysis and Computer Vision*, Chellappa R., Theodoridis S., Ed., vol. 6, pp. 3-74, 2017.
- [88] S. Schwarz et al., “Emerging MPEG Standards for Point Cloud Compression”, *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133-148, Mar. 2019.
- [89] M. Gross, H. Pfister, *Point-based Graphics*, in The Morgan Kaufmann series in Computer Graphics, Morgan Kaufmann publishers, 2007.
- [90] L. Levkovich-Maslyuk, A. Ignatenko, A. Zhirkov, A. Konushin, I. K. Park, M. Han, Y. Bayakovski “Depth Image-Based Representation and Compression for Static and Animated 3-D Objects”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 7, pp. 1032-1045, July 2004.
- [91] “glTF Overview.” Khronos Group. <https://www.khronos.org/glTF/> (accessed Nov. 11, 2020).
- [92] G. Lafruit, A. Schenkel, C. Tulvan, M. Preda, Y. Lu, “MPEG-I Coding performance in Immersive VR/AR applications”, *IBC 2018, International Broadcasting Convention: IET: Best of IBC 2018, The Institution of Engineering and Technology*, pp. 23-27, 13 Sept. 2018.
- [93] MPEG-I 3DG, “OpenV3C – Multi-platform open-source implementation of the V-PCC”, *ISO/IEC JTC 1/SC 29/WG 11 N19375, Online MPEG meeting*, Apr. 2020.
- [94] A. Collet et al., “High-Quality Streamable Free-Viewpoint Video”, *ACM Trans. Graphics (SIGGRAPH)*, vol. 34, no. 4, 2015.
- [95] E. Zerman, C. Ozcinar, P. Gaoy, A. Smolic, “Textured Mesh vs Coloured Point Cloud: A Subjective Study for Volumetric Video Compression”, *12th Int. Conf. on Quality of Multimedia Experience (QoMEX)*, 2020.
- [96] I. Sodagar, “The MPEG-DASH Standard for Multimedia Streaming Over the Internet”, *IEEE MultiMedia*, vo. 18 , no. 4, pp. 62-67, Apr. 2011.
- [97] J. van der Hooft, T. Wauters, F. De Turck, Ch. Timmerer, H. Hellwagner, “Towards 6DoF HTTP Adaptive Streaming Through Point Cloud Compression”, *MM ’19*, Nice, France, Oct., 2019.
- [98] P.-L. Hsieh et al., “Unconstrained real-time performance capture”, in *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [99] M. Zollhöfer et al., “State of the Art on Monocular 3D Face Reconstruction, Tracking, and Applications”, *Comput. Graph. Forum*, vol. 37, pp. 523-550, 2018.
- [100] A. Tewari et al., “High-Fidelity Monocular Face Reconstruction based on an Unsupervised Model-based Face Autoencoder”, *IEEE Trans. On Pattern Analysis and Machine Intelligence (PAMI)*, 2018.

- [101] A. Tkach, A. Tagliasacchi, E. Remelli, M. Pauly, A. Fitzgibbon, “Online generative model personalization for hand tracking”, *ACM Trans. On Graphics*, vol. 36, no. 6, 2017.
- [102] T. Alldieck, M. Magnor, B. Bhatnagar, C. Theobalt, and G. Pons-Moll, “Learning to reconstruct people in clothing from a single RGB camera”, in *Proc. Computer Vision and Pattern Recognition (CVPR)*, June 2019, pp. 1175–1186.
- [103] G. Pavlakos et al., “Expressive body capture: 3d hands, face, and body from a single image”, in *Proc. Computer Vision and Pattern Recognition (CVPR)*, Long Beach, USA, June 2019.
- [104] M. Habermann, W. Xu, M. Zollhöfer, G. Pons-Moll, and C. Theobalt, “Livecap: Real-time human performance capture from monocular video”, *ACM Trans. of Graphics*, vol. 38, no. 2, Mar. 2019.
- [105] T. Alldieck et al., “Detailed human avatars from monocular video”, in *Proc. Int. Conf. on 3D Vision (3DV)*, 2018.
- [106] D. Mehta et al., “VNect: Real-time 3D human pose estimation with a single RGB camera”, In *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, 2017.
- [107] A. Kanazawa et al., “End-to-End recovery of human shape and pose”, In *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [108] J. T. Barron et al., “Shape, illumination, and reflectance from shading”, in *Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 2015.
- [109] V.F. Abrevaya, S. Wuhrer, and E. Boyer, “Spatiotemporal Modeling for Efficient Registration of Dynamic 3D Faces”, in *Proc. Int. Conf. on 3D Vision (3DV)*, Verona, Italy, Sep. 2018, pp. 371–380.
- [110] M. Habermann et al., “NRST: Non-rigid surface tracking from monocular video”, in *Proc. GCPR*, 2018.
- [111] C. Vondrick et al., “Generating videos with scene dynamics”, in *Proc. Int. Conf. on Neural Information Processing Systems (NIPS)*, 2016.
- [112] P. Fechteler, A. Hilsmann, and P. Eisert, “Markerless Multiview Motion Capture with 3D Shape Model Adaptation”, *Computer Graphics Forum*, vol. 38, no. 6, pp. 91–109, Mar. 2019.
- [113] T. Zhou et al., “Learning data-driven reflectance priors for intrinsic image decomposition”, in *Proc. Int. Conf. on Computer Vision (ICCV)*, 2015.
- [114] L. Lettry, K. Vanhoey, L. van Gool, “DARN: A deep adversarial residual network for intrinsic image decomposition”, 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), March 2018.
- [115] R. Konstantinos et al., “Deep reflectance maps”, in *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [116] T. F. Cootes et al., “Active appearance models”, in *Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 23, no. 6, 2001.
- [117] L. Hu et al., “Avatar digitization from a single image for real-time rendering”, in *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, 2017.
- [118] S. Lombardi et al., “Deep appearance models for face rendering”, in *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, 2018.

- [119] K. Bousmalis et al., “Unsupervised pixel-level DA with generative adversarial networks”, in *Proc. Proc. Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [120] T.-C. Wang et al., “Video-to-video synthesis”, in *Proc. 32nd Int. Conf. on Neural Information Processing Systems (NIPS)*, pp. 1152-1164, 2018.
- [121] C. Finn et al., “Unsupervised learning for physical interaction through video prediction”, in *Proc. Int. Conf. on Neural Information Processing Systems (NIPS)*, 2016.
- [122] S.M. Ali Eslami et al., “Neural scene representation and rendering”, In *Science*, vol. 360, no. 6394, 2018.
- [123] Z. Zang et al., “Deep generative modeling for scene synthesis via hybrid representations”, in arXiv:1808.02084, 2018.
- [124] C. Bregler, M. Covell, and M. Slaney, “Video Rewrite: Driving Visual Speech with Audio”, in *ACM SIGGRAPH*, 1997.
- [125] A. Schodl, R. Szeliski, D. Salesin, and I. Essa, “Video Textures”, in *ACM SIGGRAPH*, 2000.
- [126] C. Malleson et al., “Facedirector: Continuous control of facial performance in video”, in *Proc. Int. Conf. on Computer Vision (ICCV)*, Santiago, Chile, Dec. 2015.
- [127] G. Fyffe, A. Jones, O. Alexander, R. Ichikari, and P. Debevec, “Driving highresolution facial scans with video performance capture”, *ACM Transactions on Graphics (TOG)*, vol. 34, no. 1, Nov. 2014.
- [128] J. Serra, O. Cetinaslan, S. Ravikumar, V. Orvalho, and D. Cosker, “Easy Generation of Facial Animation Using Motion Graphs”, *Computer Graphics Forum*, 2018.
- [129] F. Xu et al., “Video-based Characters - Creating New Human Performances from a Multiview Video Database”, in *ACM SIGGRAPH*, 2011.
- [130] A. Hilsmann, P. Fechteler, and P. Eisert, “Pose space image-based rendering”, *Computer Graphics Forum (Proc. Eurographics 2013)*, vol. 32, no. 2, pp. 265–274, May 2013.
- [131] W. Paier, M. Kettern, A. Hilsmann, and P. Eisert, “Hybrid approach for facial performance analysis and editing”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 784–797, Apr. 2017.
- [132] C. Bregler, M. Covell, and M. Slaney, “Video-based Character Animation”, In *Proc. the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2005.
- [133] P. Hilton, A. Hilton, and J. Starck, “Human Motion Synthesis from 3D Video”, In *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2009.
- [134] C. Stoll, J. Gall, E. de Aguiar, S. Thrun, and C. Theobalt, “Video-based reconstruction of animatable human characters”, *ACM Transactions on Graphics (Proc. SIGGRAPH ASIA 2010)*, vol. 29, no. 6, pp. 139–149, 2010.
- [135] D. Casas, M. Volino, J. Collomosse, and A. Hilton, “4d video textures for interactive character appearance”, *Computer Graphics Forum (Proc. Eurographics)*, vol. 33, no. 2, Apr. 2014.
- [136] M. Volino, P. Huang, and A. Hilton, “Online interactive 4d character animation”, in *Proc. Int. Conf. on 3D Web Technology (Web3D)*, Heraklion, Greece, June 2015.

- [137] A. Boukhayma and E. Boyer, "Video based animation synthesis with the essential graph", in *Proc. Int. Conf. on 3D Vision (3DV)*, Lyon, France, Oct. 2015, pp. 478–486.
- [138] A. Boukhayma and E. Boyer, "Surface motion capture animation synthesis", *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 6, pp. 2270–2283, June 2019.
- [139] J. Regateiro, M. Volino, and A. Hilton, "Hybrid skeleton driven surface registration for temporally consistent volumetric," in *Proc. Int. Conf. on 3D Vision (3DV)*, Verona, Italy, Sep. 2018.
- [140] C. Chan, S. Ginosar, T. Zhou, and A. Efros, "Everybody dance now", in *Proc. Int. Conf. on Computer Vision (ICCV)*, Seoul, Korea, Oct. 2019.
- [141] D. Mehta et al., "Vnect: Real-time 3d human pose estimation with a single RGB camera", in *Proc. Computer Graphics (SIGGRAPH)*, vol. 36, no. 4, July 2017.
- [142] L. Liu et al., "Neural rendering and reenactment of human actor videos", *ACM Trans. of Graphics*, 2019.
- [143] H. Kim et al., "Deep video portraits", *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, p. 163, 2018.
- [144] A. J. Davidson, "Real-Time Simultaneous Localisation and Mapping with a Single Camera", *IEEE Int. Conf. on Computer Vision (ICCV)*, 2003.
- [145] Georg Klein and David Murray, "Parallel Tracking and Mapping for Small AR Workspaces", in *Proc. International Symposium on Mixed and Augmented Reality (ISMAR'07)*, 2007.
- [146] N. Radwan, A. Valada, W. Burgard, "VLocNet++: Deep Multitask Learning for Semantic Visual Localization and Odometry", in *IEEE Robotics and Automation Letters*, vol. 3, no. 4, Oct. 2018
- [147] L. Sheng, D. Xu, W. Ouyang, X. Wang, "Unsupervised Collaborative Learning of Keyframe Detection and VisualOdometry Towards Monocular Deep SLAM", *ICCV 2019*.
- [148] M. Bloesch, T. Laidlow, R. Clark, S. Leutenegger, A. Davison, "Learning Meshes for Dense Visual SLAM", *ICCV 2019*.
- [149] N.-D. Duong, C. Soladié, A. Kacète, P.-Y. Richard, J. Royan, "Efficient multi-output scene coordinate prediction for fast and accurate camera relocalization from a single RGB image", in *Computer Vision and Image Understanding*, vol. 190, Jan. 2020.
- [150] O. Schreer et al., "Capture and 3d video processing of volumetric video", *2019 IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan, 2019, pp. 4310-4314, doi: 10.1109/ICIP.2019.8803576.
- [151] M. Kazhdan, H. Hoppe, "Screened Poisson Surface Reconstruction", *ACM Transactions on Graphics (TOG)*, vol. 32, no. 3, 2013, doi: 10.1145/2487228.2487237.
- [152] M. Garland, P. S. Heckbert, "Surface simplification using quadric error metrics", in *SIGGRAPH '97, Proc. of the 24th annual conference on Computer graphics and interactive techniques*, New York, USA, 1997, pp. 209-216, doi: 10.1145/258734.258849
- [153] R. Diaz, et al., "Region Dependent Mesh Refinement for Volumetric Video Workflows", *2019 International Conference on 3D Immersion (IC3D). IEEE*, 2019.
- [154] W. Morgenstern, A. Hilsman, P. Eisert, "Progressive non-rigid registration of temporal mesh sequences", In *Proc. Europ. Conf.on Visual Media Production (CVMP)*, London, UK, 2019.

- [155] S. Gül, et al., "Interactive Volumetric Video from the Cloud", *Int. Broadcasting Convention (IBC)*, Amsterdam, Netherlands, Sept. 2020.

4.4 3D sound processing algorithms

Currently, three general concepts exist for storing, coding, reproducing, and rendering spatial audio, all based on multichannel audio files: channel based, Ambisonics based, and object based. A concise overview of the currently used formats and platforms is given in [156] and [157].

4.4.1 Channel-based audio formats and rendering

The oldest and, for XR, already a bit outdated method of spatial audio is the channel-based reproduction. Every audio channel in a sound file directly relates to a loudspeaker in a pre-defined setup. Stereo files are the most famous channel-based format. Here, the left and right loudspeakers are supposed to be set up with an angle of 60°. More immersive formats are the common 360-degree surround formats such as 5.1 and 7.1, as well as the 3D formats like Auro 3D 9.1 - 13.1. All these formats use virtual sound sources, also referred to as phantom sources. This means they send correlated signals to two or more loudspeakers to simulate sound sources between the loudspeaker positions.

Thus, for classical formats such as stereo, 5.1, and 7.1, the rendering process happens before the file is stored. During playback the audio only needs to be sent to the correct loudspeaker arrangement. Therefore, the loudspeakers have to be located correctly to perceive the spatial audio correctly. Dolby Atmos and Auro 3D extend this concept by also including the option for object-based real-time rendering.

To reproduce new audio sources between the pre-defined loudspeakers, different approaches can be used. In general, they all satisfy the constraint for equal loudness, which means that the energy of a source stays the same regardless of their positions. Vector-based amplitude panning (VBAP) spans consecutive triangles between three neighbouring loudspeakers [158]. A position of a source is described by a vector from the listener position to the source position and the affected triangle is selected on the basis of this vector. The gain factors are calculated for the loudspeakers spanning the selected triangle under the previously mentioned loudness constraint. This is a very simple and fast calculation. By contrast, distance-based amplitude panning (DBAP) utilises the Euclidean distance from a source to the different speakers and makes no assumption about the position of the listener [159]. These distances build up a ratio calculating a gain factors, again under the previously-mentioned loudness constraint. By contrast with VBAP, mostly all loudspeakers are active for a sound source in DBAP.

Both of these methods create virtual sound sources between loudspeaker positions. This causes some problems. Firstly, the listener has to be at special places (the so-called sweet spots) to get the correct signal mixture, allowing only a few persons to experience the correct spatial auralisation. Because in VBAP a source is only played back by a maximum of three loudspeakers, this problem is much more present in VBAP than in DBAP. Secondly, a virtual sound source matches the ILD and ITD cues of human audio perception correctly (see section 4.2.1), but it might be in conflict with reproduction of the correct HRTF and can therefore cause spatially-blurred and spectrally-distorted representations of the acoustic situation.

4.4.2 Ambisonics-based formats and rendering

Another method of storing 3D audio is the Ambisonics format (also see section 4.2.4). The advantage of Ambisonics-based files over channel-based files is their flexibility with respect to a playback on any loudspeaker configuration. However, the necessity for a decoder also increases the complexity and the amount of computation. There are currently two main formats used for Ambisonics coding; they differ in the channel ordering and weighting: AmbiX (SN3D encoding) and Furse-Malham Ambisonics (maxN encoding).

In contrast to the VBAP and DBAP rendering methods of channel-based formats (see section 4.4.1), which implement a spatial auralisation from a hearing-related model approach, Ambisonics-based rendering and wave field synthesis (see section 4.4.4) use a physical reproduction model of the wave field [160]. There are two common, frequently-used approaches for designing Ambisonic decoders. One approach is to sample the spherical harmonic excitation individually for the given loudspeaker's positions. The other approach is known as mode-matching. It aims at matching the spherical harmonic modes excited by the loudspeaker signals with the modes of the Ambisonic sound field decomposition [160]. Both decoding approaches work well with spherically and uniformly distributed loudspeaker setups. However, non-uniform distributed setups, require correction factors for energy preservation. Again, see [160] for more details. 3D Rapture by Blue Ripple Sound is one of the current state-of-the-art HOA decoders for XR applications. Other tools are IEM AllRADecoder, AmbiX by Matthias Kronlachner and Harpex-X.

4.4.3 Binaural rendering

Most XR applications use headsets. This narrows down the playback setup to the simple loudspeaker arrangement of headphones. Hence, a dynamical binaural renderer achieves spatial aural perception over headphones by using an HRTF-based technique (described in 4.2.1). The encoded spatial audio file gets decoded to a fixed setup of virtual speakers, arranged spherically around the listener. These virtual mixings are convolved with a direction-specific Head Related Impulse Response (HRIR). Depending on the head position, the spatial audio representation is rotated before being sent to the virtual speakers. New methods propose a convolution with higher-order Ambisonics HRIRs without the intermediate step of a virtual speaker down mix [160]. When the proper audio formats and HRIRs with a high spatial resolution are used, a very realistic audible image can be achieved. Facebook (Two Big Ears) and YouTube (AmbiX) developed their own dynamical binaural renderer using first- and second-order Ambisonics extensions [161][162].

4.4.4 Object based formats and rendering

The most recent concept for 3D audio formats uses an object-based approach. Every sound source is assigned to its own channel with dynamic positioning data encoded as metadata. Hence, in contrast to the other formats, exact information about location, angle, and distance to the listener are available. This allows maximum flexibility during rendering, because, in contrast to the previously-mentioned formats, the position of the listener can easily be changed relatively to the known location and orientation of the source. However, for complex scenes, the number of channels and, with it, the complexity are growing considerably, and, similar to Ambisonics, a special decoding process is needed, with the amount of computation increasing proportionally to the number of objects in the scene. Furthermore, complex sound sources such as reverberation patterns caused by reflections in the environment cannot yet be

represented accurately in this format, because they depend on complex scene properties rather than the source and listener positions only. Furthermore, there is currently no standardised format specialised for object-based audio. In practice, the audio data is stored as multichannel audio files with an additional file storing the location data.

One well-known rendering concept for object-based audio formats is the Wave Field Synthesis (WFS) developed by Fraunhofer IDMT. It enables the synthesis of a complete sound field from its boundary conditions [163]. In theory, a physically correct sound field can be reconstructed with this technology, eliminating all ILD, ITD, HRTF, and sweet spot related artefacts (see section 4.2.1). In contrast to other rendering methods, the spatial audio reproduction is strictly based on locations and not on orientations. Hence, it even allows for positioning sound sources inside the sound field. Supposing that multiple Impulse Responses (IR) of the environment are known or can be created virtually using ray tracing models, WFS even enables one to render any acoustic environment onto the sound scene.

For a correct physical sound field in the range of the human audible frequency span, a loudspeaker ring around the sound field with a distance between the loudspeakers of 2 cm is needed [164]. As these conditions are not realistic in practice, different approximations have been developed to alleviate the requirements on loudspeaker distance and the number of required loudspeakers.

4.4.5 Combined applications

State-of-the-art-formats combine the qualities of the previously mentioned concepts depending on the use case. Current standards are Dolby Atmos and AuroMAX for cinema and home theatres, Two Big Ears by Facebook for web-based applications and the MPEG-H standard for generic applications. MPEG-H 3D Audio, developed by Fraunhofer IIS for streaming and broadcast applications, combines basic channel-based, Ambisonics-based, and object-based audio, and can be decoded to any loudspeaker configuration as well as to binaural headphone [165].

Besides being used in cinema and TV, 3D auralisation can also be used for VR. In particular, the VR players used in game engines are suitable tools for the creation of 3D auralisation. These players offer flexible interfaces to their internal object-based structure allowing the integration of several formats for dynamic 3D sound spatialization. Most game engines already support a spatial audio implementation and come with preinstalled binaural and surround renderers. For instance, Oculus Audio SDK is one of the standards being used for binaural audio rendering in engines like Unity and Unreal. Google Resonance, Dear VR, and Rapture3D are sophisticated 3D sound spatialisation tools, which connect to the interfaces of common game engines and even to audio specific middleware like Wwise and Fmod providing much more complex audio processing.

In general, VR player and gaming engines use an object-based workflow for auralisation. Audio sources are attached over interfaces to objects or actors of the VR scene. Assigned metadata are used to estimate localisation and distance-based attenuation as well as reverberation and even Doppler effects. The timing of reflection patterns and reverberations is calculated depending on the geometry of the surrounding, their materials as well as the positions of the sound sources and the listener. Filter models for distance-based air dissipation are applied, as well as classical volume attenuation. Sound sources contain a directivity pattern, changing their volume depending on their orientation to the listener. Previously-

mentioned middleware can extend this processing further to create a highly unique and detailed 3D auralisation.

The whole object-based audio scene is then usually rendered into a HOA format, where environmental soundscapes not linked to specific scene objects (e.g. urban atmosphere) can be added to the scene. The whole HOA scene can be rotated in accordance to the head tracking of the XR headset and is then rendered as a binaural mixture as described in section 4.4.3.

4.4.6 References

- [156] “Virtual Reality audio formate – Pros & Cons.” VRTONUNG. <https://www.vrtonung.de/en/virtual-reality-audio-formats/> (accessed Nov. 11, 2020).
- [157] “360-Grad Videos für Virtual Reality Plattformen und VR-player.” VRTONUNG. <https://www.vrtonung.de/en/spatial-audio-support-360-videos/> (accessed Nov. 11, 2020).
- [158] V. Pulkk, “Spatial sound generation and perception by amplitude panning techniques”, PhD thesis, Helsinki University of Technology, 2001.
- [159] T. Lossius, P. Baltazar, T.de la Hogue, “DBAP–distance-based amplitude panning”, in *Proc. Of Int. Computer Music Conf. (ICMC)*, 2009.
- [160] F. Zotter, H. Pomberger, M. Noisternig, “Ambisonic decoding with and without mode-matching: A case study using the hemisphere”, in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, Vol. 2, 2010.
- [161] Facebook Audio 360. <https://facebookincubator.github.io/facebook-360-spatial-workstation/> (accessed Nov. 11, 2020).
- [162] “Use spatial audio in 360-degree and VR videos.” Youtube Help. <https://support.google.com/youtube/answer/6395969?co=GENIE.Platform%3DDesktop&hl=en> (accessed Nov. 11, 2020).
- [163] T. Ziemer, “Wave Field Synthesis”, in *Springer Handbook of Systematic Musicology*, R. Bader, Ed., Springer Berlin Heidelberg, 2018.
- [164] R. Rabenstein and S. Spors, “Spatial aliasing artefacts produced by linear and circular loudspeaker arrays used for wave field synthesis”, in *120th Audio Engineering Society Convention*, May 2006.
- [165] Fraunhofer IIS. <https://www.iis.fraunhofer.de/en/ff/amm/broadcast-streaming/mpegh.html> (accessed Nov. 11, 2020).

4.5 Interactive technologies for virtual flavour

A very interesting domain in XR technologies is the simulation of taste as another cue beside audio, visual and haptic. In the section below, the current state of taste simulation is given.

The molecules of food are chemicals detected by taste receptors in the mouth, and the olfactory receptors in the nose. There are five primary tastes: salty, sour, bitter, sweet and umami (from the Japanese for “tasty” - which corresponds roughly to the taste of glutamate) [166]. How we perceive food is also influenced by its texture, smell (both orthonasal (“sniffed in”) and retronasal (“from the food in the mouth”), temperature, looks, cost, and environmental factors, such as where we are eating and with whom, etc, eg [167][168][169].

In 2003, Iwata et al [170] presented their three-sense food simulator: a haptic interface to mimic the taste, sound and feeling of chewing real food. A mouth device simulated the force of the food type, a bone vibration microphone provided the sound of biting, while chemical simulation of taste was achieved via a micro injector which squirted the chemicals into the mouth. Very recently, Miyashita demonstrated the “Norimaki taste display” [171] using five gels to recreate basic tastes. Although highly novel, these devices do not include mouthfeel or aroma, key components of flavour. Work from Ranasinghe et al. [172] has shown that it is possible to simulate the sensation of some of the primary tastes by direct electrical and thermal stimulation of the tongue. This work has led to the development of virtual cocktail device [173]. However, this device can only simulate a few flavours. Electrical stimulation has also been used to attempt the simulation of smell, with limited success so far [174]. In 2010, Narumi et al. [175] showed how cross-sensory perception can influence enjoyment of food by superimposing virtual colour onto a real drink, while the MetaCookie+ project [176] changed the perceived taste of a cookie using visual and auditory stimuli. How multisensory stimuli, in particular visuals, audio, smell and motion, may affect a real experience (singularly or in combination) has been studied extensively, eg [177], including their impact on flavour perception, eg [166][169].

Virtual Reality has recently been used to see how an environment can affect the perception of flavour [178][179], however the tastes used in these studies (berry-flavoured beverage, blue cheese) were real and not simulated. The concept of “virtual flavour” was showcased by Chalmers at the British Science Festival on 10 September 2019, (see Figure 19) [180][181]. Their FlaVR concept comprises a soft “mouth-guard-like” device in the mouth for delivering taste, and a small tube just in front of the user’s nose for delivering smell. Flavour information (similar to a recipe) is extracted by software from a previously prepared flavour database, in harmony with the experience, and used to create the virtual sample including visuals, taste, mouthfeel, and aroma “on the fly” at the right precision [182] and deliver this to the user.

The inclusion of taste and smell within virtual environments has the potential to significantly enhance the immersion and indeed “authenticity” of any virtual experience [183]. Humans perceive the real world with all our senses. Failure to include any of these senses risks misrepresenting reality in the virtual experience [184].



Figure 19: FlaVR at the British Science Festival in September 2019.

4.5.1 References

- [166] B. Piqueras-Fiszman, C. Spence C. (eds), “Multisensory Flavor Perception: From Fundamental Neuroscience Through to the Marketplace”, Woodhead Publishing, 2016.
- [167] J. Delwiche, “The impact of perceptual interactions on perceived flavour”, *Food Q & P*, 15, 2004.
- [168] E. Rolls, “Taste, olfactory, and food reward value processing in the brain”, *Prog Neurobiol*, 127, 2015.
- [169] C. Spence, B. Piqueras-Fiszman, “The Perfect Meal: The multisensory science of food and dining”, 2017.
- [170] H. Iwata, H. Yano, T. Uemura, T. Moriya, “Food simulator”, In ICAT’03: Proceedings of the 13th International Conference on Artificial Reality and Telexistence, IEEE, 2003.
- [171] H. Miyashita, “Norimaki Synthesizer: Taste Display Using Ion Electrophoresis in Five Gels”, ACH CHI, 2020.
- [172] N. Ranasinghe, A. Cheok, R. Nakatsu, E. Yi-Luen Do, “Simulating the sensation of taste for immersive experiences”, ImmersiveMe 2013, ACM Multimedia, 2013.
- [173] N. Ranasinghe, T.N.T. Nguyen, Y. Liangkun, E. DoEllen, Y. Do, “Vocktail: A Virtual Cocktail for Pairing Digital Taste, Smell, and Color Sensations”, MM 2017, October 2017.
- [174] S. Hariri, N. Mustafa, K. Karunanayaka, A. D. Cheok, “Electrical Stimulation of Olfactory Receptors for Digitizing Smell”, HAI ’16, Singapore, October 2016.
- [175] T. Narumi, M. Sato, T. Tanikawa, M. Hirose, “Evaluating cross-sensory perception of superimposing virtual color onto real drink”, 1st Augmented HCI, 2010.
- [176] T. Narumi, S. Nishizaka, T. Kajinami, T. Tanikawa, M. Hirose, “MetaCookie+”, IEEE VR, 2011.
- [177] G. Calvert, C. Spence, B. Stein, *The multisensory handbook*. MIT Press, 2004.
- [178] Y. Chen et al. “Assessing the Influence of Visual-Taste Congruency on Perceived Sweetness and Product Liking in Immersive VR”, *Food* 9(4), April 2020.
- [179] A. Stelick, A. Penano, A. Riak, R. Dando, “Dynamic Context Sensory Testing—A Proof of Concept Study Bringing Virtual Reality to the Sensory Booth”, *Journal of Food Science*, 2018.
- [180] “Time for Tea”, British Science Festival, September 2019.
- [181] A. Chalmers, J. Gain, “Royal Academy of Engineering grant IAPP18-1989”, 2019.
- [182] E. Doukakis, K. Debattista, T. Bashford-Rogers, D. Dhokia, A. Asadipour, A. Chalmers, H. Harvey, “Audio-Visual-Olfactory Resource Allocation for Tri-modal Virtual Environments”, *Transactions on Visualization and Computer Graphics*, Vol.25 (5), May 2019, pp.1865–1875.
- [183] M. Melo, G. Gonçalves, P. Monteiro, H. Coelho, J. Vasconcelos-Raposo, M. Bessa, “Do Multisensory stimuli benefit the virtual reality experience? A systematic review”, *IEEE Transactions on Visualization and Computer Graphics*, doi: 10.1109/TVCG.2020.3010088.
- [184] A. Chalmers, D. Howard, C. Moir, “Real Virtuality: A step change from Virtual Reality”, Spring Conference on Computer Graphics (SCCG’09), pp 15-22, ACM SIGGRAPH Press, 2009.

4.6 Input and output devices

The user acceptance of immersive XR experiences is strongly connected to the quality of the hardware used, in particular of the input and output devices, which are generally the ones available on the consumer electronics market. In this context, the hardware for immersive experiences can be divided in four main categories:

- In the past, immersive experiences were presented using complex and expensive devices, systems such as 3D displays or multi-projection systems like “Cave Automatic Virtual Environment” (CAVE) (see section 4.6.1).
- Nowadays, especially since the launch of the Oculus DK1 in March 2013, most VR applications used head-mounted displays (HMDs) or VR headsets such that the user is fully immersed in a virtual environment, i.e. without any perception of the real world around him/her (see section 4.6.2).
- By contrast, AR applications seamlessly insert computer graphics into the real world, by using either (1) special look-through glasses like HoloLens or (2) displays/screens (of smartphones, tablets, or computers) fed with real-time videos from cameras attached to them (see section 4.6.3).
- Most VR headsets and AR devices use haptic and sensing technologies to control the visual presentation in dependence of the user position, to support free navigation in the virtual or augmented world and to allow interaction with the content (see section 4.6.4).

4.6.1 Stereoscopic 3D displays and projections

Stereoscopic 3D (S3D) has been used for decades for the visualisation of immersive media. For a long time, the CAVE (Cave Automatic Virtual Environment) technology was its most relevant representative for VR applications in commerce, industry, and academia, among others [185][186]. A single user enters in a large cube, where all, or most, of the 6 walls are projection screens made of glass, which imagery or video is projecting on, preferably in S3D. The user is tracked and the imagery adjusted in real-time, such that he/she has the visual impression of entering a cave-like room showing a completely new and virtual world. Often, the CAVE multi-projection system is combined with haptic controllers to allow the user to interact with the virtual world. Appropriate spatial 3D sound can be added to enhance the experience, whenever this makes sense.

More generally, S3D technologies can be divided in two main categories: glasses-based stereoscopy (where “glasses” refers to special 3D glasses) and auto-stereoscopy.

The glasses-based systems include those typically found in 3D cinemas. Their purpose is to separate the images for the left and the right eye. These glasses can be classified as passive or active. Passive glasses use a variety of image-separating mechanisms, mainly polarisation filters or optical colour filters (which include the customary anaglyphic technique, typically implemented through the ubiquitous red & blue plastic lenses). Active glasses are based on time multiplexing using shutters. Passive glasses are cheap, lightweight, and do not require any maintenance. Active glasses are much more expensive, heavier, and require their batteries to be changed.

By contrast, auto-stereoscopic 3D displays avoid the need to wear glasses. The view separation is achieved by special, separating optical plates directly bonded to the screen. These plates are designed to provide the left and right views for a given viewing position of the user, or for multiple such positions. In the last case, several viewers can possibly take place at the “sweet spots” where the 3DS visual perception is

correct. In this way, auto-stereoscopic display can provide a fixed number of views, such as 1, 3, 5, 21, and even more. For a given screen, the resolution decreases as the number of views increases. The above plates are generally implemented using lenticular filters placed in vertical bands, at a slight angle, on the display. This is similar to printed (static) photos that show a 3D effect, for one or more views. Some advanced auto-stereoscopic displays are designed to track a single user and to display the correct S3D view independently of his/her position.

The most sophisticated displays are the so-called light-field displays. In theory, they are based on a full description of the light in a 7-dimensional field. Among other things, such a display must be able to fully control the characteristics of the light in each and every direction in a hemisphere at each of its millions of pixels.

Of course, for each of the above types of “3D” visualisation systems, one must have the corresponding equipment to provide i.e. the corresponding cameras. For example, one needs, in the case of real images (as opposed to synthetic, computer-made images), a light-field camera to provide content for a light-field display. A more detailed description about the different S3D display technologies is provided by Urey et al. [187].

Since the 1950s, S3D viewing has seen several phases of popularity and a corresponding explosion of enthusiasm, each triggered by a significant advance in technology. The last wave of interest (roughly from 2008 to 2016) was triggered by the arrival of digital cinema, which allowed for an unprecedented control of the quality of S3D visualisation. Each such wave came with extreme and unwarranted expectation. During, the last wave, TV manufacturers succeeded for a while in convincing every consumer to replace their conventional TV with a new one allowing for S3D viewing. However, today, it is hard to find a new TV offering such capability.

Most international consumer-equipment manufacturers have stopped their engagement in 3D displays. It is only in 3D cinema and in some niche markets that stereoscopic displays have survived. This being said, S3D remains a key factor of immersion, and this will always be the case. Today, most quality XR systems use S3D.

Nevertheless, in case of auto-stereoscopy, some recent progress has been made possible by high-resolution display panels (8K pixels and beyond) as well as by OLED technology and light-field technology. An example pointing to this direction is the current display generation from the Dutch company Dimenco (for a while part of the Chinese company KDX [188]), called Simulated Reality Display and demonstrated successfully at CES 2019 [189]. Similar to the earlier tracked auto-stereoscopic 3D displays, as for example Fraunhofer HHI’s Free3C Display [190] launched as a very first research system almost 15 years ago, the Simulated Reality Display is equipped with additional input devices for eye- and hand-tracking to enable natural user interaction. The main breakthrough, however, is the usage of panels of 8K and more providing a convincing immersive S3D experience from a multitude of viewpoints. Several other European SMEs, like SeeFront, 3D Global, and Alioscopy, offer similar solutions.

4.6.2 VR Headsets

In contrast to the former usage of the CAVE technology (see section 4.8.1), today most VR applications focus on headsets. Since the acquisition of Oculus VR by Facebook for 2 billion US dollars in 2014, the sales

market of VR headsets has been steadily growing [191]. At the gaming platform Steam, the yearly growing rate of monthly-connected headsets is even up to 80% [192].

There are many different types of VR headsets ranging from smartphone-based mobile systems (e.g. Samsung Gear VR) through console-based systems (e.g. Sony PlayStation VR) and PC-based Systems (e.g. HTC Vive Cosmos and Facebook Oculus S), to the new generation of standalone systems (e.g. Facebook Oculus Quest and Lenovo Mirage Solo). In this context, the business strategy of Sony is noteworthy. The company has strictly continued to use its advantages in the gaming space and, with it, to pitch PlayStation VR to their customers. Unlike HTC Vive and Oculus Rift, users in the high-performance VR domain only need a PlayStation 4 instead of an expensive gaming PC.

Figure 20 shows the enormous progress VR headsets made during the last five years. The first Oculus DK1 had a resolution of only 640 x 800 pixels per eye, a horizontal field of view of 90 degrees, and a refresh rate of 60 frames per second. These characteristics were far away from the ones needed to meet the requirements of the human visual system, i.e. the resolution of 8000 x 4000 pixels per eye, a horizontal field of view of 120 degrees, and the refresh rate of 120 frames per second. By comparison, state-of-the-art headsets like the Oculus Rift S and HTC Vive Cosmos have a resolution of up to 1440 x 1700 pixels per eye, a horizontal field of view over 100 degrees, and at a refresh rate of 80-90 frames per second. This is certainly not yet enough compared to the properties of the human visual system, but this shows that the three main display parameters (resolution, field of view and refresh rate) have been improving. Besides the main players, there are plenty of smaller HMD manufacturers offering various devices and sometimes special functionalities. For instance, the Pimax 8K headset provides the highest resolution on the market, with about 4000 x 2000 pixels per eye, i.e. half of what is needed, and the Valve Index provides the highest refresh rate on the market, with up to 144 Hz, i.e. even more than what is needed.

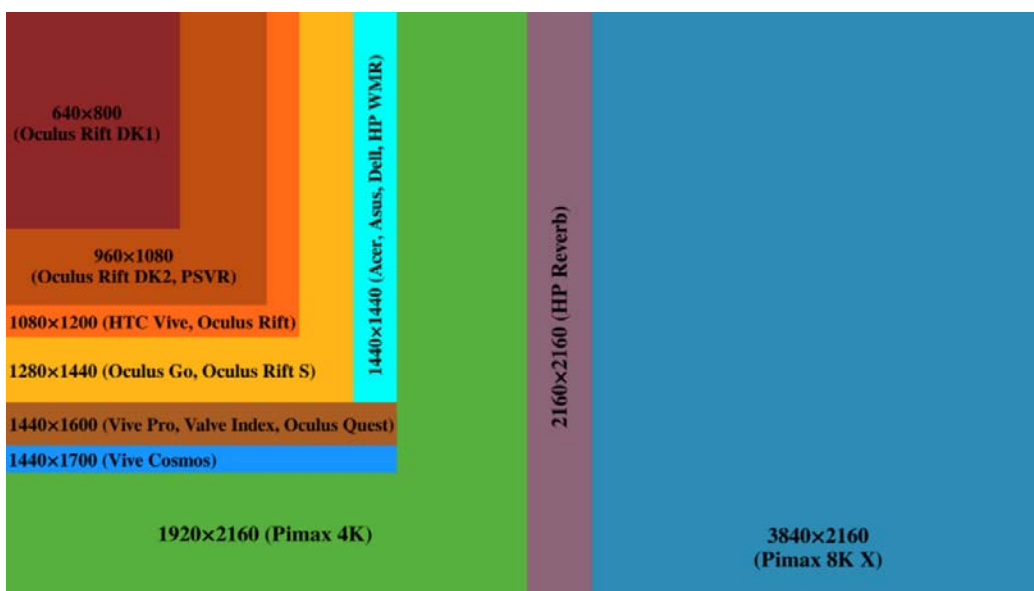


Figure 20: Comparison chart of VR headset resolutions [193].

Another interesting market development is the new generation of untethered, standalone VR headsets that were launched 2019 (e.g. Oculus Quest). These headsets are very promising for the new upcoming VR ecosystems requiring a movement with six degrees of freedom (6DoF) without external tracking and computing systems, but with high VR performance comparable to the ones of tethered headsets. Systems like Oculus Quest have image-based, inside out-tracking systems as well as sufficient computing power on board, and they need neither cable connections to external devices nor external, outside-in tracking systems. They nevertheless have VR performances that are comparable to the ones of their tethered counterpart. Because of their ability to provide, at lower cost, excellent performance with a simple setup in any environment, they address completely new groups of VR users.

4.6.3 AR Systems

Two main classifications are generally used for AR systems. In the first classification of AR systems, one classifies the systems according to the strategy for combining virtual elements with the user's perception of the real environment:

- Video see-through (VST): First, the AR system captures the real environment with vision sensors, i.e. cameras. Second, the digital content (representing the virtual elements of interest) are then combined with the images captured by these vision sensors. Third, the resulting image is displayed to the user thanks to an opaque screen. Smartphones and tablets fall into this category;
- Optical see-through (OST): The AR system displays digital content directly on a transparent screen allowing the user to perceive the real world naturally. These transparent screens are mainly composed of a transparent waveguide that transports the light emitted by a microscreen, which is placed around the optical system and outside the field of view, in such a way that the image on the microscreen arrives on the retina of the user. The physical properties of the materials composing these waveguides theoretically limit the field of view of such a system to 60°. Other solutions are based on a beam-splitter, which is a kind of semi-transparent mirror that reflects the image of the microscreen. A system using beam-splitter is generally bulkier than a system using waveguide. A beam-splitter can offer a wide field of view, but with a relatively small accommodation distance. AR headsets and glasses using transparent displays fall into this category;
- Projective see-through (PST): The AR system project the digital content directly on the surface of the elements of the real environment. This technique, called projection mapping, is widely used to create shows on building facades, and it can also be used to display assembly instructions in manufacturing.

In the second classification of AR systems, one classifies the systems according to the position of the display with respect to the user:

- Near-eye display: The display is positioned a few centimetres from the user's retina, and is integrated either into a relatively large headset or into glasses with a lighter form factor;
- Handheld display: The display is held in the user's hands. Handheld AR display systems are widely used through smartphones and tablets, but they do not allow the user's hands to be free;
- Distant display: The display is placed in the user's environment, but it is not attached to the user. These systems require to track the user to ensure a good registration of the digital content on the real world.

Today, the most frequently used hardware devices for AR applications are smartphones and tablets (handheld video-see-through displays). In this cases, special development toolkits like the Apple ARKit at

Apple's iOS allow the augmentation of live camera views with seamlessly integrated graphical elements or computer animated characters. Today, approximately 48 million US broadband households currently have access to Apple's ARKit platform via a compatible iPhone.

As a result, the commercialisation of the successive version of Google Glass and more recently the Microsoft HoloLens, AR glasses and headsets (near-eye see-through displays) are beginning to spread out, mainly targeting the professional market.

Indeed, the first smart glasses were introduced by Google in 2012. These monocular smart glasses that simply overlay the real-world vision with graphical information have often been considered more like a head-up display than a real AR system, because they do not provide true 3D registration. At that time, smart glasses generated a media hype leading to a public debate on privacy. Indeed, these glasses, often worn in public spaces, continuously recorded the user's environment through to their built-in camera. In 2015, Google quietly dropped these glasses from sale and relaunched a "Google Glass for Enterprise Edition" version in 2017 aiming at factory and warehouse usage scenarios. However, among all AR platforms they have tested, consumers report the highest level of familiarity with the Google Glass, even though Google stopped selling these devices in early 2015 [194].

To this day, the best-known representative of high-end stereoscopic 3D AR glasses is the HoloLens from Microsoft, which inserts graphical objects or 3D characters under right perspective seamlessly into the real-world view with true 3D registration. Another such high-end device is the one being developed by Magic Leap, a US company that was founded in 2014 that has received a total funding of more than 2 billion US dollars. Despite this astronomical investment, the launch in 2019 of the Magic Leap 1 glasses did not meet the expectations of AR industry experts, although some of their specifications seemed than those of the existing HoloLens. Furthermore, in February 2019, Microsoft announced the new HoloLens 2 and the first comparisons with the Magic Leap 1 glasses seem to confirm that Microsoft currently dominates the AR field. Magic Leap itself admits that they have been leapfrogged by HoloLens 2 [195]. Rumours indicate that Microsoft purposely delayed the commercialisation of HoloLens 2 until the arrival of the Magic Leap 1 glasses, precisely to stress their dominance of the AR headset market.

Although HoloLens 2 is certainly the best and most used high-end stereoscopic AR headset, it is still limited in terms of image contrast, especially in brighter conditions, field of view, and battery life. The problem is that the all complex computing like image-based inside-out tracking and high-performance rendering of graphical elements has to be carried out on-board. The related electronics and the needed power supply have to be integrated in smart devices with extremely small form factors.

One alternative is to offload the bulk of computing to another device like a smartphone and use the AR glasses primarily for pose tracking and display of the rendered view. First example of this approach is the Nreal Light glasses [196] that are tethered via USB-C to either a high-end recent Android phone or an Nreal computing unit. The lightweight glasses (only 88 g) enable prolonged use and targets consumer market as opposed to bulkier devices like HoloLens whose target applications are primarily in enterprise domain. Another solution might be the combination with upcoming wireless technology, i.e. the 5G standard, and its capability to outsource complex computations to the network edge while keeping the low latency and fast response needed for interactivity (see section 4.7).

Very recent VR headsets integrate two cameras, with one in front of each user's eye. Each such camera captures the real environment as seen by each eye and displays the corresponding video on the

corresponding built-in screen. Such near-eye video see-through systems can address both AR and VR applications and are thus considered as mixed-reality (MR) systems.

All these AR systems offering a true 3D registration of digital content on the real environment are using processing capabilities, built-in cameras and sensing technology that have been originally developed for handheld devices. In particular, the true 3D registration is generally achieved by using inside-out tracking, as is implemented in the technique called “Simultaneous Localisation and Mapping (SLAM)” (see section 4.3).

Parks Associates reported in April 2019 that the total installed base of AR head-mounted devices will rise from 0.3 M units in 2018 to 6.5 M units by 2025 [197]. In the future, AR applications will certainly use more head-mounted AR devices, but these applications will most likely be aimed at applications in industry for quite some time.

4.6.4 Sensing and haptic devices

Sensing systems are the key technologies in all XR applications. A key role of sensing is the automatic determination of the user’s position and orientation in the given environment. In contrast to handheld devices like smartphones, tablets, laptops, and gaming PC’s, where the user navigation is controlled manually by mouse, touch pad or game controller, the user movement is automatically tracked in case of VR or AR headsets or even of former VR systems like CAVEs (see section 4.6.1). The first generations of VR headsets (e.g. HTC Vive) use external tracking systems for this purpose. For instance, the tracking of HTC Vive headset is based on the Lighthouse system, where two or more base stations arranged around the user’s navigation area emit laser rays to track the exact position of the headset. Other systems like Oculus Rift use a combination of onboard sensors like gyroscope, accelerometer, magnetometers and cameras to track head and other user movements.

High-performance AR systems like HoloLens and recently launched standalone VR headsets use video-based inside-out tracking. This type of tracking is based on several on-board cameras that analyse the real- world environment, often in combination with additional depth sensors. The user position is then calculated in relation to previously analysed spatial anchor points in the real 3D world. By contrast, location-based VR entertainment systems (e.g. The Void [198]) use outside-in tracking, the counterpart to inside-out tracking. In this case many sensors or cameras are mounted on the walls and ceiling of a large-scale environment that may cover several rooms. In this case, the usual headsets are extended by special markers or receivers that can be tracked by the outside sensors. Some more basic systems even use inside-out tracking for location-based entertainment. In this case many ID-markers are mounted on the walls, floor, and ceiling, while on-board cameras on the headset determine its position relatively to the markers (e.g. Illusion Walk [199]).

Apart from position tracking, other sensing systems track special body movements automatically. The best know example is that of hand-tracking systems that allow the user to interact in a natural interaction way with the objects in the scene (e.g. Leap Motion by Ultraleap [200]). Usually, these systems are external, accessory devices that are mounted on headsets and are connected via USB to the rendering engine. The hand tracker of Leap Motion, for instance, uses infrared-based technologies, where LEDs emit infrared light to detect hands and an infrared camera to track and visualise them. However, in some recently-launched systems like standalone VR or AR headsets (e.g. Oculus Quest and HoloLens 2), hand (and even finger) tracking are already fully integrated.

Another example of sensing particular body movements is the eye and gaze tracker that can be used to detect the user's viewing direction and, with it, which scene object attracts the user's attention and interest. A prominent example is Tobii VR, which has also been integrated in the new HTC Vive Pro Eye [201]. It supports foveated rendering to render those scene parts of the scene with more accuracy where the user is looking at than for other parts. Another application is natural aiming, where the user can interact with the scene and its objects by just looking in particular directions, i.e. via his/her gaze.

Beside the above sensing technologies, which are quite natural and, now, often fully integrated in headsets, VR and AR applications can also use a variety of external haptic devices. In this context, the most frequently used devices are hand controllers, which are usually delivered together with the specific headset. Holding one controller in one hand, or two controllers in the two hands, users can interact with the scene. The user can jump to other places in the scene by so-called "teleportation", and can touch and move scene objects. For these purposes, hand controllers are equipped with plenty of sensors to control and track the user interaction and to send them directly to the render engine.

Another important aspect of haptic devices is the force feedback that gives the user the guarantee that a haptic interaction has been noticed and accepted by the system (e.g. in case of pushing a button in the virtual scene). Hand controllers usually give tactile feedback (e.g. vibrations), often combined with an acoustic and/or visual feedback. More sophisticated and highly-specialised haptic devices like the Phantom Premium from 3D Systems allow an extremely accurate force feedback [202]. Other highly specialised haptic devices with integrated force feedback are data gloves (e.g. Avatar VR).

The most challenging situation is force feedback for interaction with hand tracking systems like Leap Motion. Due to the absence of hand controllers, it is limited to acoustic and visual feedback without any tactile information. One solution to overcome this drawback is to use ultrasound. The most renowned company in this field was the company Ultrahaptics, which has now merged with the above mentioned hand-tracking company Leap Motion, with the resulting company now being called Ultraleap [200]. Their systems enable mid-air haptic feedback via an array of ultrasound emitters usually positioned below the user's hand. While the hand is tracked with the integrated Leap-Motion camera module, the ultrasound feedback can be generated at specific 3D positions in mid-air hand position. Ultrahaptics has received \$85.9M total funding. This shows the business value of advanced solutions in the domain of haptic feedback for VR experiences.

Apart from location-based VR entertainment, a crucial limitation of navigating in VR scenes is the limited area in which the user can move around and is being tracked. Therefore, most VR application offer the possibility to jump into new regions of the VR scene by using e.g. hand controllers; as indicated above, this is often referred to as "teleportation". Obviously, teleportation is an unnatural motion, but this is a reasonable trade-off today. However, to give the user a more natural impression of walking around, several companies offer omni-directional treadmills (e.g. Sensorial XR [203], Cyberith Virtualizer [204] or KAT VR [205]).

4.6.5 References

- [185] Cruz-Neira, Carolina and Sandin, Daniel J. and DeFanti, Thomas A., Kenyon, Robert V. and Hart, John C., "The CAVE: Audio Visual Experience Automatic Virtual Environment", *Commun. ACM*, vol. 35, no. 6, pp. 64–72, June 1992.

- [186] S. Manjrekar and S. Sandilya, D. Bhosale and S. Kanchi, A. Pitkar and M. Gondhalekar, "CAVE: An Emerging Immersive Technology - A Review", in *2014 UK Sim-AMSS 16th International Conference on Computer Modelling and Simulation*, 2014.
- [187] H. Urey, K. V. Chellappan, E. Erden, and P. Surman, "State of the Art in Stereoscopic and Autostereoscopic Displays", in *Proc. the IEEE*, vol. 99, no. 4, pp. 540-555, April 2011, doi: 10.1109/JPROC.2010.2098351.
- [188] "Dimenco back in Dutch hands". Bits and Chips., <https://bits-chips.nl/artikel/dimenco-back-in-dutch-hands/> (accessed Nov. 12, 2020).
- [189] "Simulated Reality 3D Display Technology". Dimenco, <https://www.dimenco.eu/simulated-reality> (accessed Nov. 12, 2020).
- [190] K. Hopf, P. Chojecki, F. Neumann, and D. Przewozny, "Novel Autostereoscopic Single-User Displays with User Interaction", in *SPIE Three-dimensional TV, Video, and Display V*, Boston, MA, USA, 2006.
- [191] "TrendForce Global VR Device Shipments Report, 2017-2019." Statista, <https://www.statista.com/statistics/671403/global-virtual-reality-device-shipments-by-vendor/> (accessed Nov. 12, 2020).
- [192] "Analysis: Monthly-connected VR Headsets on Steam Pass 1 Million Milestone." Road to VR, <https://www.roadtovr.com/monthly-connected-vr-headsets-steam-1-million-milestone/> (accessed Nov. 12, 2020).
- [193] "Comparison of virtual reality headsets." Wikipedia, https://en.wikipedia.org/wiki/Comparison_of_virtual_reality_headsets (accessed Nov. 12, 2020).
- [194] Virtual Dimension Center (VDC). Whitepaper, Head Mounted Displays & Data Glasses, Applications and Systems. 2016
- [195] "Magic leap admits they have been leapfrogged by HoloLens 2". MSPoweruser. <https://mspoweruser.com/magic-leap-admits-they-have-been-leapfrogged-by-hololens-2/> (accessed Nov. 12, 2020).
- [196] "Nreal Light MR glasses". Nreal, <https://www.nreal.ai/light> (accessed Nov. 12, 2020).
- [197] "Augmented Reality: Innovations and Lifecycle". Parks Associates, <https://www.parksassociates.com/report/augmented-reality> (accessed Nov. 12, 2020).
- [198] The VOID. <https://www.thevoid.com/> (accessed Nov. 12, 2020).
- [199] Illusion Walk. <https://www.illusion-walk.com/> (accessed Nov. 12, 2020).
- [200] Ultraleap. <https://www.ultraleap.com/> (accessed Nov. 12, 2020).
- [201] Tobii VR. <https://vr.tobii.com/> (accessed Nov. 12, 2020).
- [202] 3D Systems. <https://www.3dsystems.com/scanners-haptics#haptic-devices> (accessed Nov. 12, 2020).
- [203] Sensorial XR. <https://sensorialxr.com/> (accessed Nov. 12, 2020).
- [204] Cyberith. <https://www.cyberith.com/> (accessed Nov. 12, 2020).
- [205] KAT VR. <https://kat-vr.com/> (accessed Nov. 12, 2020).

4.7 Render engines and authoring tools

A key technology required are tools to create AR and VR experiences. The most common platforms for the creation of 3D environments and real-time rendering on a large variety of devices are the following are Unity [206] and Unreal [207]. Both applications offer a 3D development environment in which games, AR and VR applications and other interactive applications can be developed. They also support real-time rendering for all common operating systems such as Linux, Windows and iOS. The MetaVRse engine aims to be a fully web-based design and development tool to create immersive 3D/XR experiences across virtually any OS, browser, or device [208]. InstaVR offers VR application development for all relevant VR cameras (360 degree video) supporting the Oculus family as well as WebVR [209].

Volumetric video will become one of the key technologies in the near future to create realistic digital representations of humans. Due to different representation formats by major volumetric video studios (see sec. 4.1.4 for more details), the modification of volumetric assets is quite a challenge. In order to re-create new combinations of human performances from captured assets, the finish start-up Sens of Space is currently developing an authoring tool to allow arbitrary modifications of volumetric video [210].

In the education sector, the platform Cospaces Edu [211] lets students build 3D creations, animate them and explore them in Virtual or Augmented Reality. In the area of creation 360-degree videos, Google Tour Creator [212] allows people to build immersive 360-degree video or tours. Furthermore, Apelab developed various software tools being integrated in 2019 in the product Zoe giving teachers a simple way to get started in using Spatial Learning in K-12 and Higher Education to make the students learning experiences more engaging and adapted to the future of education [213].

4.7.1 References

[206] Unity. <https://unity.com/> (accessed Nov. 12, 2020).

[207] Unreal Engine. <https://www.unrealengine.com/en-US/> (accessed Nov. 12, 2020).

[208] MetaVRse. <https://metavrse.com/> (accessed Nov. 14, 2020).

[209] InstaVR. <https://www.instavr.co/> (accessed Nov. 22, 2020).

[210] Sense of Space. <https://www.senseofspace.io/>

[211] CoSpaces Edu. <https://cospaces.io/edu/about.html> (accessed Nov. 12, 2020).

[212] Tour Creator. <https://arvr.google.com/tourcreator/> (accessed Nov. 12, 2020).

[213] Apelab. <https://www.apelab.io/> (accessed Nov. 14, 2020).

4.8 Cloud services

The low latency and high bandwidth provided by 5G communication technologies will drive the development of XR technologies by distributing complex computation requiring very low latencies (on the order of a few milliseconds) into the centralised cloud or the edge cloud.

4.8.1 Remote rendering

Remote rendering for very high resolution and high frame rate VR & AR headsets is currently one of the main usage of edge cloud technology for XR [214][215]. Indeed, the 1 to 3 milliseconds of latency induced by the distribution of calculations on the edge cloud is one possibility to preserve a “motion-to-photon”

(M2P) latency under 20ms by significantly reducing the network round-trip time. It is well-known that an increase in M2P latency may cause an unpleasant user experience and motion sickness [216][217]. Therefore, moving the volumetric content to an edge server geographically closer to the user is an important optimisation for improving the user's Quality of Experience (QoE).

High-quality volumetric videos can be represented as meshes that consist of millions of polygons. Rendering such representations in real-time is currently very challenging for mobile devices, whose GPUs are much less capable than desktop/server GPUs. Moreover, unlike 2D video and omnidirectional content that can be decoded using dedicated hardware, decoding of volumetric videos can only be performed in software today, resulting in high computational overhead that can quickly drain the battery of mobile XR devices. Thus, XR remote rendering allow the users to immerse themselves in CAD models of several hundred million polygons using mobile AR or VR devices that can hardly display more than 100,000 polygons in real-time [218].

In the same way that cloud gaming is changing the business model of the video game industry, it may be that cloud VR and AR offerings may expand in the coming years and promote the adoption of XR for mass market. In any case, Huawei is massively relying on 5G and edge cloud technologies applied to XR [219], and could become a leader in the field in the coming years. In Europe, telecommunication operators such as Deutsche Telekom or Orange are preparing this capability [220].

Several companies have recently launched cloud-based XR rendering platforms. NVIDIA CloudXR [221] is built on NVIDIA RTX™ GPUs and provides an SDK that allows streaming of XR experiences. Using NVIDIA GPU virtualisation software, CloudXR targets efficient scaling by allowing multiple users to share GPU resources. Azure™ Remote Rendering [222] is a cloud service by Microsoft that enables rendering high-quality volumetric content in the cloud and stream it to end devices (currently HoloLens 2 and Windows 10 PCs). Target use cases include industrial plant management and design review for assets (such as truck engines) that require visualisation of highly complex 3D models with millions of polygons. Unreal Pixel Streaming [223] is a plugin for Unreal Engine™ (UE) that allows running a packaged UE application on a cloud server. Pre-rendered frames from the UE application can be directly streamed to web browsers using a WebRTC P2P communication framework. Users can interact with the scene on their browsers sending keyboard, mouse or touch events.

On the research side, significant effort has been going into the design and optimisation of remote rendering systems for efficient delivery of XR content. Initial works focused on edge cloud-based rendering of VR content [224], but the attention has been shifting to streaming of volumetric videos for MR use cases [225][226]. In collaboration with Deutsche Telekom, Fraunhofer HHI developed a prototype system for interactive low-latency streaming of animatable volumetric meshes using a 5G edge cloud server [227]. The system uses WebRTC streaming for low-latency network transmission and hardware video encoding to reduce the compression delay. Prediction of the user's 6DoF head motion is another important optimisation that may potentially eliminate a significant portion of the effective M2P latency. However, mispredictions of head motion may potentially degrade the user's QoE. Therefore, recent works started to investigate more accurate and robust prediction techniques based on advanced models [228][229].

4.8.2 AR Cloud

AR is announced as a breakthrough poised to revolutionise our daily lives in the next 5 to 10 years. But to reach the tipping point of real adoption, an AR system will have to run anywhere at any time. Along these lines, many visionaries present AR as the next revolution after smartphones, where the medium will become the world.

Thus, a persistent and real-time digital 3D map of the world, the ARCloud, will become the main software infrastructure in the next decades, far more valuable than any social network or PageRank index [230]. Of course, the creation and real-time updating of this map built, shared, and used by every AR users will only be possible with the emergence of 5G networks and edge computing. This map of the world will be invaluable, and big actors such as Apple, Microsoft, Alibaba, Tencent, but especially Google that already has a map of the world (Google Map), are well-positioned to build it.

The AR cloud raises many questions about privacy, especially when the risk of not having any European players in the loop is significant. Its potential consequences on Europe's leadership in interactive technologies are gargantuan. With that in mind, it is paramount for Europe to immediately invest a significant amount of research, innovation, and development efforts in this respect. In addition, it is necessary now to prepare future regulations that will allow users to benefit from the advantages of ARCloud technology while preserving privacy. In this context, open initiatives such Open ARCloud [231] or the XRSI Privacy Framework [232], as well as standardisation bodies such as the Industry Specification Group "Augmented Reality Framework" at ETSI [233] are already working on specifications and frameworks to ensure ARCloud interoperability.

4.8.3 References

- [214] S. Shi, V. Gupta, M. Hwang, R. Jana, "Mobile VR on edge cloud: a latency-driven design", in *Proc. Of the 10th ACM Multimedia Systems Conference*, pp. 222-231, June 2019.
- [215] "Cloud AR/VR Whitepaper." GSMA. <https://www.gsma.com/futurenetworks/wiki/cloud-ar-vr-whitepaper> (accessed Nov. 12, 2020).
- [216] B. D. Adelstein, T. G. Lee, and S. R. Ellis, "Head tracking latency in virtual environments: psychophysics and a model", in *Proc. the Human Factors and Ergonomics Society Annual Meeting*, Los Angeles, CA, USA: SAGE Publications, vol. 47, no. 20, pp. 2083-2087, 2003.
- [217] R. S. Allison, L. R. Harris, M. Jenkin, U. Jasiobedzka and J. E. Zacher, "Tolerance of temporal delay in virtual environments." in *Proc. IEEE Virtual Reality 2001*, Yokohama, Japan, 2001, pp. 247-254, doi: 10.1109/VR.2001.913793.
- [218] "Holo-Light AR edge computing." HOLO-LIGHT. <https://holo-light.com/pledger-next-level-edge-computing/> (accessed Nov. 12, 2020).
- [219] "Preparing For a Cloud AR/VR Future." Huawei report. https://www-file.huawei.com/-/media/corporate/pdf/x-lab/cloud_vr_ar_white_paper_en.pdf (accessed Nov. 12, 2020).
- [220] "Podcast Terry Schussler (Deutsche Telekom) on the importance of 5G and edge computer for AR." The AR Show. <https://www.thearshow.com/podcast/043-terry-schussler> (accessed Nov. 12, 2020).
- [221] NVIDIA CloudXR. <https://developer.nvidia.com/nvidia-cloudxr-sdk> (accessed Nov. 12, 2020).

- [222] Azure Remote Rendering. <https://azure.microsoft.com/en-us/services/remote-rendering/> (accessed Nov. 12, 2020).
- [223] Unreal Pixel Streaming. <https://docs.unrealengine.com/en-US/Platforms/PixelStreaming> (accessed Nov. 12, 2020).
- [224] S. Mangiante, G. Klas, A. Navon, G. Zhuang, R. Ju, and M. F. Silva, "VR is on the edge: How to deliver 360 videos in mobile networks", in *Proc. of the Workshop on Virtual Reality and Augmented Reality Network*, pp. 30-35, 2017.
- [225] F. Qian, B. Han, J. Pair and V. Gopalakrishnan, "Toward practical volumetric video streaming on commodity smartphones." in *Proc. of the 20th International Workshop on Mobile Computing Systems and Applications*, pp. 135-140, 2019.
- [226] Jeroen van der Hooft, T. Wauters, F. Turck, C. Timmerer, and H. Hellwagner, "Towards 6DoF HTTP adaptive streaming through point cloud compression", in *Proc. of the 27th ACM International Conference on Multimedia*, pp. 2405-2413, 2019.
- [227] S. Gül, D. Podborski, T. Buchholz, T. Schierl, C. Hellge, "Low-latency Cloud-based Volumetric Video Streaming Using Head Motion Prediction", in *Proc. of the 30th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV '20)*, Association for Computing Machinery, Istanbul, Turkey, June 2020.
- [228] X. Hou, J. Zhang, M. Budagavi and S. Dey, "Head and Body Motion Prediction to Enable Mobile VR Experiences with Low Latency", in *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, 2019, pp. 1-7).
- [229] S. Gül, S. Bosse, D. Podborski, T. Schierl, C. Hellge, "Kalman Filter-based Head Motion Prediction for Cloud-based Mixed Reality", In *Proc. of the 28th ACM International Conference on Multimedia (ACMMM)*, Oct. 2020.
- [230] Charlie Fink's, "Metaverse. An AR Enabled Guide to VR & AR", 2018.
- [231] Open AR cloud. <https://www.openarcloud.org/> (accessed Nov. 12, 2020).
- [232] XRSI. <https://xrsi.org/publication/the-xrsi-privacy-framework> (accessed Nov. 12, 2020).
- [233] ETSI. <https://www.etsi.org/committee/arf> (accessed Nov. 12, 2020).

4.9 Conclusion

The presented technology domains are considered as the most relevant ones for interactive XR technologies and applications. Since the first version of this document in November 2019, many more aspects have been added in order to cover the whole spectrum of XR technologies.

5 XR applications

In this section, the most relevant domains and the most recent developments for XR applications are discussed in some detail. The above domains were selected, based on (1) the market watch presented in sec. 3.2 and (2) the main players in the AR & VR industry in sec. 3.5.

5.1 Advertising and commerce

AR has reached already a quite mature level in several areas especially for home furnishing. Some offered functionalities are measuring a room, and placing & scaling objects such as furniture. A list of example applications is given here [234][235], such as IKEA Place, Houzz, Augmented Furniture, Wayfair, Homestyler amongst others. All these applications use the integrated registration and tracking technology that is available for iOS devices (ARKit) and Android devices (ARCore). Two examples are shown in Figure 21.

In the real-estate domain, AR is used to prove users with an experience using 3D models while they are looking for properties to rent or buy. The users get an instant feel of how the property of interest is going to look like, whether the property already exists, or must still be built or completed. The benefits include eliminating or reducing travel time to visit properties, virtually visiting a larger number of properties, providing a personal experience, testing furniture, and likely signing a purchase-and-sale agreement faster. For sellers and real-estate agents, the main benefits are less time on the road and faster purchase decisions. See the following references: [236][237][238] (see the Onirix App in Figure 22, left).

In the food & beverage industry, AR is used to allow users to preview their potential order like in Jarit [239] (see the Jarit App in Figure 22, right).

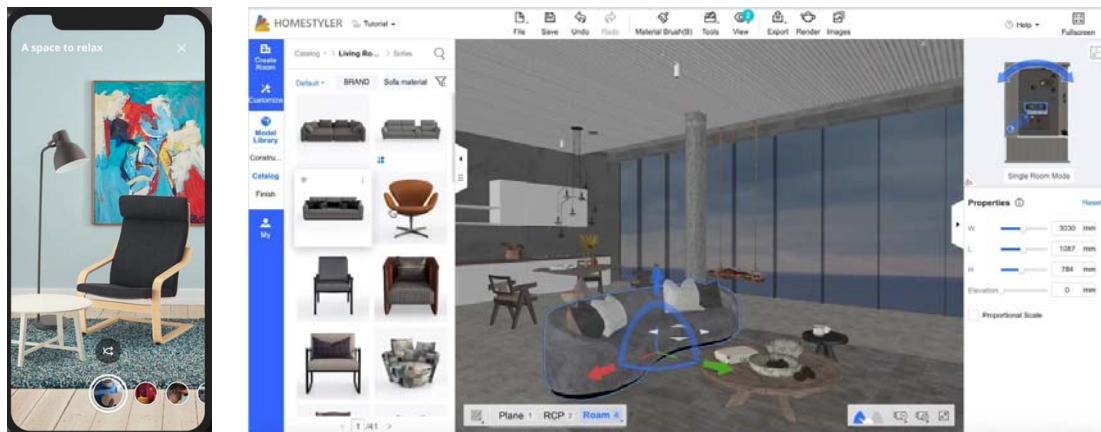


Figure 21: AR applications in furniture domain: IKEA app (left), Homestyler (right).

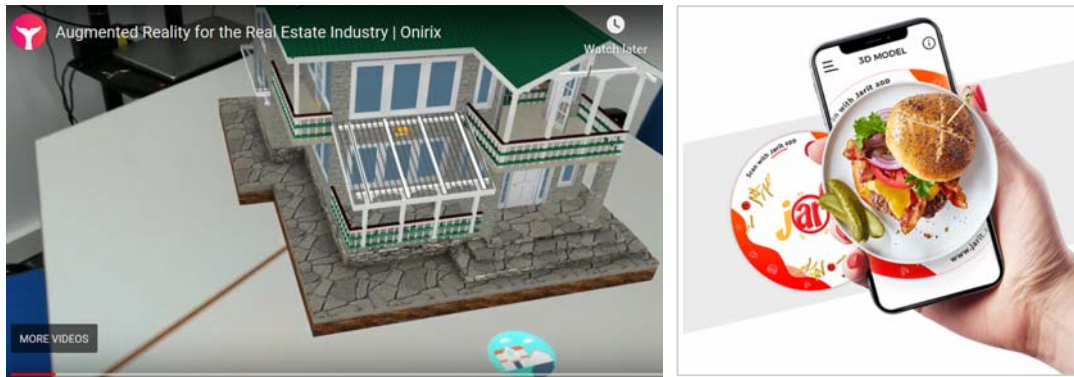


Figure 22: AR applications in retail and food domain: real estate app by Onirix (left), food preview app by Jarit (right).

In the fashion industry, AR and VR becomes a relevant technology for various applications. The main objective is to bridge off-line experience and on-line buying experience. Several platforms are available addressing the fashion market such as Obsess, and Virtusize.

Modiface, acquired by L’Oreal in March 2018, is an AR application that allows one to simulate live 3D make-up. The company ZREALITY developed a virtual show room, where designers and creators can observe fashion collections anywhere and anytime [240]. Different styles can be combined and jointly discussed. Clothing can be presented in a photo-realistic way (see Figure 23).



Figure 23: Sample views of ZREALITY virtual show room.

The glasses retailer Warby Parker recently presented an online-try-on Augmented Reality app to allow the user to try different models of glasses [241]. Based on face scanning capabilities of Apple’s iPhone X, the users receive personalised product suggestions from the App.

Cosmetic company Sephora uses AR technology to allow customers to try out different looks and eye, lips and cheek products as well as colours right on their own digital face [242]. This is a powerful way to boost sales and to give customers a fun way to try out new looks. Another company that uses augmented reality to inspire purchases is Chrono24 with its AR app Virtual Showroom [243]. The company has developed a

virtual try-on experience where prospective customers can try out different styles and models. In Figure 24, the eco-system for the fashion industry is depicted. The major players for the development of AR & VR applications are listed.



Figure 24: The fashion eco-system.

5.1.1 References

- [234] "14 best Augmented Reality furniture apps". Nadia Kovach. <https://thinkmobiles.com/blog/best-ar-furniture-apps/> (accessed Nov. 14, 2020).
- [235] "Augmented Reality in Furniture". Nadia Kovach. <https://thinkmobiles.com/blog/augmented-reality-furniture/> (accessed Nov. 30, 2020).
- [236] Onirix. <https://www.onirix.com/learn-about-ar/augmented-reality-in-real-estate/> (accessed Nov. 12, 2020).
- [237] Obsess. <https://www.obsessar.com/> (accessed Nov. 12, 2020).
- [238] Virtusize. <https://www.virtusize.com/site/> (accessed Nov. 12, 2020).
- [239] Jarit. <https://jarit.app> (accessed Nov. 12, 2020).
- [240] ZREALITY. <https://www.zreality.com/vr-mode/> (accessed Nov. 12, 2020).
- [241] Warby Parker. <https://www.warbyparker.com/app> (accessed Nov. 12, 2020).
- [242] Sephora. <https://www.sephora.sg/pages/virtual-artist> (accessed Nov. 14, 2020).
- [243] Chrono24. <https://www.chrono24.com/info/apps.htm#augmented-reality> (accessed Nov. 12, 2020).

5.2 Cultural Heritage

Cultural heritage has always been an important aspect of human society and technological advances are often used in order to preserve, protect and make accessible cultural heritage to the general audience. In the last years, research has been made in developing innovative systems that focus on cultural heritage. Europe has already made some steps in expanding the research agenda to include cultural heritage. For example, the European Project eHERITAGE [244] had as a goal to develop a center of excellence in virtual heritage by exploiting recent advancements in the field of virtual reality and intelligent systems. In [245], Carrozino et al. carried out a comparative study on innovative XR systems in cultural heritage during the H2020 project eHERITAGE. In Figure 25, Figure 26, Figure 27 and Figure 28, we show the four systems that were developed and evaluated:



Figure 25: Mobile AR exposition.

User looking at the painting “Still Nature” by Romanian painter S. Luchian. #



Figure 26: VR Book: Researchers’ Night participants using the digital book.

The book was drafted in Vienna in 1768 and ensured the uniform application of Criminal Law in Austria and Bohemia.



Figure 27: Holographic Stand: participants looking at the holographic display.



Figure 28: Bow Simulator: Students trying the haptic bow.

The different systems were compared at application level and classified based on common features such as Interaction, Manipulability, Ease of Use and others. The interaction level for example differs from

application to application. Looking at some visual content as shown in Figure 25 and Figure 26 is less interactive than using material in some visual context as shown in Figure 27 and Figure 28.

European research institutes such as Fraunhofer are also contributing to the innovation of cultural heritage systems. The Omnicam-360 and the 3D Human Body Reconstruction technology of Fraunhofer HHI were used to permanently digitise and research worldwide cultural objects and artefacts in the “Cultural Heritage Expo” [246]. In this way, art and cultural objects can be accessed at anytime from anywhere. Additionally, CultLab3D, developed by Fraunhofer IGD, specialises in 3D scanning technologies. It focuses on offering an autonomous 3D scanning pipeline for fast and economic mass digitisation [247]. One of the main applications is that of 3D digitisation of cultural heritage artefacts (see Figure 29).



Figure 29: 3D digitisation of cultural heritage artefacts as developed by CultLab3D.

A number of museums have been offering either VR experiences [248][249][250][251][252][253] or Mixed Reality [254] experiences to their audience in addition to their exhibition. Even though many VR products were developed the last few years in the content of a museum, there is still space to explore and define how the digital products enhancing the museum visit should be. The research project museum4punkt0 [255] connects seven cultural institutions from different regions in Germany and test digital products for new types of learning, experiencing, and participation in museums.

There is some progress made in the tourism area as well. The Luxembourgish company URBAN TIMETRAVEL created a virtual reality bus tour, which was to be presented at ITB 2020, where the tourists can experience the city of Luxembourg in 1867 [256]. The system makes use of real-time location and mixed reality technology in order to provide the tourist with an immersive cultural experience (see Figure 30).



Figure 30: Luxembourg in 1867 experience as developed by URBAN TIMETRAVEL.

A few years ago, Google launched a browsing application called “Google Arts & Culture” [257] with which you could virtually visit many museums all over the world. It also supported Google Cardboard DIY VR headset to take 360-degree tours of some of the featured museums, heritage sites and landmarks. There are many partners [258] to Google Arts & Culture among others *British Museum* in London, *Van Gogh Museum*, *Musée d’Orsay*, *Acropolis Museum* and *Pergamon Museum*.

Finally, mixed reality technology cannot only be used to enhance existing art but also to create art itself. Joseph Bates on his paper “*Virtual Reality, Art, and Entertainment*” [259] in 1992 mentioned that the public is beginning to understand that virtual reality portends a new medium, new entertainment, a new and very powerful type of art. After almost two decades later, the virtual reality field became mature enough in order for the artists to start using it. Famous artist like Olafur Eliasson [260] start using augmented reality to create art with.

5.2.1 References

- [244] eHeritage. <http://www.eheritage.org/>
- [245] M. Carrozzino, G. Voinea, M. Duguleana, R. Boboc and M. Bergamasco, “Comparing innovative XR systems in culture heritage. A case study”, *ISPRS - International Archives of the Photogrammetry. Remote Sensing and Spatial Information Sciences*, pp. 373-378. doi: 10.5194/isprs-archives-XLII-2-W11-373-2019.
- [246] Fraunhofer HHI. <https://www.hhi.fraunhofer.de/en/press-media/news/2018/fraunhofer-hhi-technologies-at-the-cultural-heritage-expo.html> (accessed Nov. 12, 2020).
- [247] CultLab3D. <https://www.cultlab3d.de/> (accessed Nov. 12, 2020).
- [248] Louvre Museum. <https://www.louvre.fr/en/leonardo-da-vinci-0/realite-virtuelle> (accessed Nov. 12, 2020).
- [249] National Museum of Finland. <https://www.helsinki.fi/en/national-museum-of-finland-virtual-reality> (accessed Nov. 12, 2020).

- [250] National Museum of Natural History. <https://naturalhistory.si.edu/visit/virtual-tour> (accessed Nov. 12, 2020).
- [251] French National Museum of Natural History. <https://www.mnhn.fr/en/explore/virtual-reality/journey-into-the-heart-of-evolution> (accessed Nov. 12, 2020).
- [252] The Natural History Museum. <https://www.nhm.ac.uk/discover/news/2018/march/explore-the-museum-with-sir-david-attenborough.html> (accessed Nov. 12, 2020).
- [253] staedel museum. <https://www.staedelmuseum.de/en/offerings/time-machine> (accessed Nov. 12, 2020).
- [254] VR Focus. <https://www.vrfocus.com/2018/01/petersen-automotive-museum-showcases-mixed-reality-exhibit/> (accessed Nov. 12, 2020).
- [255] Museum4punkt0. <https://www.museum4punkt0.de/en/> (accessed Nov. 12, 2020).
- [256] Urban Timetravel. <https://www.urbantimetravel.com/> (accessed Nov. 12, 2020).
- [257] Google Arts&Culture. <https://about.artsandculture.google.com/> (accessed Nov. 12, 2020).
- [258] Google Arts&Culture. <https://artsandculture.google.com/partner> (accessed Nov. 12, 2020).
- [259] Joseph Bates, *Virtual Reality, Art, and Entertainment*, MIT Press, pp. 133-138, 1992.
- [260] Acute Art. <https://app.acuteart.com/> (accessed Nov. 12, 2020).

5.3 Education and Research

The potential of extended reality (XR) within the field of education and research is manifesting itself as a powerful multifunctional toolkit, used for the dissemination of knowledge and the interactive participation in educational and research contexts.

While the medium has not found its way into mainstream academia yet, it appears as a highly promising and diverse tool that could improve education and research on multiple levels. The following sections will delineate different use cases of extended reality technologies, their impact and benefits, as well as the gaps that still need to be bridged.

5.3.1 XR as teaching medium

The first approach is to embed XR experiences in the curriculum and apply them as a teaching medium. Just like watching a documentary, doing observatory field work or reading a book is part of the educational program, the embedding of a specific VR or AR experience can be part of a curriculum.

VR and AR can be a fun and engaging way to bring educational content to students with a clear didactic, experimental or presentational goal in mind. It can be a powerful supportive tool alongside traditional teaching methods or function as a stand-alone module. Examples of applications in academia range from stepping inside unique worlds for field trips and excursions [261][262][263], learning about abstract concepts and processes [261][264] or training specific skills in a safe environment [265].

5.3.1.1 Virtual Reality

Virtual reality enables students to go to places and practise in contexts that are not easily accessible in real life, because they might be too costly or dangerous. It enables teachers to provide contextual learning to students and connect educational content to experience, for example through virtual trips to remote

locations, boosting empathising with other cultures [266][267]. Another example can be found in technical and practical skills training which simulate dangerous environments, or closed-off environments (e.g. training for firefighters, [268] or the practice of medical procedures [269]).

5.3.1.2 Augmented Reality

Head-mounted displays for virtual reality provide an immersive audio-visual space for education and research purposes, and additionally remove noise and interruptive signals from the external world, allowing users to experience endless possibilities of events, goals and contexts. Augmented reality techniques on the other hand are more relevant for learning in a physical context. It can be used to bring virtual content into the classroom. Some of the examples of application include the study of virtual archaeological objects [270] or the exploration of a virtual anatomical model of the body [271] (see Figure 31). As AR allows for a feeling of presence in the real world, users can still communicate naturally through speech and body language, which allows for collaborative learning when manipulating virtual objects [272].



Figure 31: AR can help medical students understand complex anatomy.

Another application of AR lies in the opportunity for supporting practical education remotely, as was done at the Imperial College London, where chemical engineering students could take part in lab-based experiences remotely through augmented reality [273].

5.3.2 Teaching students to autonomously build Experiences by developing media literacy and creation skills

Aside from experiencing content in AR and VR as part of the curriculum, XR technologies also enable students to create their own XR experiences. Educational institutions are adopting this approach ever more frequently, particularly in technical courses involving programming and interaction design [274]. By having students actively work with the medium, students' media literacy can be developed (see Figure 32). Motivating students to review strengths and weaknesses of the medium helps them to form a conceptual and critical understanding of the impact of the medium in general and to understand the relevance of the medium for specific objectives they might want to reach during their studies or future careers [275].



Figure 32: Students increase their media literacy by creating their own VR experiences.

5.3.3 Possibilities for learning and teaching

As can be seen from XR's framework of application, it becomes evident that the medium can support the teaching of both applied and practical knowledge. This gives educators the chance to apply authentic assessment principles when testing students but at the same time, helps students to anchor their knowledge, and prove their critical thinking and problem-solving skills in meaningful situations [276].

To explain how exactly this can be done, the following section will provide an outline of the use cases of XR in the field of higher education and how it offers unique learning opportunities:

- **Applying abstract theory to real-life situations.** Due to XR's potential to visualise or make real-life situations experienceable, it could, for instance, become an asset for visualising molecules or simulating psychological theory in a simulation using digital actors;
- **Allowing experimentation and creativity.** XR offers a space, where students can engage in and create new knowledge. This can be done in novel virtual spaces without real-life repercussions, thereby fostering free exploration and experimentation inside of a safe environment;
- **Serving as an active learning medium or as a supportive tool.** XR technologies offers flexible ways in which it can be integrated into the curriculum. Through elaborate simulated environments, it can serve as a means to train learners through scenarios, contextualised problems or collaborative work.

On the other hand, XR technologies can also be used to support traditional teaching methods by introducing practical aspects, such as virtual field trips or interactive visualisations, before or after explaining theoretical information (see Figure 33);



Figure 33: Cellverse uses VR in a blended learning setting to teach students about cell functions.

- **Simulate environments that were traditionally inaccessible.** In situations where a novel space or tool can be recreated in XR to prepare or teach about the real-world equivalent. XR can enrich existing curriculums when real life alternatives for learning experiences are too dangerous, difficult, expensive or plainly impossible to implement (for example, an excursion to Mars, an opportunity to control a Mars rover or learning to get in and out of a spacesuit as shown in Figure 34);
- **Developing media literacy or design and technical skills.** XR can be used as a target medium in course assignments by supporting students to design and build XR experiences themselves.

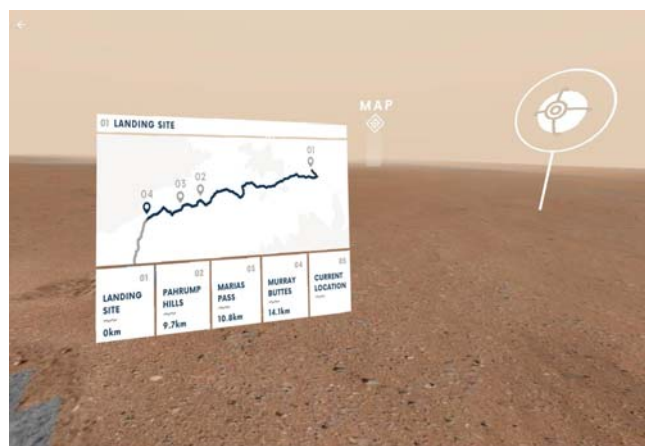


Figure 34: Exploring Mars: VR can give access to normally unreachable places.

5.3.4 Approach for realisation

Academic institutes and individual teachers take different approaches to the implementation of XR in their educational curricula. Some decide to explore freely available or low-cost applications, others acquire more advanced applications offered by commercial companies.

Academic institutes also create their own XR applications, specifically targeted at the needs of their curricula. Often, this process is costlier and there appears to be a lack of content sharing between institutes, preventing applications from being developed further or from becoming accessible for a wider target audience [277].

When the objective is to let students build experiences themselves, teachers can choose to introduce students to different development methods, depending on the desired learning outcomes. These could include more professional development engines like Unity and Unreal or more low-key tools, targeted at less technologically proficient users such as CoSpaces for computer generated virtual and augmented reality experiences and Google Tour Creator for interactive 360° videos.

5.3.5 Effect on students and learning outcomes

While the use of XR technologies in education is still fairly new, there are already many promising studies about the effect on students' learning outcomes.

Apart from giving students access to expensive facilities that are not available in all schools (e.g. labs), virtual learning environments have also proven to result in a significantly positive impact on students' enjoyment and their intrinsic motivation [278]. VR can engage students through multiple sensory stimuli, thus increasing cognitive stimulation. The use of VR for cognitive rehabilitation already demonstrates cognitive improvements in people with mild cognitive impairment or the elderly [279][280]. With regard to education, there is a growing consensus in research that the use of advanced 3D visualisations can enhance the learning experience of students [281]. In Figure 35, left, a virtual chemistry lab is shown [282] and in the right, an AR use case is presented.

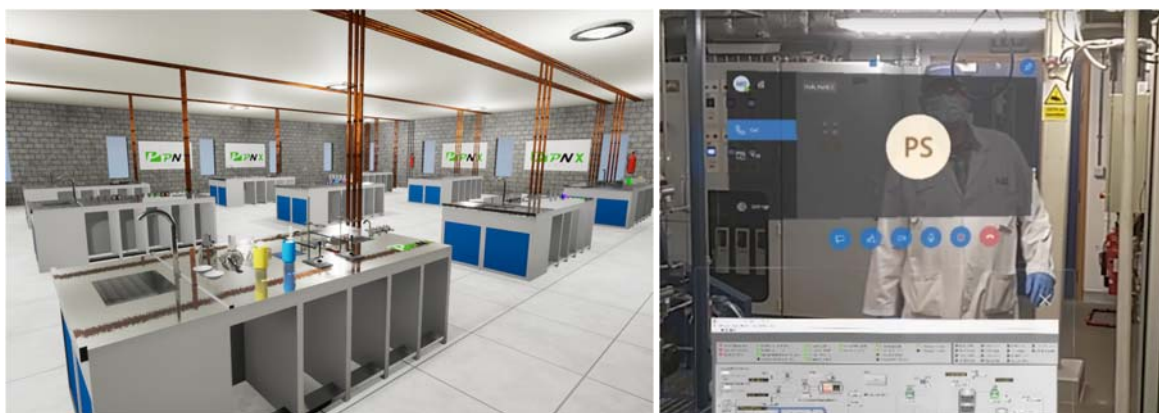


Figure 35: A virtual chemistry lab (left) and teaching practical lab skills in AR (right).

Other factors that increase motivation in students are goal-oriented and collaborative learning [283]. While these aspects can also be covered in traditional learning settings, XR simulations can more easily

incorporate elaborate and diverse scenarios that are more life-like and also allow students to collaborate remotely. Active learning as opposed to passive learning has shown to have a positive impact on students' memory, including in the use of VR [284]. Virtual reality interactions can therefore lead to improved memory retention, especially for learning tasks that involve spatial or navigational information [285][286]. However, XR can also be used in the classroom to improve critical thinking abilities [287]. XR's interactive aspect lets students engage with the objects while constructing their own understanding of concepts. Such an approach to learning can increase the understanding of, for instance, mathematical concepts, especially in lower-performing students [261].

Virtual scenarios can simulate problems in a safe environment, where students can learn from their mistakes without causing harm or experiencing embarrassment. Virtual reality is already known to be successful in reducing anxiety in social settings and could therefore be used to prepare students for real-life interactions [288]. Students can largely benefit from the stress and anxiety reducing effects of immersive virtual experiences in order to focus more on their studies [289].

Given the cost of XR, not all educational institutions can afford to implement these, and more thorough research about long-term effects of VR is needed. What also needs to be considered is that the use of XR is still limited to very few students, therefore requiring larger studies. Future research also needs to compare traditional learning methods with XR, more specifically on whether there is a significant improvement in student performance and how, based on these studies, XR needs to be adapted to different classrooms [290].

5.3.6 XR in research

When considering XR in the area of scientific research it's important to distinguish between different motivations that researchers can have. This is important to ensure clarity in discussing the potential of the medium for research, for example when drafting the research agenda for funding programmes. Drafting from the writer's own experiences in research and educational projects involving XR (the Centre for Innovation at Leiden University), we distinguish between research into XR and using XR as a research method.

5.3.6.1 Research into XR

Research into XR involves conducting studies about the potential of the medium but is also aimed at technical developments of the medium. Qualities of the medium for example relate to researching the potential of the medium for different application areas and various contexts, analysing human cognition in AR and VR experiences, or the impact on society and ergonomics. It involves a wide range of fields such as engineering, psychology, medicine, educational science and humanities. Knowledge acquired about these topics, helps to understand the potential as well as the limitations of the medium, to define a more responsible use of the medium, concerning for example ethics and privacy. Research into specific technologies, such as screen technology and artificial intelligence could contribute to the technological development of the medium.

5.3.6.2 XR as research method

Another way XR plays a role in research, is the possibility to use augmented and virtual environments in experimental setups involving (human) participants. The endless possibilities that underlie creating virtual

environments enable researchers to expose participants to environments and interactions that are otherwise too dangerous, unethical or too expensive to realise. It becomes possible to put people into customised environments and expose them to specific emotions, cultural events etc. Next, it can serve as an alternative to simulate otherwise hard-to-produce situations.

Despite the perception that virtual environments are a safe and convenient alternative that protects participants from real-life consequences and harm, it is crucial to remember that virtual environments can also have adverse consequences for research participants. Although simulations can protect participants from physical harm, researchers still need to respect the rights and dignity of the research participants from a psychological perspective. Given the high impact that virtual environments can have on its users, it is important to expand research ethics in order to establish guidelines that can ensure the proper and safe development and application of virtual simulations.

5.3.7 Facilitating the use and creation of XR in education and research

Researchers, teachers, students and education facilitators now cross financial barriers and are implementing VR and AR as valuable research support, education facilitation and creation mediums within aforementioned contexts. Looking at examples mentioned above, it seems that application areas are explored thoroughly and a more varying range of hardware is being used. Nonetheless, continuous and sustainable implementation of XR in research and education still encompasses some challenges.

First of all, those challenges are related to practical conditions needed to use the medium in the first place. While educational institutes usually have a special department dedicated to IT facilitation, often matters, such as hardware acquisition, device management and user support have not yet found a place in formal structures. Also, in cases where IT departments do take responsibility over these issues, they are confronted with a large set of new questions, for instance: *Who should have access to device?, How should charging be handled? How should training and support be offered to new users?* While hygiene has previously been the elephant in the room for organisations working with shared devices, the intimate nature of the medium confronts facilitators with the question of how to ensure hygiene safety if devices are not owned by students or teachers themselves.

Next to practical facilitation, a knowledge gap on the responsible use of the medium and its applications is rising. As the medium is slowly surpassing the stage of novelty, many new groups of teachers and educational facilitators are starting to implement the medium in their curricula. While this could be considered a good thing, it is important that institutes and governments are aware of the potential negative effects of the medium and draft frameworks and regulations that take into account components like ethics, privacy and health.

To accommodate some of these challenges, institutions can adopt various approaches. The introduction of XR technologies can first happen on smaller scales and specific use cases, as has been done in the past. This allows educators and researchers to investigate the effects of the XR and understand where improvements need to be made before the XR technologies can be used on a larger scale. Another alternative would also be to use VR as a proxy [291]. This method would allow instructors to demonstrate theoretical knowledge in virtual environments, while students are able to observe. Such an approach would require fewer resources, less training and would also make it easier to guarantee the safe and

ethical use of the medium. Students will be able to become more familiar with the medium, and the integration of XR could happen in a controlled manner.

Educators and students should be actively involved in the process of introducing XR into the curriculum. This can happen on multiple levels, including the design process of educational XR environments, feedback sessions on the current state of education as opposed to the desired outcomes that XR could bring to the classroom, or participation in the development of XR applications.

5.3.8 Conclusion

As is the case with other fields in which XR is being introduced and implemented, we can observe evidence of XR leading to significant improvements in both education and academia. Current research indicates that XR technologies can lead to significant improvements in student engagement, motivation and performance. XR offers clear benefits to both students and educators in terms of achieving learning outcomes and assessment. We also observe that research into XR and, using XR as a research method is a relatively new phenomenon, yet expect this area to grow proportionately once the education and XR field gains traction.

Yet, we are also aware of the still novel status that XR occupies in the field of education and academia which requires more studies to be conducted in order to dive deeper into longer-term learning benefits, privacy issues, ethical problems and inclusivity of XR. Gathered insights will prove important to gauge the effectiveness of XR and the need for this might become more evident if XR moves from being a relatively fringe technology to a more widely implemented technology in both academia and education.

5.3.9 References

- [261] E. Hu-Au and J.J. Lee, "Virtual reality in education: a tool for learning in the experience age", *International Journal of Innovation in Education*, vol. 4, no. 4, pp. 215-226, 2017
- [262] D. Schipper. "Fieldwork Techniques: a virtual excavation. Centre for Innovation." Centre for Innovation. <https://www.centre4innovation.org/stories/fieldwork-techniques-a-virtual-excavation/> (accessed Nov. 12, 2020).
- [263] E. Evans. "Virtual museum and monument tours: how to explore the wonders of history from your home." HistoryExtra. <https://www.historyextra.com/magazine/virtual-remote-museum-exhibition-tours-how-explore-history-from-home/> (accessed Nov. 12, 2020).
- [264] S. W. Greenwald. "Electrostatic Playground: A multi-user virtual reality physics learning experience." MIT Media Lab. <https://www.media.mit.edu/projects/vr-physics-lab/overview/> (accessed Nov. 12, 2020).
- [265] EON Reality. "A Virtual Lab for Chemistry Students." EON Reality. <https://eonreality.com/a-virtual-lab-for-chemistry-students/> (accessed Nov. 12, 2020).
- [266] UN VIRTUAL REALITY. "Syrian Refugee Crisis – UN Virtual Reality." United Nations Virtual Reality (UNVR). <http://unvr.sdgactioncampaign.org/cloudoversidra/#.X5FZF5LitPZ> (accessed Nov. 12, 2020).
- [267] "Under the Canopy - A VR Experience." Conversation International. <https://www.conversation.org/stories/virtual-reality/amazon-under-the-canopy> (accessed Nov. 12, 2020).

- [268] R.M. Clifford, H. Khan, S. Hoermann, M. Billingham, R.W. Lindeman, “Development of a multi-sensory virtual reality training simulator for airborne firefighters supervising aerial wildfire suppression”, in *2018 IEEE Workshop on Augmented and Virtual Realities for Good (VAR4Good)*, pp. 1-5, 2018.
- [269] H.G. Colt, S.W. Crawford, O. Galbraith III, “Virtual reality bronchoscopy simulation: a revolution in procedural training”, *Chest*, vol. 120, no. 4, pp. 1333-1339, 2001.
- [270] B.J. Fernández-Palacios, A. Rizzi, F. Nex, “Augmented reality for archaeological finds”, in *Euro-Mediterranean Conference*, Springer Berlin Heidelberg, 2012, pp. 181-190.
- [271] J. Kroese, and L.U.M.C. “Seeing clearly: How augmented reality can help medical students understand complex anatomy.” Centre for Innovation.
<https://www.centre4innovation.org/stories/augmented-reality-app-leiden-medical-students-transplants/> (accessed Nov. 12, 2020).
- [272] J. Martín-Gutiérrez, P. Fabiani, W. Benesova, M.D. Meneses, C.E. Mora, “Augmented reality to promote collaborative and autonomous learning in higher education”, *Computers in human behavior*, vol. 51, pp. 752-761, 2015.
- [273] M. MacKay. “Lab-based teaching re-imagined using augmented reality.” Imperial News.
https://www.imperial.ac.uk/news/202013/lab-based-teaching-re-imagined-using-augmented-reality/?mc_cid=e1ad0a431b&mc_eid=%5BUNIQID%5D (accessed Nov. 12, 2020).
- [274] “Student- Created VR Experiences – It is Easier Than You Think!.” The Infused Classroom.
<https://www.hollyclark.org/2019/10/30/student-created-vr-experiences-it-is-easier-than-you-think/>, (accessed Nov. 12, 2020).
- [275] R. Hobbs, K. Donnelly, J. Friesem, M. Moen, “Learning to engage: How positive attitudes about the news, media literacy, and video production contribute to adolescent civic engagement”, *Educational Media International*, vol. 50, no. 4, pp. 231-246, 2013.
- [276] V. Villarroel, D. Boud, S. Bloxham, D. Bruna, C. Bruna, “Using principles of authentic assessment to redesign written examinations and tests”, *Innovations in Education and Teaching International*, vol. 57, no. 1, pp. 38-49, 2020.
- [277] T. Ginn. “XR ERA - Extended Reality for Education and Research in Academia.” XR ERA.
<https://xrera.eu/state-of-xr/> (accessed Nov. 12, 2020).
- [278] G. Makransky, S. Borre-Gude, R. Mayer, “Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments”, *Journal of Computer Assisted Learning*, vol. 35, no. 6, pp.691-707, 2019.
- [279] I. Tarnanas, A. Tsolakis, M. Tsolaki, “Assessing virtual reality environments as cognitive stimulation method for patients with MCI”, in *Technologies of Inclusive Well-Being*, Springer Berlin Heidelberg, pp. 39-74, 2014.
- [280] P. Gamito, J. Oliveira, C. Alves, N. Santos, C. Coelho, R. Brito, “Virtual Reality-Based Cognitive Stimulation to Improve Cognitive Functioning in Community Elderly: A Controlled Study”, *Cyberpsychology, Behavior, and Social Networking*, vol. 23, no. 3, pp.150-156, 2020.

- [281] G. Keenaghan, I. Horvath, "Using Game Engine Technologies for Increasing Cognitive Stimulation and Perceptive Immersion", *Smart Technology Based Education and Training 2014*, Crete, Greece, vol. 262, 2014.
- [282] "Chemistry Virtual Labs". PNX. <http://pnxlabs.com/university-labs/chemistry-lab.html> (accessed Nov. 30, 2020).
- [283] H. D. Song, B-L. Grabowski, "Stimulating intrinsic motivation for problem solving using goal-oriented contexts and peer group composition", *Educational Technology Research and Development*, vol. 54, no. 5, pp. 445-466, 2006.
- [284] H. Sauzéon et al., "The Use of Virtual Reality for Episodic Memory Assessment", *Experimental Psychology*, vol. 59, no. 2, pp.99-108, 2012.
- [285] G. Plancher et al., "The influence of action on episodic memory: A virtual reality study", *Quarterly Journal of Experimental Psychology*, vol. 66, no. 5, pp.895-909, 2013.
- [286] K.Z. Huang, C. Ball, J. Francis, R. Ratan, J. Boumis, J. Fordham, "Augmented versus virtual reality in education: an exploratory study examining science knowledge retention when using augmented reality/virtual reality mobile applications", *Cyberpsychology, Behavior, and Social Networking*, vol. 22, no. 2, pp.105-110, 2019.
- [287] J. Ikhsan, K. Sugiyarto, T. Astuti, "Fostering Student's Critical Thinking through a Virtual Reality Laboratory", *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 14, no. 08, pp. 183, 2020.
- [288] D. R. Camara, R.E. Hicks, "Using virtual reality to reduce state anxiety and stress in University students: An experiment", *GSTF Journal of Psychology (JPsych)*, vol. 4, no. 2, 2020.
- [289] R.K. Chesham, J.M. Malouff, N.S. Schutte, "Meta-analysis of the efficacy of virtual reality exposure therapy for social anxiety", *Behaviour Change*, vol. 35, no. 3, pp. 152-166, 2018.
- [290] J.K. Crosier, S.V. Cobb, J.R. Wilson, "Experimental comparison of virtual reality with traditional teaching methods for teaching radioactivity", *Education and Information Technologies*, vol. 5, no. 4, pp. 329-343, 2000.
- [291] N.M. McDonnell, "VR By Proxy – Media and Learning." Media&Learning. <https://media-and-learning.eu/type/featured-articles/vr-by-proxy/> (accessed Nov. 12, 2020).

5.4 Industry 4.0

Starting already in the mid-80s, the professional world had already identified a set of possible uses for AR that range from product design to the training of various operators. But in the last few years, with the arrival of smartphones equipped with advanced sensors (especially 3D sensors) and more powerful computing capabilities, and with the arrival of powerful AR headsets (such as the HoloLens from Microsoft), a considerable number of proof-of-concepts have been developed, demonstrating indisputable returns on investment, in particular through gains in productivity and product quality. Furthermore, one is now beginning to see more and more large-scale deployments in industry. A revolution on industry, also called Industry 4.0, is happening which will radically change the way products are made and managed. Technologies in the scope of Industry 4.0 include among others smart factories, smart products, connected products, digital twin, robotics, virtual reality (VR) and augmented reality (AR) [292] as depicted in Figure 36. In this section, we will attempt to identify and characterise the main uses for AR in Industry 4.0 and construction.



Figure 36: Industry 4.0.

5.4.1 Assembly

Whether it concerns a full task schedule, a sheet of assembly instructions, an assembly diagram, or a manufacturer's manual, the use of AR makes it possible to present an operator with information about the tasks to be done in an intuitive way and with little ambiguity (e.g. matching contextual information with the system being assembled is easier and more natural than using a paper plan). In this context, the augmentations are generally presented sequentially so as to reflect the various steps of the assembly (or disassembly) process as demonstrated in Figure 37.

The information and augmentations integrated into the real world may relate to:

- Number and name of the current step;
- Technical instructions on the task to be performed and the tools or resources to use;
- Safety instructions (at risk areas, personal protective equipment (PPE) to use, etc.);
- Showing, what elements are to be assembled;
- Precise location of the elements to be assembled;
- Path to follow in order to assemble a component;
- Physical action to perform (inserting, turning etc.).



Figure 37: AR application for assembly in industrial environment.

These operator assistance solutions are particularly attractive for businesses with either one of the following characteristics:

- High levels of turnover or seasonal employment, because they reduce the time needed to train the operators;
- Few repetitive tasks, where instructions are never the same and can be continuously displayed in the field of view of the operator.

One benefit of such AR-based assistance solutions lies in the ability to track and record the progress of assembly operations. Their use may therefore help the traceability of operations.

In a context of assembly assistance, it is generally desirable to implement hands-free display solutions. For this reason, the solution deployed is generally one of the following:

- A fixed screen that is placed near the workstation, and that is combined with a camera positioned so as to provide an optimal viewpoint of the objects to be assembled that is understandable, sufficiently comprehensive, and not obscured by the user's hands;
- A head-mounted device (goggles);
- A projection system for use cases, where ambient light is not a problem.

One of the major challenges of these applications lies in the goal of accurately placing the augmentations with respect to the current assembly. Depending on the usage context, the desired accuracy may vary from about a millimetre to several centimetres. For this reason, the methods and algorithms used for locating the AR device in the 3D space also vary, but all include a method that makes it possible to spatially register the digital models (augmentations) with the real world (object to be assembled).

The main issues and technological obstacles for this type of use case particularly relate to improving the process of creating task schedules in AR, and locating moving objects. This is due to the fact that many of these industrial processes are already digitised, but require new automation tools to correctly adapt them to AR interfaces.

Finally, with respect to the acceptability by users of the above AR devices, the needs expressed relate to improving the ergonomics and usability of the display devices, and especially ensuring that the optical devices cause no harm to the operator as a result of intensive use over a long term.

The expected benefits of the use of AR technologies for assembly tasks include improving the quality of the task performed (following procedures, positioning, etc.), doing the job with fewer errors, saving time on complex or regularly changing assembly tasks, and accelerating acquisition of skills for inexperienced operators.

5.4.2 Quality control

As a result of its ability to realistically incorporate a digital 3D model in the real world, AR makes it possible to assist in the assembly control process. The visual comparison offered to the human makes it easier to search for positioning defects or reference errors as shown in Figure 38. At the same time, an assembly completeness check is possible. If a fault is detected, the operator can generally take a photo directly from the application used and fill out a tracking form, which will later be added back to the company's information system.

The current state of technology does not make it possible, today, to automatically detect defects. Therefore, current technology constitutes a (non-automatic) visual assistance system, but one that can accelerate the control process while increasing its performance and exhaustiveness.

For quality control assistance, the needs for hands-free display solutions are rarely expressed. Additionally, the potentially long time that verifications take, together with the need to enter textual information, generally steer solutions toward tablets or fixed screens. In some cases, it may be wise to use projective systems (with a video projector) because they have the benefit of displaying a large quantity of information all at once, and thereby do not require that the operator that is looking for defects examine the environment through his or her tablet screen, the field of view of which is limited to a few tens of degrees.



Figure 38: AR application for quality control in industrial environment.

The technological obstacles and challenges for this use case primarily relate to seeking accurate enough augmentation positioning to be compatible with the positioning control task to be performed. Location algorithms must both (1) estimate the device's movements sufficiently accurately, and (2) enable precise spatial co-referencing between the digital model and the real object.

The expected benefits of the use of AR technologies for quality control tasks are improving assembly quality and reducing control time.

5.4.3 Field servicing and maintenance

AR may be a response to the problems encountered by operators and technicians in the field. Such workers must, for instance, complete maintenance tasks (corrective or preventive) or inspection rounds in large environments with a lot of equipment present. Based on the operator's level of expertise and the site of intervention, those activities may prove particularly complex.

In this context, AR enables valuable assistance to the user in the field by presenting him/her with information drawn from the company's document system, in an intuitive and contextualised manner. This includes maintenance procedures, technical documentation, map of inspection rounds, data from sensors, etc. as visualised in Figure 39.

Information from the supervisory control and data acquisition (SCADA) system may also be viewed in AR, so as to view values drawn from sensors and connected industrial objects (Industrial IoT). This data may thereby be viewed in a manner that is spatially consistent with respect to the equipment, or even to the sensors.

In large environments and in environments particularly populated with equipment, AR combined with a powerful location system may also be used like a GPS to visually guide the operator to the equipment that they need to inspect or service.

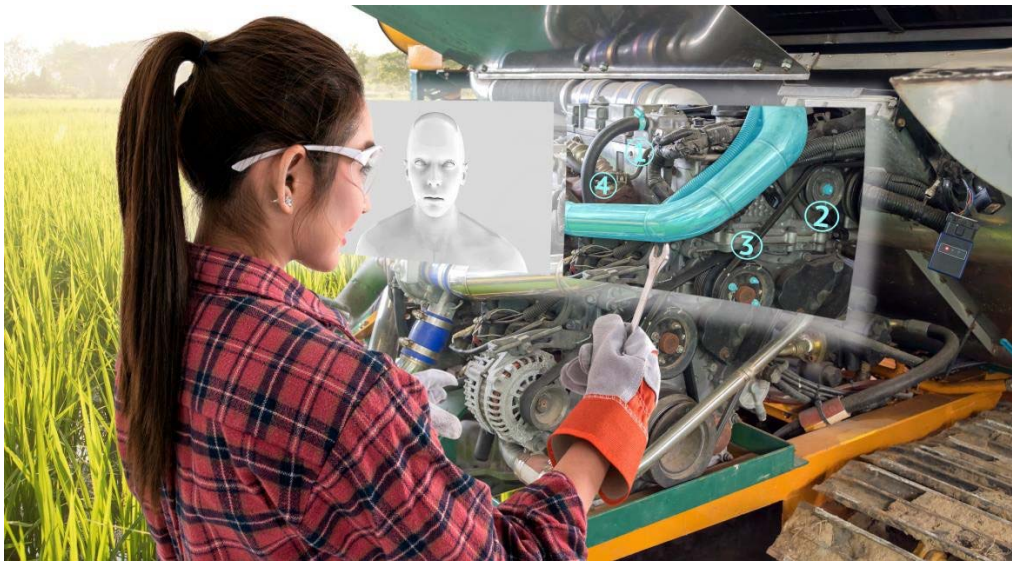


Figure 39: AR application for maintenance in industrial environment.

Finally, the most advanced AR may include remote assistance systems. These systems enable the operator in the field to share what they are observing with an expert (via video transmission, possibly in 3D), and to receive information and instructions from the expert, in the form of information directly presented in his/her visual field, and furthermore in a temporally and spatially consistent way with the real world in the best of cases.

The desired devices are generally head-mounted devices (HMDs), i.e. goggles or headsets, so as to leave the user's hands free. Despite this fact, the solutions currently deployed are most commonly based on tablets, because these devices are more mature and robust than currently available HMDs.

To assist technicians in the field, there are numerous solutions that may be incorrectly be called and viewed as being AR solutions. Without questioning its usefulness, a system based on an opaque heads-up display that shows instruction sheets to the user cannot be considered to be an AR system. If there is no spatial-temporal consistency between the augmentations and the real world, it is not AR.

The expected benefits of the use of AR technologies for maintenance tasks are (1) easing the access to all information in digital forms related to the management of life of industrial equipment, (2) reducing the mobilisation of operators who are trained and familiar with the equipment, (3) reducing errors during operations, (4) facilitating the update of procedures, and (5) tracing operations.

5.4.4 Factory planning

When designing new production equipment, testing its integration into its intended environment through VR or AR may be valuable. These simulations make it possible, for example, to identify any interference between the future equipment and its environment, to assess its impact on flows of people and materials, and to confirm that users will be safe, when moving machinery operates such as in the case of an industrial robot. To perform such simulations, it is necessary to be able to view, in a common frame of reference, both (1) the future production equipment, available digitally only, and (2) the environment that it will fit

into, which is either, real and physical, or represented by the digital twin of the factory. On the one hand, 3D scanning solutions make it possible to digitise the environment in 3D, but the compromise between the accuracy of the scan and the acquisition time make them of little use for a simple, quick visualisation in VR. On the other hand, AR technologies, offer the ability to incorporate the digital model into the actual environment without any additional, prior 3D scanning (see Figure 40).

From a hardware perspective, these applications generally require a wide field of view to enable the user to perceive the modelled production equipment in full. As of this writing, head-mounted devices do not meet this requirement, unlike hand-held devices (e.g. tablets). From a software perspective, as the real environment is usually not available in digital form, and accuracy takes precedence over aesthetics, the digital model's positioning is generally done using visual markers made up of geometric patterns with a high degree of contrast.

These simulations are helpful mainly at identifying potential problems of integration at a very early stage in the process of designing production machinery. These problems may thereby be anticipated and resolved long before the installation, which greatly reduces the number of adaptations needed on-site. These on-site modifications are generally very expensive, because, without even taking into account the direct cost of the modifications and the need to service the machine under more delicate conditions than in the shop, production must also be adapted or even stopped during servicing. Problems must therefore be identified and corrected as early as possible in the design cycle.

Furthermore, viewing the future means of production is an extremely powerful way of communicating, which, among other things, enables better mediation between designers and operators.

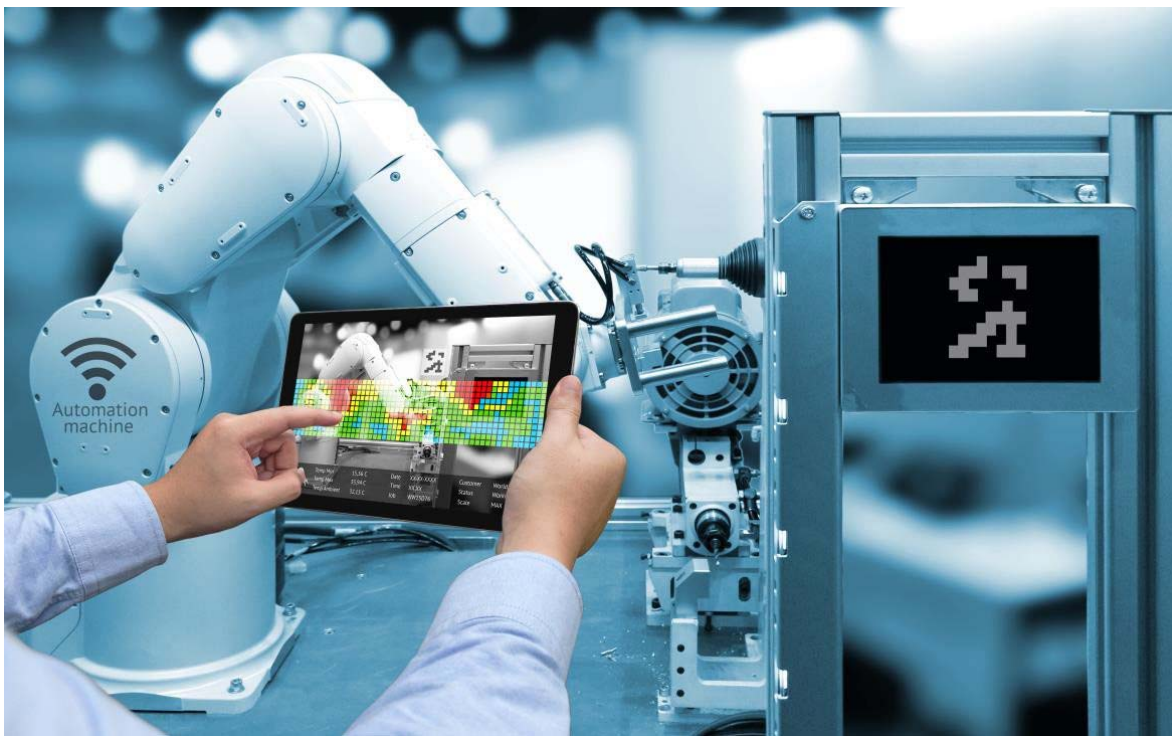


Figure 40: AR application for planning in industrial environment.

5.4.5 Logistics

A possible Industry 4.0 application is the effective management of warehouse operations in order to keep up with the supply chain needs by exploiting technological progress [293][294]. This will reduce the inventory; make the response time faster dealing better with the rapid increase in e-commerce transactions. The sales of smart glasses in 2017, according to ABI Research, reached the value 52.9 million dollars and about one out of four smart glasses were bought by the logistics industry [295].

Although the use of AR is still emerging in the field of logistics, it does appear to be a promising source of time savings. Potential uses of AR in warehouse operations are:

- Receiving:
 - Indicate the unloading dock to incoming truck driver;
 - Check received goods against delivery note;
 - Show where to put the items/how to arrange them in the waiting zone.
- Storing:
 - Inform an operator about a new allocated task;
 - Display the storage location of incoming items;
 - Display picture and details of the item to be stored;
 - Indicate route to storage location;
 - Indicate picker's current status as well as next step of the process;
 - Check locations requiring replenishment while storing.
- Picking:
 - Inform an operator about a new task allocated to him;
 - Display picture and details of the item to be picked;
 - Display the storage location of the item to be picked;
 - Display picking route;
 - Highlight the physical location with the item required;
 - Inform about errors and disruptions;
 - Scan the item's barcode to assign to picking car or to see more information;
 - Highlight where to put each item on the picking cart for sorting while picking;
 - Give information to prevent congestion in aisles;
 - Monitor picker's condition and performance;
- Shipping:
 - Show what type of cardboard to use;
 - Show the best way to place picked items in a package;
 - Indicate the right location/pallet for the shipment;
 - Show where to place each order on a pallet/in a truck according to type of orders, destination, fragility;
 - Indicate appropriate loading area;
 - Check/Count products/orders to be loaded on a truck.

Research has shown that in warehouse operations, the order-picking process typically account for approximately 55% of the total operational activity and traveling activity comprise the remaining 45% [296]. That is why technological advance focuses on being used in the picking process. A possible AR

application for a smart warehouse would be a sophisticated way of picking an order, which would reduce the operational time of picking the order by providing the fastest route [297]. An AR device displays the order-picking instructions, renders the virtual navigation and virtually marks the positions of the items.

In this context, AR enables superior anticipation of the order schedule and load management by connecting with management systems. The visual assistance made possible by AR enables workers to find their way around the site more quickly, using geolocation mechanisms that are compatible with the accuracy requirements of a large-scale indoor location scenario.

The visual instructions may also contain other information that is useful for completing the task, like the number of items in the order or the reference “numbers” of the parts ordered. Once the task is complete, the validation by the device causes stock management and operations supervisory system - often referred to as a Warehouse Management System (WMS) - to be updated in real time.

AR systems that make it possible to assist operators during the handling of pallets have also been created as shown in Figure 41.

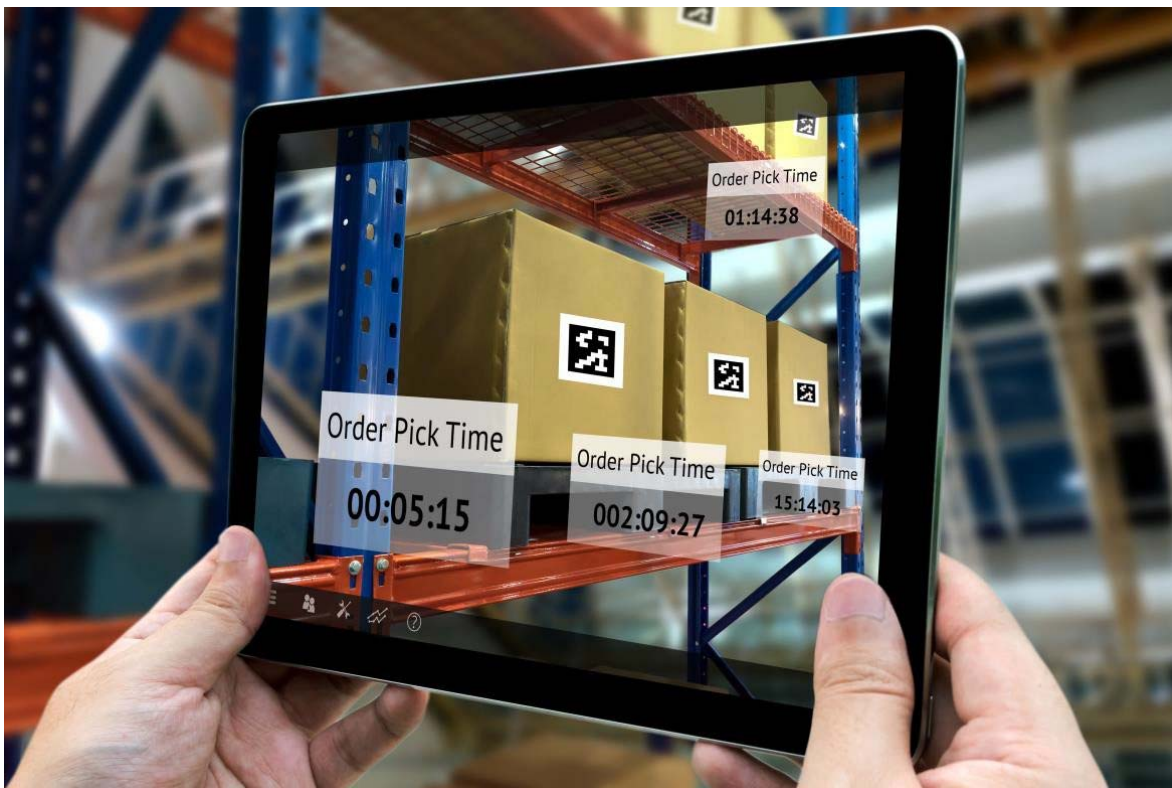


Figure 41: AR application for logistics.

To guide moving operators, the preferred technology consist in heads up displays. Through their use, guidance instructions are presented in the operator's natural field of view without him/her needing to be constrained in what they're doing. This is crucial for these highly-manually operations, as these displays leave the user's hands free. For interaction, most commonly to indicate that an order preparation step

has been completed, assistance systems generally use voice interaction (based on speech-recognition technology), or an interaction device linked to the relevant computer (like a smartphone attached to the forearm or a smart watch attached to the wrist). To remedy the generally short battery life of AR devices, a wired connection to the computer or additional batteries may be needed to achieve a usage duration compatible with a full work shift.

An AR solution is expected to limit errors while also saving time, particularly for novice staff.

A European project called SafeLog works on safe human-robot interaction in logistic applications for highly flexible warehouses. Many academic and research institutes are partners on this project, among others Swisslog and Fraunhofer IML [298][299].

In such a warehouse, as shown in Figure 42, a harmonic coexistence of robots and human is aimed. Humans are wearing a special vest, which sends signals to the robots and update them about the current location of the humans. This has an effect for the robots slowing down or even stopping when workers are nearby. Additionally, humans are wearing special glasses that allow them for example to see the path to the racks to pick up a specific item or allow them to see robots behind racks that would not be visible without the glasses. Figure 43 shows Safelog concept exhibition at Logimat trade fair in Stuttgart, Germany, 2019.



Figure 42: Warehouse concept by European Project SafeLog.



Figure 43: SafeLog concept exhibition at Logimat trade fair in Stuttgart, Germany (2019).

Using visual guiding for picking tasks can reduce the time needed but choosing the best guiding technique is not a trivial task. In Figure 44, a classification of attention guiding techniques is shown.

Technique	DoF	Supported Phases	Flexibility	Integration	Investment
Highlight	3D pos	Maneuvering	Multiple targets	in-situ & occlusions	see-through hw
Minimap	2D pos	[Navigation]	Multiple targets	HUD	all mobile devices
Arrow	2D/3D rot	Orientation	One target	HUD	all mobile devices
SWave	3D pos, 3D rot	Orientation, Maneuvering	One target	in-view & in-situ, no occlusions	3D glasses
Line	3D pos, 3D rot	Orientation, Maneuvering	One target	in-view & in-situ, no occlusions	see-through hw
Attention Funnel	3D pos, 3D rot	Orientation, Maneuvering	One target	in-view & in-situ, no occlusions	3D glasses
3D Path	3D pos, 3D rot	All	One target	in-view & in-situ, occlusions	3D glasses, spatial understanding

Figure 44: Classification of Attention Guiding Techniques [233].

Review of the above techniques give some insights when choosing a technique [300]:

- Orientation cues are required to make sure users quickly find the correct direction to go;
- Users benefit from information about other targets such that they are as fast or even faster when showing multiple targets in contrast to showing the best way to each target one after another;
- Users tend to prefer guiding techniques which leave some autonomy to them.

In a worker-oriented direction, an interesting question would be what makes an order picking support system not accepted by the worker. Research shows that seven barriers can play a role into a rejection of adoption [301]:

1. An overwhelmingly high subjective task load;
2. Loss of autonomy;

3. Loss of social interaction;
4. Negative influences from co-workers;
5. High complexity in handling the technology;
6. A lack of training;
7. A lack of maturity of the technology.

Therefore, it would be very important that the technology becomes mature enough and intuitive before using it in warehouse management operations.

5.4.6 Transportation

In the previous section, AR technology possibilities in warehouse management operations were discussed. Here we give some directions and ideas on how AR technology can be used in optimisation of transportation in areas such as completeness checks, international trade, driver navigation and freight loading as proposed in [302]:

- *Completeness Checks:* Currently, this process requires manual counting or time-consuming barcode scanning with a handheld device. An AR-equipped collector could quickly glance at the load to check if it is complete;
- *International Trade:* Before a shipment, an AR system could assist in ensuring the shipment complies with the relevant import and export regulations, or trade documentation has been correctly completed. After shipment, AR technology can significantly reduce port and storage delays by translating trade document text such as trade terms in real time;
- *Dynamic Traffic Support:* It's estimated that traffic congestion costs Europe about 1% of gross domestic product (GDP) each year [303]. Therefore, it is crucial to improve punctuality. AR driver assistance apps could be used to display information in real time in the driver's field of vision;
- For example, WayRay [304], a Swiss company has created a suite holographic augmented reality displays that turn the entire car windshield into a dynamic space that can display real-time navigation information and visual tools for Advanced Driver Assistance Systems (ADAS) (see Figure 45). It's expected that future iterations will incorporate V2X (Vehicle to Everything) technology, and will share information gleaned from transport and smart city applications such as traffic control, weather, and road alerts;
- *Freight Loading:* Freight transportation by air, water and road makes extensive use of digital data and planning software for optimised load planning and vehicle utilisation. The bottleneck is often the loading process itself. AR devices could help by replacing the need for printed cargo lists and load instructions. At a transfer station, for example, the loader could obtain real-time information on their AR device about which pallet to take next and where exactly to place this pallet in the vehicle. The AR device could display loading instructions identifying suitable target areas inside the vehicle.

Last-mile Delivery and Last-meter Navigation could also benefit from AR technology. Last-mile Delivery refers to the final step in the supply chain and often is the most expensive one. There has never been a time of greater change for the "last mile". Consumers order more things online, expecting more control and faster deliveries [305].



Figure 45: A holographic augmented reality display as developed by WayRay [304].

- **Parcel loading and drop-off:** Each driver could receive critical information about a specific parcel by looking at it with their AR device. The device could then calculate the space requirements for each parcel in real time, scan for a suitable empty space in the vehicle, and then indicate where the parcel should be placed, taking into account the planned route. In this way, the search process would be much more convenient and significantly accelerate every drop-off. In addition, AR could help to reduce the incidence of package damage. One of the key reasons why parcels get damaged today is that drivers need a spare hand to close their vehicle door, forcing them to put parcels on the ground or clamp them under their arm. With an AR device, the vehicle door could be closed 'hands-free' – the driver could give a voice instruction or make an eye or head movement.

Last-meter Navigation starts when the vehicle door is shut and the correct parcel is in the driver's hands and the driver has to find a specific building (see Figure 46).



Figure 46: Last-meter navigation prototype Error! Reference source not found..

AR could be extremely helpful here; AR could identify the correct building and entrance as well as indoor navigation. A learning system is able to add user-generated content to the AR map [306].

5.4.7 Training

VR and AR applied to the field of training is of great benefit and interest for learners, thanks to its visualisation and natural interaction capabilities. The use of VR & AR for training enables one to understand phenomena and procedures (see the example in Figure 47).

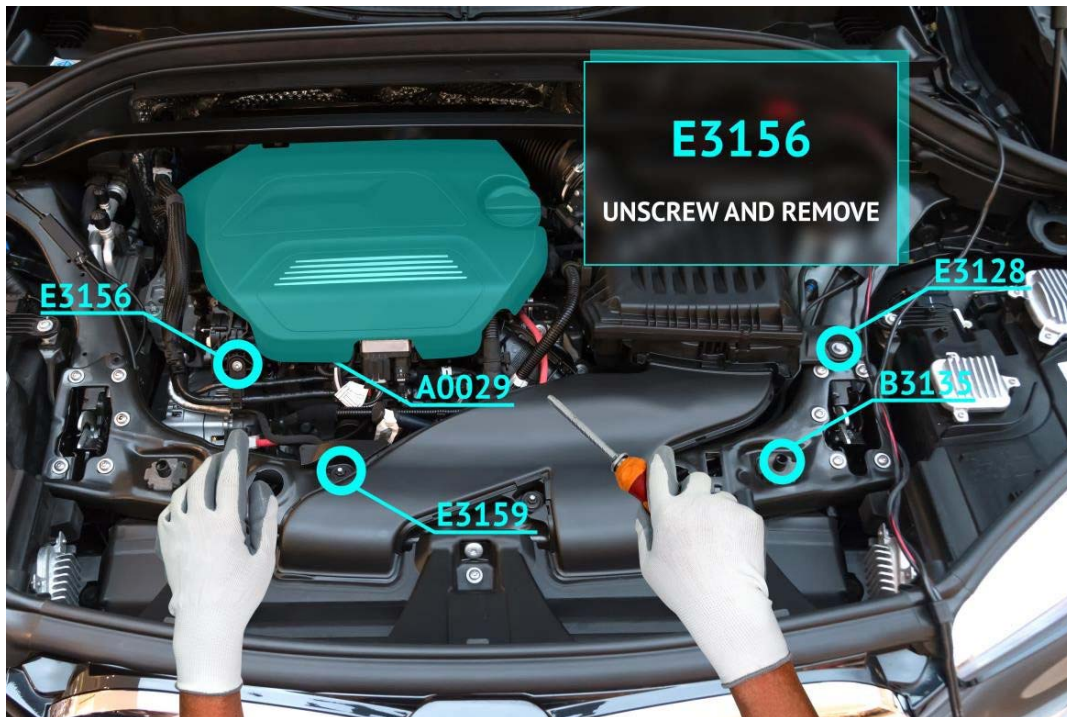


Figure 47: AR application for training in industrial environment.

For instance, this use offers a learner an "X-ray vision" of a piece of equipment allowing him/her to observe its internal operation. It also makes it possible to learn how to carry out complex procedures, by directly showing the different steps of assembly of an object, and, by extension, the future movements to perform. Instant visual feedback of how well the learned action was performed is also possible (speed of movement, positioning of a tool, etc.). The use of VR & AR also offers the chance to train without consuming or damaging materials, as in the case, e.g. for welding and spray-paint. Finally, it enables risk-free learning situations for tasks that could be hazardous in real life (such as operating an overhead crane).

For training purposes, many types of display are used, with the choice depending on the goal and condition of use:

- VR headset;
- CAVE, visiocube;
- Tablet, smartphone;

- Screen with camera;
- AR goggles.

The expected benefits of the use of XR technologies for training task are reduction in the costs and duration of training, and improvement in the quality of training and of memorisation of knowledge acquired.

5.4.8 References

- [292] D. Neiding. "Steps to prepare for Industry 4.0." Today's Motor Vehicles. <https://www.todaysmotorvehicles.com/article/industry-40-overview-to-getting-started/>
- [293] Stoltz, Marie-Hélène et al., "Augmented reality in warehouse operations: opportunities and barriers", *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 12979-12984, 2017.
- [294] A. Cirulis and E. Ginters, "Augmented reality in logistics", *Procedia Computer Science*, vol. 26, pp. 14-20, 2013.
- [295] O. Bay. "Logistics Leading the way in Augmented Reality Usage and Adoption." ABI Research. <https://www.abiresearch.com/press/logistics-leading-way-augmented-reality-usage-and-/>
- [296] J. J. Bartholdi, III and S. T. Hackman, *Warehouse and Distribution Science: Release 0.96*, Supply Chain and Logistics Institute, Atlanta.
- [297] U. K. Latif, and S. Y. Shin, "OP-MR: the implementation of order picking based on mixed reality in a smart warehouse", *The Visual Computer* 36, 2019, doi: 10.1007/s00371-019-01745-z .
- [298] SafeLog Project. <http://safelog-project.eu/> (accessed Nov. 12, 2020).
- [299] D. Puljiz, G. Gorbachev and B. Hein, "Implementation of augmented reality in autonomous warehouses: challenges and opportunities." *arXiv preprint arXiv:1806.00324*, 2018.
- [300] P. Renner, and T. Pfeiffer, "AR-glasses-based attention guiding for complex environments: requirements, classification and evaluation", in *Proc. of the 13th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, 2020.
- [301] J. Haase, and D. Beimborn, "Acceptance of Warehouse Picking Systems: A Literature Review." In *Proc. of the 2017 ACM SIGMIS Conference on Computers and People Research*, 2017.
- [302] Glockner, H. et al., *Augmented reality in logistics. Changing the way we see logistics - a DHL perspective*, [Online]. Available: http://www.dhl.com/content/dam/downloads/g0/about_us/logistics_insights/csi_augmented_reality_report_290414.pdf (accessed Nov. 12, 2020).
- [303] "Transport 2050: The major challenges, the key measures." European Commission. https://ec.europa.eu/commission/presscorner/detail/ga/Memo_11_197 (accessed Nov. 12, 2020).
- [304] Wayray. <https://wayray.com/> (accessed Nov. 12, 2020).
- [305] "The Future of the Last-Mile Ecosystem." World Economic Forum, [Online]. Available: http://www3.weforum.org/docs/WEF_Future_of_the_last_mile_ecosystem.pdf (accessed Nov. 12, 2020).

[306] “Augmented Reality in Logistics.” DHL Global Technology Conference 2015, [Online]. Available: https://na.eventcloud.com/file_uploads/b05d26158820d377ca7a022173486cb0_T.6_InnovationinPractise-AugmentedRealityinLogistics.pdf (accessed Nov. 12, 2020).

5.5 Health and medicine

In an analysis published at the ISMAR conference, Long Chen reported that the number of publications on AR addressing applications in health has increased 100-fold from the 2-year period of 1995-1997 to the 2-year period of 2013-2015, thus separated by 18 years [307]. At the 2017 edition of the Annual Meeting of the Radiological Society of North America (RSNA), Dr Eliot Siegel, Professor and Vice President of Information Systems at the University of Maryland, explained that the real-time visualisation of imagery from X-ray computed tomography (CT) and magnetic-resonance imaging (MRI) via VR or AR systems could revolutionise diagnostic methods and interventional radiology. The dream of offering doctors and surgeons the superpower of being able to see through the human body without incision is progressively becoming a reality. Four use cases are described below, i.e., training and learning, diagnostic and pre-operative uses, intra-operative uses, and post-operative uses.

5.5.1 Training and learning

Using AR and VR for training and learning is of major interest to trainees and students in a wide variety of fields (medical and others) thanks to its natural visualisation and interaction capabilities. This use allows for an understanding of phenomena and procedures, which is facilitated by an integration in the real world of virtual elements that can be scripted (see Figure 48).

For virtually all applications of training and learning, both VR and AR are relevant, and they provide specific benefits.

For both VR and AR, the benefits are as follows:

- Offer the learner a transparent view of an equipment or organ to observe its internal operation/functioning;
- Allow the learning, without risk (for people and material), of technical gestures for complex and/or dangerous procedures;
- Offer instant visual feedback regarding the quality of the gesture (speed of a movement, positioning of a tool, etc.).

As for AR, despite its lower maturity, it offers additional advantages compared to VR:

- The insertion of virtual elements in the real world facilitates their acceptance of the system/technology by the user and allows a longer use, compared to an immersion in a purely virtual scene;
- The capability of interactions between real and virtual elements, e. g. on a manikin, opens up the range of possibilities for broader and more realistic scenarios.

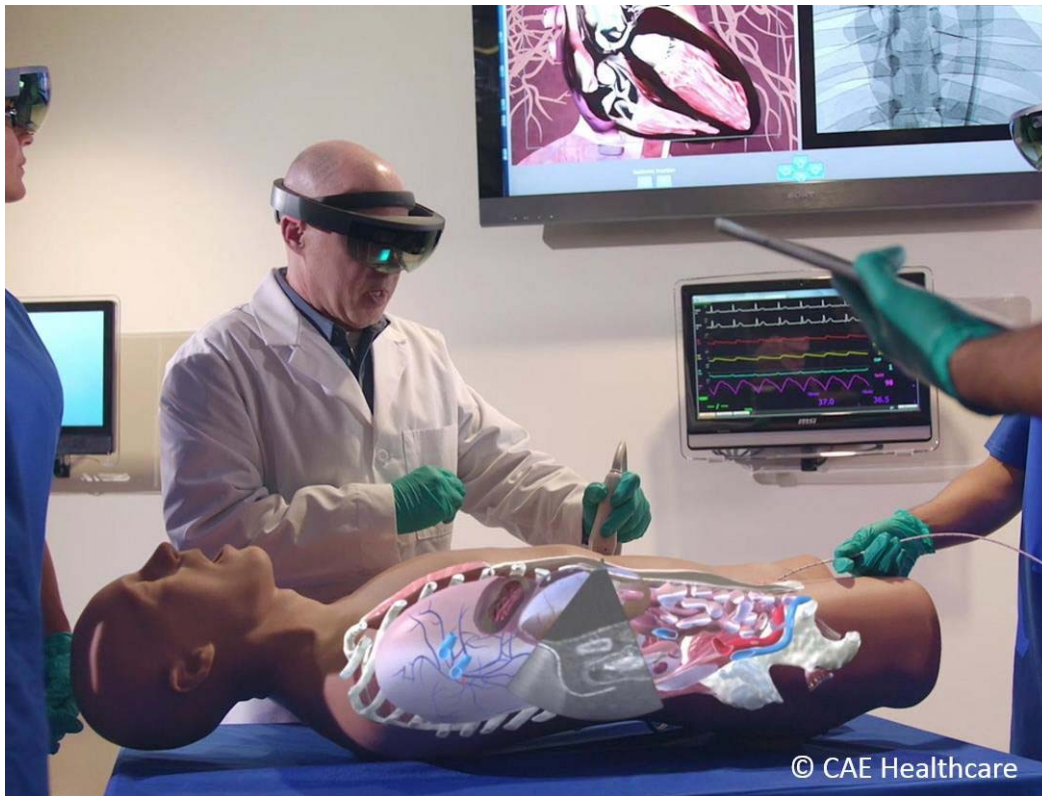


Figure 48: Example of training and learning use of AR in the medical domain.

To date, many proofs-of-concepts have been developed, but few immersive training solutions have actually been deployed in the medical field. The developments in this field can be classified in two main categories:

- Non-AR solutions to explore a complex 3D model (showing anatomy, pathologies, etc.) via a tablet or a headset, using advanced 3D visualisation and interaction techniques to manipulate objects (rotation, zoom, selection, cutting, excising, etc.);
- AR solutions allowing interactions between the real environment and the virtual scene, for example AR on a manikin. These solutions require higher accuracy, which can be achieved through the use of either visual markers or sensors.

The expected benefits of the use of XR technologies for medical training and learning are the easier and more effective knowledge acquisition, the reduction of cost and time through self-training (such as via tablets) and programmable scenarios (including through the use of a physical dummy), and the reduction in medical errors due to non-compliance with procedures.

5.5.2 Diagnostic and pre-operative uses

The diagnostic and pre-operative planning phases are generally based on the interpretation of previously acquired patient imaging data, such as from conventional radiography, X-ray computed tomography (CT), or magnetic resonance imaging (MRI). This data most often consists of 3D stacks of 2D sections allowing the reconstruction of a 3D image of the explored area, thus composed of 3D pixels, called "voxels". In

addition to simple interpretation, these 3D images can also be used to define a treatment plan, such as the trajectory of a tool, or the precise positioning of a prosthesis.

The visualisation and manipulation of this 3D data via a computer and a 2D screen sometimes present certain difficulties, especially for complex pathologies, and such interaction with the data is not very natural.

XR technologies can be extremely relevant for providing a more natural exploration of complex 3D medical imagery (see Figure 49 as an example). As for the previously described uses for training and learning, AR and VR applied to the present diagnostic and pre-operative uses share many of the same advantages. However, AR allows for better acceptance and is more suitable for collaboration and dialogue between, say, the radiologist, surgeon, and prosthesis manufacturer. Moreover, XR technologies may also be of interest to facilitate the dialogue between a doctor and his/her patient before a complex procedure. This step is indeed critical for pathologies with a strong emotional character, such as in paediatric surgery.

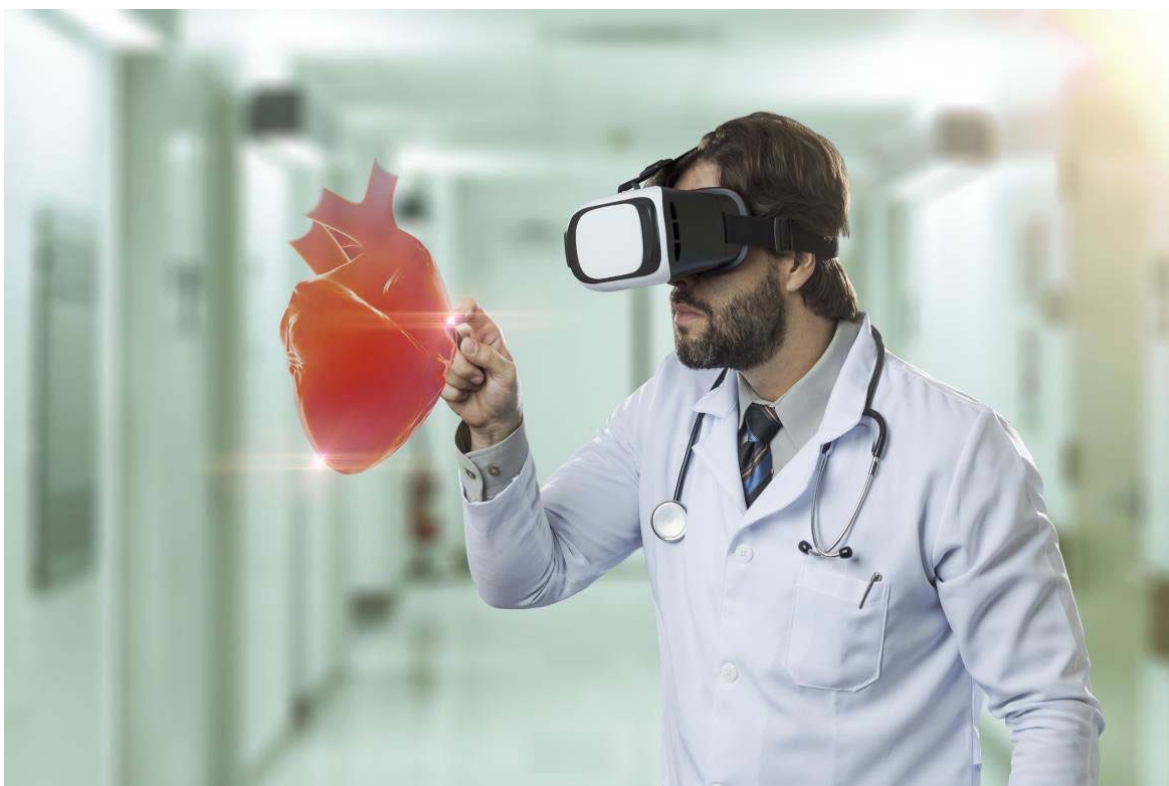


Figure 49: Example of pre-operative use of AR/VR.

The insertion of 3D imaging data into an AR or VR scene can be done in two ways:

- *By converting the patient's image data (based on voxels) into surface models (based on triangles):* This involves a segmentation step, which must be fast and accurate, and must be performed with the minimum manual intervention. The main 3D imaging software packages on the market provide advanced tools for segmentation and conversion. Fast and easy to handle, surface models have

certain limitations, particularly regarding the quality of the rendering, and the precision of the model, which depends on the initial segmentation (with appropriate thresholding, smoothing, etc.);

- *By rendering directly the patient's image data (based on voxels):* This technology allows a better 3D rendering thanks to the Volumetric Rendering algorithms, and offers more visualisation flexibility, such as display of cuts, and dynamic adjustment of the threshold. However, it is less used in AR-based solutions because it is more complex to implement and consumes more computing resources. Nevertheless, VR raytracing is gaining popularity, thanks to the last generation of NVIDIA's RTX architectures supporting real-time ray tracing scenarios.

XR solutions applied to diagnostic and pre-operative uses could increase reliability of complex procedures through a better understanding by the surgeon of the patient's anatomy and pathology in 3D before the actual surgical operation, a better collaboration between the team specialists during treatment planning, and a better communication with the patient before a complex procedure.

5.5.3 Intra-operative usage

Intra-operative uses of XR generally consist in helping the practitioner's gesture during the intervention by providing him/her with more information. Indeed, all interventional procedures require special visual attention from the surgeon and his/her team. In some cases, such as orthopaedic surgery, the doctor observes the patient and his instruments directly. In other procedures, such as minimally invasive surgery or interventional radiology, attention is focused on real-time images provided by room equipment (endoscope, ultrasound scanner, fluoroscopy system, etc.). Since the surgeon needs to maintain eye contact with the body of the patient, whether a direct contact or through real-time images, AR solutions are much more appropriate than VR solutions.

AR makes it possible to enrich the visual information available to the doctor by adding relevant and spatially-registered virtual information. This information can typically come from pre-operative 3D imagery, a preliminary planning step, and/or real-time imagery from various intra-operative imagers (such as US, CT, MRI):

- Information from 3D imagery, superimposed on the view of the patient, can be used to show, via transparency effects, structures that are not visible to the doctor's naked eye (internal organs, or organs hidden by other structures). This application is often referred to as the "transparent patient";
- Planning information, such as an instrument trajectory or the optimal location of a prosthesis, allows the practitioner to monitor in real time the conformity of his action with the treatment plan, and to correct it if he/she deviates from it;
- Information that allows instruments to be virtually "augmented" may also be relevant. Examples include, the display in real time of a 3D line in the extension of the axis of a biopsy needle, or the cutting plane extending the current position of an orthopaedic, bone-cutting-saw;
- The images from various sources of complementary information can be merged virtually to provide as much information as possible in a single image. For example, some endoscopes are equipped with an ultrasonic probe for real-time image acquisition near the current position. Displaying the recalibrated ultrasound cut plane on the endoscopic video allows the clinician to see, not only the tissue surface, but also the internal structures not visible in the endoscopic video.

It is important to note that, in some of these use cases involving medical imagery, the image format/geometry provided by the various imaging equipment (X-ray, CT, MRI, US, PET) cannot easily be

mixed with the classical "optical" view that the surgeon has of the patient. It takes a lot of training on the part of the surgeon to relate what he/she sees in the 3D coordinate frame of the patient in the real world, and what he/she sees in the 2D or 3D coordinate frame of the images of various modalities. In the cases where it is too complex to overlay the medical images on the screen of an AR headset or of a tablet, the medical imagery will continue to be presented on screens, where the view is not registered with the patient.

Intra-operative uses share some of the difficulties of pre-operative uses, particularly regarding the accuracy and fidelity of the virtual model. The main difficulty is the registration, which often requires very high accuracy. Indeed, a shift of the superimposed model could in some cases lead to misinterpretation and errors during the gesture.

This spatial registration is made particularly difficult in the case of "soft" organs such as the liver, or moving organs such as the heart or lungs, where the pre-operative model may not correspond to reality at the time of the gesture. AR applications must then use biomechanical modelling techniques, allowing the handling of organ deformations. In addition, in the case of motion, the accuracy of the time synchronisation of the two sources combined by AR has a direct impact on the accuracy of the spatial registration.

Techniques have been developed for tackling the problem of image-guided navigation taking into account organ deformation, such as the so-called "brain shift" encountered in neurosurgery upon opening of the skull. Some of these techniques use finite-element methods (FEMs), as well as their extension known as the extended finite-element method (XFEM) to handle cuts and resection. However, these techniques are very demanding in terms of computation.

The use of AR solutions for intra-operative uses provides a better reliability and precision of the intervention procedures thanks to the additional information provided to the practitioner, and this use can reduce the duration of surgery (see Figure 50 and Figure 51).



Figure 50: Example of intra-operative use of AR/VR.



Figure 51: Example of intra-operative use of AR.

During a surgical operation, a surgeon needs to differentiate between (1) healthy tissue regions, which have to be maintained, and (2) pathological, abnormal, and/or damaged tissue regions, which have to be removed, replaced, or treated in some way. Typically, this differentiation—which is performed at various times throughout the surgery—is based solely on his/her experience and knowledge, and this entails a significant risk because injuring important structures, such as nerves, can cause permanent damage to the patient’s body and health. Nowadays, optical devices—like magnifying glasses, surgical microscopes and endoscopes—are used to support the surgeon in more than 50% of the cases. In some particular types of surgery, the number increases up to 80%, as a three dimensional (3D) optical magnification of the operating field allows for more complex surgeries.

Nonetheless, a simple analogue and purely optical magnification does not give information about the accurate scale of the tissue structures and characteristics. Such systems show several drawbacks as soon as modern computer vision algorithms or medical augmented reality (AR)/ mixed reality (MR) applications can be applied. The reasons are now listed.

First, a beam splitter is obligatory to digitise the analogue input signal, resulting in lower image quality in terms of contrast and resolution. Second, the captured perspective differs from one of the surgeon’s field-of-view. Third, system calibration and pre-operative data registration is complicated and suffers from low spatial accuracy. Besides these limiting imaging factors, current medical AR systems rely on external tracking hardware, e.g. electro-magnetic tracking (EMT) or optical tracking systems based on infrared light using fiducial markers. These systems hold further challenges, since EMT can suffer from signal interference and optical tracking systems need an obstructed line-of-sight to work properly. The configuration of such a system is time-consuming, complicated, and error prone, and it interfere with, and even easily interrupt, the ongoing surgical procedure.

Furthermore, digitisation is of increasing importance in surgery and this will, in the near future, offer new possibilities to overcome these limitations. Fully-digital devices will provide a complete digital processing chain enabling new forms of integrated image processing algorithms, intra-operative assistance, and “surgical-aware” XR visualisation of all relevant information. The display technology will be chosen depending on the intended surgical use. While digital binoculars will be used as the primary display for

visualisation, augmentation data can be distributed to any external 2D/3D display or remote XR visualisation unit, whether VR headsets or AR glasses.

Thus, consulting external experts using XR communication during surgery becomes feasible. Both, digitisation and XR technology will also allow for new image-based assistance functionalities, such as (1) 3D reconstruction and visualisation of surgical areas, (2) multispectral image capture to analyse, visualise, segment, and/or classify tissue, (3) on-site visualisation of blood flow and other critical surgery areas, (4) differentiation between soft tissues by blood flow visualisation, (5) real-time, true-scale comparison with pre-operative data by augmentation, and (6) intra-operative assistance by augmenting anatomical structures with enriched surgical data [308][309][310][311][312].

5.5.4 Post-operative uses

XR solutions are also of great interest for the follow-up of the patient after surgery or an interventional procedure. The main medical issue in a post-operative context is to help the patient in his recovery, while monitoring and quantifying his progress over time. For newly orthopaedic prosthetic equipped patients (hip, knee, shoulder...), there is a need to quantify the so called Range Of Motion (ROM), which qualifies the articular mobility. It is important to note that ROM exams also happen in pre-operative phases, to have a pre/post-operative follow-up. This follow-up can take place in different places, according to the needs:

- Within the hospital;
- In specialised centres (rehabilitation centres);
- Home care.

After certain types of surgery, the patient must return to normal limb mobility through a series of rehabilitation exercises. XR then provides an effective way to support the patient in his or her home rehabilitation. For example, an image can be produced from a camera filming the patient, combining the video stream of the real world with virtual information such as instructions, objectives, and indications calculated in real time and adjusted based on the movements performed. Some solutions, such as "Serious Games", may include a playful aspect, which makes it easier for the patient to accept the exercise, thus increasing the effectiveness of this exercise.

VR solutions based on serious gaming approaches are actually available on the market for patient rehabilitation. For instance, Karuna [313], KineQuantum [314] and Virtualis [315] provide VR systems for physiotherapists as well as rehabilitation structures. These types of solutions can address physical/functional rehabilitation, as well as balance disorders, phobias, or elderly care, and require no additional hardware apart from a headset connected to a computer and some hand controllers. Some devices also couple VR with dedicated hardware, like for example Ezygain [316], which introduces VR scenarios on a smart treadmill for gait rehabilitation.

Also, the Swiss Company MindMaze aims to bring 3D virtual environment to therapy for neurorehabilitation [317][318] (see Figure 52, left). The company received series A funding of 110 M USD in 2016. Another example is the US company BTS Bioengineering Corp. that offers a medical device based on VR specifically designed to support motor and cognitive rehabilitation in patients with neuromotor disorders [319] (see Figure 52, right).

The European research project VR4Rehab specifically focuses on enabling the co-creation of VR-based rehabilitation tools [320]. By identifying and combining forces from SMEs active in the field of VR, research institutions, clinics and patients, VR4Rehab aims at creating a network of exchange of information and cooperation to explore the various use of state-of-the-art VR technology for rehabilitation potential, and to answer, as well as possible, and the needs of patients and therapists. The project is partly funded by Interreg Europe [321], a transnational funding scheme to bring European regions together.

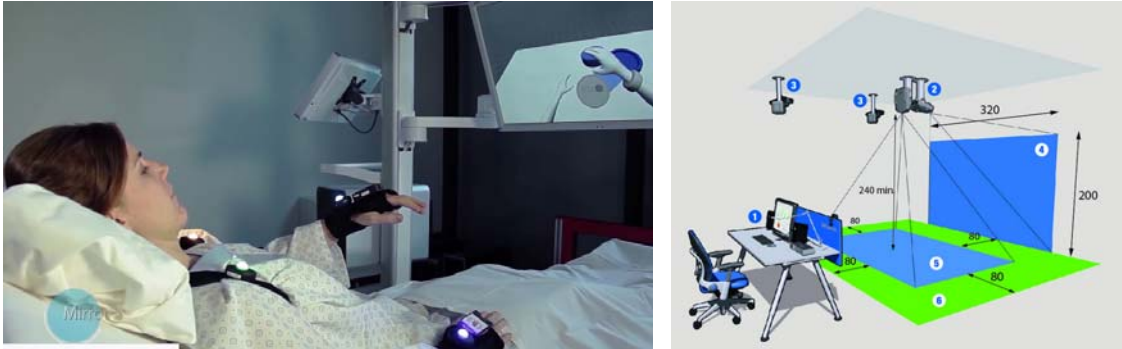


Figure 52: Mindmotion VR by MindMaze (left) and Nirvana by BTS Bioengineering (right).

The national project VReha in Germany develops concepts and applications for therapy and rehabilitation [322]. Researchers from medicine and other scientific domains, together with a medical technology company, exploit the possibilities of VR, so that patients can be examined and treated in computer-animated 3D worlds. Another example is the TeleRehabilitation project, which aims to create a rehabilitation path that combines self-rehab sessions for the patient and monitoring the rehabilitation through remote consultation with a health care professional. The proposed solution combines three different technologies: videoconferencing, VR/AR, and a 3D camera [323].

Concerning the use of AR for rehabilitation, some studies have led to real AR applications, like HoloMed [324], which has been led by the Artanim motion capture centre in Switzerland. It features a solution coupling HoloLens with professional MoCap system, enabling augmented visualisation of bone movements. They have developed an anatomical see-through tool to visualise and analyse patient's anatomy in real time and in motion for applications in sports medicine and rehabilitation. This tool will allow healthcare professionals to visualise joint kinematics, where the bones are accurately rendered as a holographic overlay on the subject (like an X-ray vision) and in real-time as the subject performs the movement. We can also talk about Altoida [325], which develops an Android/iOS app that allows testing of complex everyday functions in a gamified way, while directly interacting with a user's environment. It allows evaluation of three major cognitive areas: spatial memory, prospective memory and executive functions.

AR can also help a nurse working on in-home hospitalisation. Using glasses or a tablet filming the patient, the nurse will be able to communicate with a remotely-located doctor (telemedicine), who will help him/her via instructions added to the transmitted image. This can apply, for example, to wound monitoring at home or in a residential facility for dependent elderly people.

VR can help patients reduce pain following any trauma by diverting the patient's attention from his or her pain through an immersive experience. This technique has given very good results (1) with patients with burns over a large percentage of their body, by immersing them in a virtual polar environment, (2) with people with amputated limbs, to alleviate pain associated with phantom limbs, by displaying the missing limb thanks to VR/AR solutions, and (3) with patients with mental disorders.

For rehabilitation support systems, it is important to reproduce the patient's movement in real-time. This can be done by the video image itself, or by a more playful avatar, which only replays the movement concerned. Accurate and reliable motion reproduction can involve 3D (or RGB-D) cameras which provide, in addition to conventional video, a depth image capturing the 3D scene (see Figure 53).

The main benefits of XR solutions for post-operative uses are faster recovery through more effective and frequent home exercises, support for intervention assistance or remote monitoring; better monitoring of the patient's progress by the surgeon, and effective pain management.

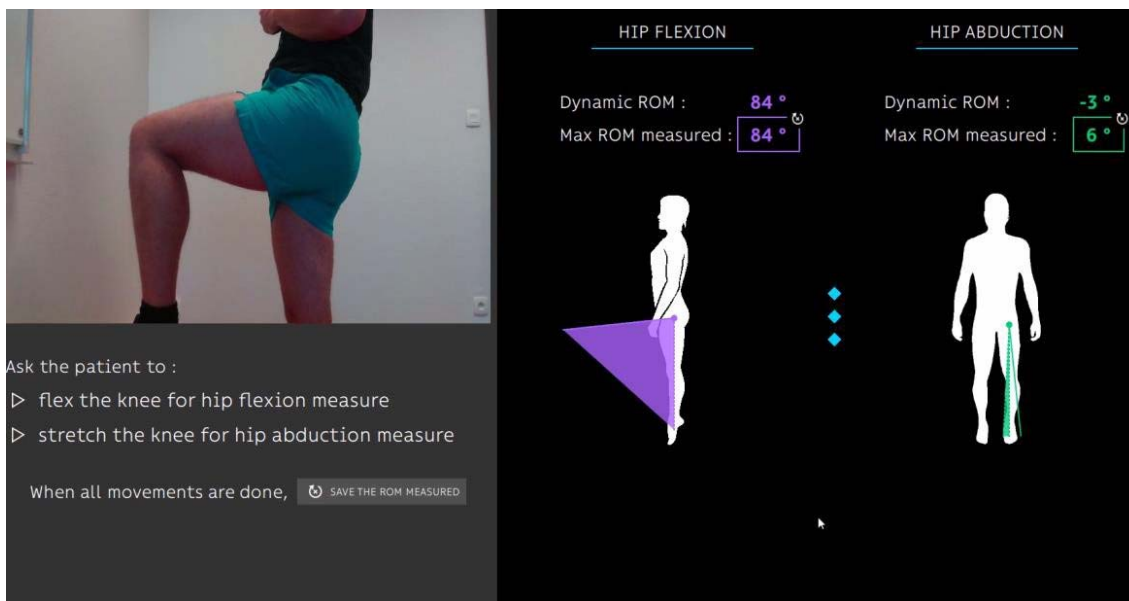


Figure 53: Example for post-operative use of VR/AR.

5.5.5 References

- [307] L. Chen, T. Day, W. Tang and N. W. John, "Recent Developments and Future Challenges in Medical Mixed Reality", *The 16th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2017.
- [308] "Medical Ray-tracing in VR". NVIDIA.
<https://on-demand.gputechconf.com/gtc/dc/2019/video/dc91185-medical-volume-ray-tracing-in-virtual-reality/> (accessed Nov. 20, 2020).
- [309] E. L. Wisotzky et al., "Interactive and Multimodal-based Augmented Reality for Remote Assistance using a Digital Surgical Microscope", *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Osaka, Japan, 2019.

- [310] B. Kossack, E. L. Wisotzky, R. Hänsch, A. Hilsmann, P. Eisert, “Local blood flow analysis and visualization from RGB-video sequences”, *Current Directions in Biomedical Engineering*, vol. 5, no. 1, pp. 373-376, 2019.
- [311] B. Kossack, E. L. Wisotzky, A. Hilsmann, P. Eisert, “Local Remote Photoplethysmography Signal Analysis for Application in Presentation Attack Detection”, in *Proc. Vision, Modeling and Visualization*, Rostock, Germany, 2019.
- [312] A. Schneider, M. Lanski, M. Bauer, E. L. Wisotzky, J.-C. Rosenthal, “An AR-Solution for Education and Consultation during Microscopic Surgery”, in *Proc. Computer Assisted Radiology and Surgery (CARS)*, Rennes, France, 2019.
- [313] Karuna. <http://www.karunalabs.com> (accessed Nov. 12, 2020).
- [314] KineQuantum. <http://www.kinequantum.com> (accessed Nov. 12, 2020).
- [315] Virtualis. <http://www.virtualisvr.com> (accessed Nov. 12, 2020).
- [316] ezyGain. <http://www.ezygain.com> (accessed Nov. 12, 2020).
- [317] Mindmaze. <https://www.mindmaze.com> (accessed Nov. 12, 2020).
- [318] Mindmotion. <https://www.mindmotionweb.com> (accessed Nov. 12, 2020).
- [319] NIRVANA. <https://www.btsbioengineering.com/nirvana/discover-nirvana/> (accessed Nov. 12, 2020).
- [320] Interreg NWE Programme. <https://www.nweurope.eu/projects/project-search/vr4rehab-virtual-reality-for-rehabilitation/> (accessed Nov. 12, 2020).
- [321] Interreg Europe. <https://www.interregeurope.eu/> (accessed Nov. 12, 2020).
- [322] VRha. <https://www.vreha-project.com/en-gb/home> (accessed Nov. 12, 2020).
- [323] “Telerehabilitation project”. <https://b-com.com/en/institute/bcom-galaxy/telerehabilitation> (accessed Nov. 20, 2020).
- [324] Artanim. <http://artanim.ch/project/holomed/> (accessed Nov. 12, 2020).
- [325] ALTOIDA. <http://www.altoida.com> (accessed Nov. 12, 2020).

5.6 Security and Sensing

The last few years, a lot of progress has been done concerning the hardware used for mixed reality experience as we discussed in Section 4.5. As the hardware advances, it will be more available and affordable, reaching out to more audience and arising new needs for security and privacy that are not discovered yet. For example, facial images can be used without approval of the person captured and used in facial matching tasks [326]. Mozilla also expressed concerns with regards to the privacy issues when using mixed reality applications [327]. For example, a malicious application could use biometric data like pupil tracking and perspiration to infer user’s political or sexual preferences.

In the survey made by Guzman et al. [326], the different security and privacy approaches to mixed reality to handle such issues were categorised as shown in Figure 54. There are five main security approaches that enclose the interaction cycle. Security and privacy process refers to protecting the input the user is providing, to protecting the data provided and to protecting the output. The way the user is interacting

with the technology should also be protected. Last, the device should be protected both physically and digitally.

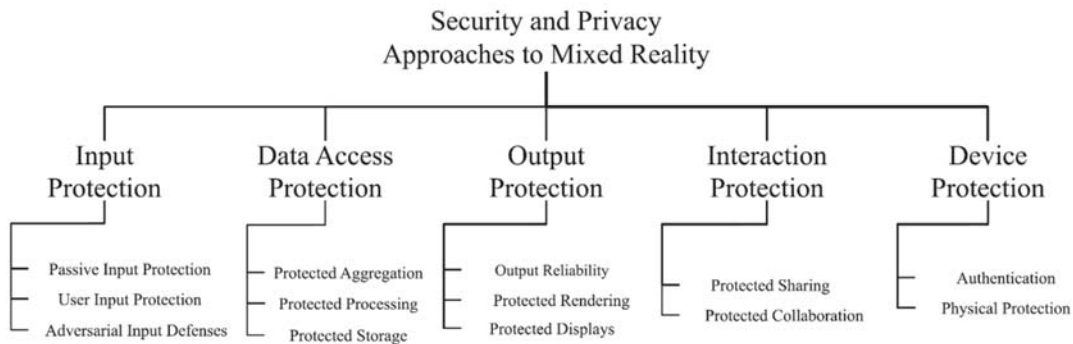


Figure 54: Security and Privacy Approaches to Mixed Reality.

In addition, current AR technology and systems could be used to enhance security and privacy [328] as shown in Figure 55. Here we see a prototype password manage application consisting of Google Chrome extension and Google Glass application. The Chrome extension modifies the browser's UI to display a QR code representing the website currently displayed to the user. Users can ask the Google Glass application to scan these QR codes and consult its password database by using the voice command "OK Glass, find password". If the user has previously stored a password for that website, the application displays the password; otherwise, the user can enrol a new password by asking the Chrome extension to generate an enrolment QR code and asking the Glass to store the new password using the "enrol password" voice command.

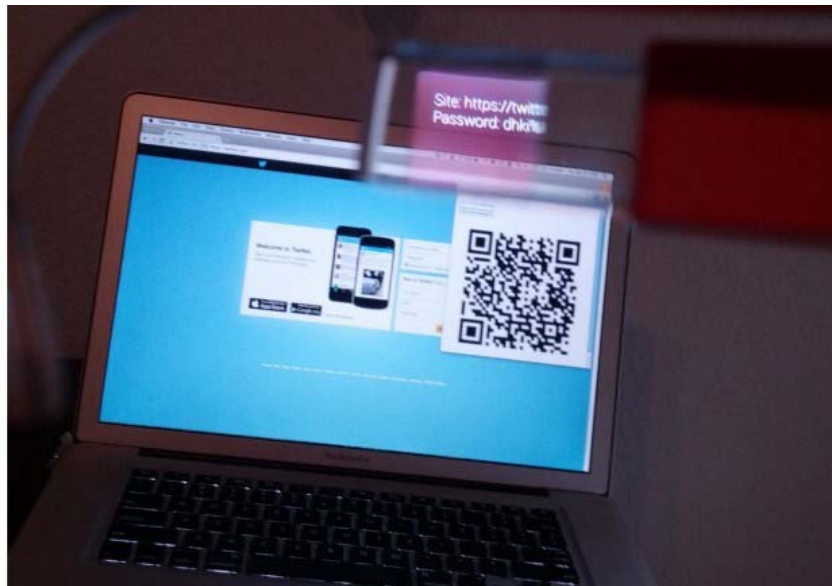


Figure 55: Use of AR to enhance security and privacy.

In addition to security issues rising when using mixed reality applications, there is also a prospect of using mixed reality for security reasons rising in real life. Next, we will discuss some platforms and studies that focus on using mixed reality to enhance security in real life.

Security staff and first responders have to deal with different levels of threats throughout their career. During their training, it is financially impossible to generate real-life threatening scenarios. AUGGMED, a mixed reality training platform, developed through a European project, addressed this issue and developed a safe, flexible training environment that can be accessed from any location by multiple agencies [329]. Mixed reality technology could be also used for cyber-physical security systems in the content of training new personnel [330]. In [331], a study was carried out, where the feasibility and usefulness of providing passive haptics in a mixed-reality environment to capture the risk-taking behavior of workers, identify at-risk workers, and propose injury-prevention interventions to counteract excessive risk-taking and risk-compensatory behavior was investigated. Figure 56 shows such an experimental setup where a mixed reality system is used to evaluate the risk-taking behavior of construction workers.

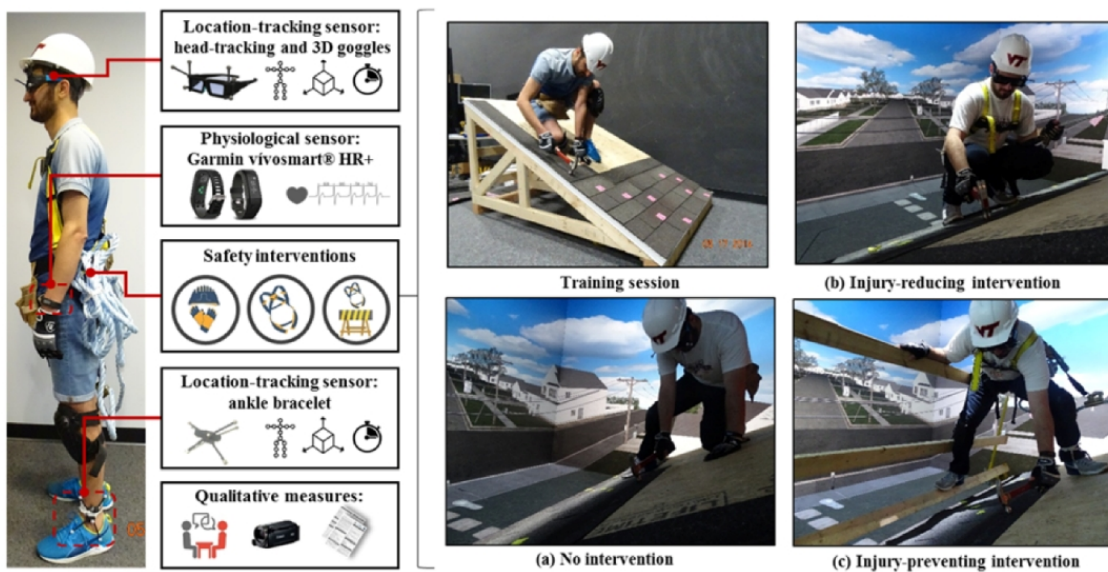


Figure 56: Experimental MR Systems to evaluate the risk-taking behaviour of construction workers.

In [332], a system is presented that exploits Mixed and Virtual Reality technologies to create a surveillance and security system that could also be extended defining emergency prevention plans in crowded environments. Recently in Japan an application was developed which contains flooding and fire smoke simulations in order to increase awareness and understanding of disaster risk [333]. In a fire-smoking scenario as shown in the Figure 57, a fire appears and smoke starts filling the room. The app prompts the user to go on hands and knees and crawl to escape.

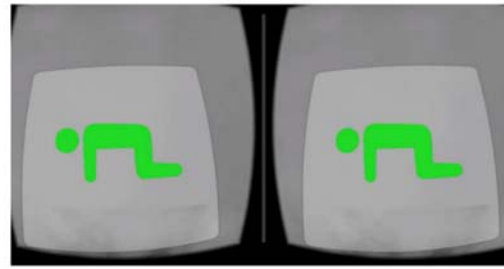


Figure 57: Smoke simulation to increase awareness and understanding of disaster risk.

Finally, we show how mixed reality technology has been used in defense system. BAE systems for example have produced the typhoon helmet, a helmet to be used by fighter pilots that help and support the pilot and let him 'see' through the body of the aircraft [334] as shown in Figure 58. Using the helmet system, the pilot can look at multiple targets, lock-on to them, and then, by voice-command, prioritise them.

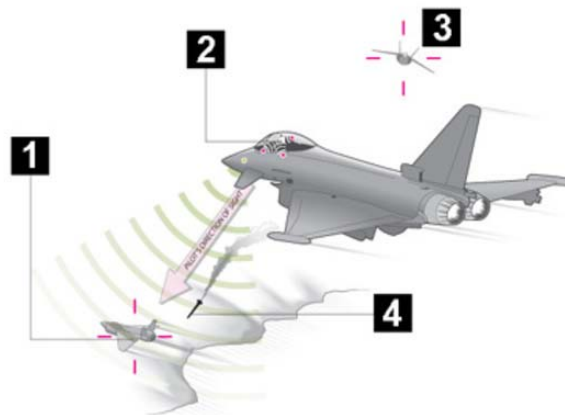


Figure 58: Typhoon helmet developed by BAE systems.

5.6.1 References

- [326] Jaybie A. de Guzman, K. Thilakarathna, and A. Seneviratne, "Security and Privacy Approaches in Mixed Reality", *ACM Computing Surveys(CSUR)*, vol. 52, no. 6, pp. 1–37, 2020.
- [327] D. Hosfelt, B. Macintyre. "Principles of Mixed Reality Permissions." Mixed Reality Blog. <https://blog.mozvr.com/principles-of-mixed-reality-permissions/> (accessed Nov. 12, 2020).
- [328] F. Roesner, T. Kohno and D. Molnar, "Security and privacy for augmented reality systems." *Communications of the ACM*, vol. 57, no. 4, pp. 88-96, 2014.
- [329] "Police and first responder training enters mixed reality." European Commission. <https://cordis.europa.eu/article/id/218536-police-and-first-responder-training-enters-mixed-reality> (accessed Nov. 12, 2020).

- [330] E. M. Raybourn and R. Trechter, "Applying Model-Based Situational Awareness and Augmented Reality to Next-Generation Physical Security Systems", *Cyber-Physical Systems Security*. Springer, Cham, 2018, pp. 331-344.
- [331] S. Hasanzadeh, N. F. Polys and J. M. de la Garza, "Presence, Mixed Reality, and Risk-Taking Behavior: A Study in Safety Interventions," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 5, pp. 2115-2125, May 2020, doi: 10.1109/TVCG.2020.2973055
- [332] D. Thalmann, P. Salamin, R. Ott, M. Gutiérrez, and F. Vexo, "Advanced mixed reality technologies for surveillance and risk prevention applications", in *Proc. of the 21st international conference on Computer and Information Sciences (ISCIS'06)*, Springer-Verlag Berlin Heidelberg, pp. 13–23, doi: https://doi.org/10.1007/11902140_2.
- [333] Tomoki Itamiya. "Disaster Scope: The Augmented Reality Floods and Smoke Simulated Experience Smartphone-Application." 2019.
- [334] BAE Systems. <https://www.baesystems.com/en/product/typhoon-helmet> (accessed Nov. 12, 2020).

5.7 Journalism & weather

A few years ago already, AR reached the news and weather reports. Graphical data as well as videos are augmenting virtual displays in TV studios and are an integral part of information delivery [335]. However, special weather apps are provided to the user with the aim that weather reports of the future will give more than just temperatures. The AccuWeather company recently announced the "Weather for Life" app, which allows someone to experience weather in VR.

In the domain of journalism, TIME has recently launched an AR and VR app, available on both iOS and Android devices, to showcase new AR and VR projects from TIME [336]. The first activation featured in TIME Immersive is "Landing on the Moon", which allows viewers to experience a scientifically and historically accurate cinematic recreation of the Apollo 11 landing in photo-real 3D on any table top at home.

5.7.1 References

- [335] IBM. <https://www.ibm.com/products/max-reality> (accessed Nov. 12, 2020).
- [336] "TIME Launches New Augmented Reality and Virtual Reality App, TIME Immersive, to Showcase Groundbreaking Visual Journalism." TIME. <https://time.com/5628880/time-immersive-app-ar-vr/> (accessed Nov. 12, 2020).

5.8 Social VR

When hearing the expression “social VR”, some people may think that it means using VR for some social actions, where “social” relates to things in the best interest of the public, like helping more fragile people, and in the sense of “social” in “social security”.

Even though the experts of the domain generally have a good intuitive feeling for what “social VR” means, one should note that there is no general agreement on a unique definition of “social VR”.

The PC Magazine Encyclopedia gives the following definition [337]:

Definition 1 of “social VR”:

“(social Virtual Reality) Getting together in a simulated world using a virtual reality (VR) system and social VR app. Participants appear as avatars in environments that can be lifelike or fantasy worlds.”

However, in his blog [338], Ryan Schultz indicates that he has searched the Internet for a good definition of “social VR” but that he has not found one that he likes. In relation to the above definition from PC Magazine, he says: “What I don’t like about this one is that it ignores platforms that are also accessible to non-VR users as well. There are quite a few of those!”

He then suggests using the following definition:

Definition 2 of “social VR”:

“Social VR (social virtual reality) is a 3-dimensional computer-generated space which must support visitors in VR headsets (and may also support non-VR users). The user is represented by an avatar. The purpose of the platform must be open-ended, and it must support communication between users sharing the same space. In almost all social VR platforms, the user is free to move around the space, and the content of the platform is completely or partially user-generated.”

Although “VR” appears in the conventional name “social XR”, one finds, in the scientific and technical literature (see below), systems that are similar but use AR instead of VR. It thus makes sense to also talk about “social AR” and, more generally, about “social XR”. In fact, below, one will see that one of the platforms of interest contains the term “XR”.

One should note that the above definitions do not limit “social VR” to gaming activities or to social exchanges between friends. Indeed, they allow for business activities, such as teleconferences and collaborative work. In fact, below, in a business/economic context, one will find the term “collaborative telepresence”, which may be a better and more-encompassing term.

5.8.1 Platforms

The following are examples of well-known “social VR” platforms (with the date of launch in parentheses):

- Second Life (2003) [339][340];
- High Fidelity (2013) [341][342];
- vTime (2015) [343][344];
- Rec Room (2016) [345][346].

A good account of the evolution of social VR from “Second Life” to “High Fidelity” is found in an article of the IEEE Spectrum of Jan 2017, which is based on a meeting of the author of the article and the founder of “Second Life” and “High Fidelity”, Philip Rosedale [347].

This article explains clearly that the key difference is that Second Life features a centralised architecture, where all the avatars and the interactions between them is managed in central servers, whereas High Fidelity features a distributed architecture, where the avatars can be created locally on the user's computer. The switch from “centralised” to “distributed” became necessary because the original platform (Second Life of 2003) did not scale up.

Philip Rosedale is convinced that, in the future, instead of surfing from one website (or webpage) to another, Internet users will surf from one virtual world to another. The transition from a page to a world is thus potentially revolutionary. This could become the “next Internet”. He is also convinced that many people will spend more time in a virtual world than in the real world.

One should also mention VR systems that allow communication in VR, such as

- Facebook Spaces [348], shut down by Facebook on 25 Oct 2019 to make way for Facebook Horizon;
- Facebook Horizon [349];
- VRChat [350];
- AltspaceVR [351].

5.8.2 Illustrations

The Figure 59 shows an example scene produced via the vTime platform. Those taking part in the platform choose their own avatars and control them as though they were in the virtual scene.



Figure 59: Illustration of interaction in a virtual space, here based upon the vTime platform.

Figure 60 shows an example scene produced via the Rec Room platform [352].



Figure 60: Illustration of interaction in a virtual space based upon the Rec Room platform.

5.8.3 A hot topic in 2019

On its website, the famed “World Economic Forum” lists the top 10 emerging technologies for 2019. One of them (#6, but without the order carrying any meaning) is “Collaborative telepresence”, sandwiched between “Smarter fertilizers” and “Advanced food tracking and packaging” [353]. Here is what the brief description says:

“6. Collaborative telepresence

Imagine a video conference, where you not only feel like you’re in the same room as the other attendees, you can actually feel one another’s touch. A mix of Augmented Reality (AR), Virtual Reality (VR), 5G networks and advanced sensors, mean business people in different locations can physically exchange handshakes, and medical practitioners are able to work remotely with patients as though they are in the same room.”

A more detailed description is found at [354].

In Sept 2019, Facebook founder M. Zuckerberg bet on the new social platform Facebook Horizon (already mentioned above) that will let Oculus users build their avatars, e.g., to play laser tag on the Moon. By contrast, in April 2019, Ph. Rosedale—creator of Second Life & founder of High Fidelity— (also mentioned above) dropped the bombshell that “social VR is not sustainable”, mainly as a result of too few people owning headsets. Thus, everything social in XR is currently a hot topic, all the more so that cheaper headsets are hitting the market, and 5G is being rolled out.

5.8.4 Mixed/virtual reality telepresence systems & toolkits for collaborative work

We give here, as a way of illustration/example, the list of MR/VR telepresence systems listed in Section 2.1 of the paper by M. Salimian [355]:

- Holoportation;

- Room2Room;
- System by Maimone and Fuchs;
- Immersive Group-to-Group;
- MirageTable.

We also give here, again as a way of illustration/example, the list of toolkits for collaborative work listed in Section 2.2 of the above paper:

- TwinSpace, SecSpace;
- Multi-User Awareness UI (MAUI);
- CAVERNsoft G2;
- SoD-Toolkit;
- PyMT;
- VideoArms, VideoDraw, VideoWhiteBoard, TeamWorkstation, KinectArms, ClearBoard;
- Proximity Toolkit, ProxemicUI.

5.8.5 Key applications and success factors

Gunkel et al. give four key use cases for “social VR”: video conferencing, education, gaming, and watching movies [356]. Furthermore, they give two important factors for the success of “social VR” experiences: interacting with the experience, and enjoying the experience.

5.8.6 Benefit for the environment

Collaborative telepresence has the huge potential of reducing the impact of business on the environment. Orts-Escolano et al. [357] state that despite a myriad of telecommunication technologies, we spend over a trillion dollars per year globally on business travel, with over 482 million flights per year in the US alone [358]. This does not count the cost on the environment. Indeed, telepresence has been cited as key in battling carbon emissions in the future [359].

5.8.7 Some terminology

Although we called this section “Interaction & social”, we indicated above that the conventional, historical term is “social VR”, which can be generalised to “social XR”. We also indicated that a good “synonymous” is “collaborative telepresence”. In some papers, such as by Misha Sra [360], one also finds “collaborative virtual environments (CVE)”. In this reference, one finds additional terminology that it is useful to be aware of.

- **Virtual environment or world** is the virtual space that is much larger than each user’s tracked space;
- **Room-scale** is a type of VR setup that allows users to freely walk around a tracked area, with their real-life motion reflected in the VR environment;
- **Physical space or tracked space** is the real-world area in which a user’s body position and movements are tracked by sensors and relayed to the VR system;
- **Shared virtual space** is an area in the virtual world where remotely located users can “come together” to interact with one another in close proximity. The shared area can be as big as the largest tracked space depending on the space mapping technique used. Each user can walk to, and in the shared area by walking in their own tracked space;

- **Presence** is defined as the sense of “being there.” It is the “...the strong illusion of being in a place in spite of the sure knowledge that you are not there” [361];
- **Co-presence**, also called “social presence” is used to refer to the sense of being in a computer generated environment with others [362][363][364][365];
- **Togetherness** is a form of human co-location in which individuals become “accessible, available, and subject to one another” [366]-. We use togetherness to refer to the experience of doing something together in the shared virtual environment. “

This is immediately followed by the remark: “While it is easy for multiple participants to be co-present in the same virtual world, supporting proximity and shared tasks that can elicit a sense of togetherness is much harder.”

5.8.8 Key topics for “social VR”

The domain of “social VR”, “collaborative telepresence”, and “collaborative virtual environment (CVE)” has already been the object of a lot of research, as is clear from the numerous references found below. All systems proposed are either at the stage of prototypes, or have limited capabilities.

Simplifying somewhat, the areas to be worked on in the coming years appear to be the following: (1) one needs to build the virtual spaces where the avatars operate and where the interaction takes place. These spaces can be life-like (like for application in business and industry) or fantasy-like; (2) one needs to build the avatars. Here too, the avatars can be life-like/photorealistic or fantasy-like. For the case of life-like avatars—thus representing a real person—one must be able to make this avatar as close as possible to the real person. This is a place where “volumetric imaging” should have a role. In one variation on this problem, one may need to scan a person in real time in order to inject a life-like/photorealistic avatar in the scene. A demonstration for this capability has been provided as part of the H2020 VR-Together project [367]; (3) one must synchronise the interaction between all avatars and their actions. This will likely require a mix of centralised and decentralised control. Of course, this synchronisation will depend on fast, low-latency communication, hence the importance of 5G; (4) social VR brings a whole slew of issue of ethics, privacy, and the like; (5) there is potential connection between social VR and both “spatial computing” and the “AR cloud”.

5.8.9 A potentially extraordinary opportunity for the future in Europe

“Collaborative telepresence” may represent the next big thing in the area of telecommunication and teleworking between people. It involves a vast, worldwide infrastructure. It involves complex technology, some still to be developed, to allow people to enjoy virtual experiences that are as close as possible to what we know in the real world, including what we perceive with all five senses.

In addition, this domain brings in a new set of consideration in privacy, ethics, security, and addiction, among others. The domain thus involves a lot of different disciplines. Since the deployment of collaborative-telepresence systems involves a lot of technologies and mostly a lot of software and algorithms, this may be an excellent and significant area for Europe to invest massively in for the next 5-10, including in research of course.

5.8.10 References

[337] PCMag. www.pcmag.com/encyclopedia/term/69486/social-vr (accessed Nov. 12, 2020).

- [338] R. Schultz. “UPDATED: What is the Best Definition of Social VR?”
<https://ryanschultz.com/2018/07/10/what-is-the-definition-of-social-vr> (accessed Nov. 12, 2020).
- [339] Second Life. <https://secondlife.com> (accessed Nov. 12, 2020).
- [340] Wikipedia. https://en.wikipedia.org/wiki/Second_Life (accessed Nov. 12, 2020).
- [341] High Fidelity. <https://www.highfidelity.com> (accessed Nov. 12, 2020).
- [342] Wikipedia. [https://en.wikipedia.org/wiki/High_Fidelity_\(company\)](https://en.wikipedia.org/wiki/High_Fidelity_(company)) (accessed Nov. 12, 2020).
- [343] vTime. <https://vtime.net> (accessed Nov. 12, 2020).
- [344] Wikipedia. https://en.wikipedia.org/wiki/VTime_XR (accessed Nov. 12, 2020).
- [345] REC ROOM. <https://recroom.com> (accessed Nov. 12, 2020).
- [346] Wikipedia. [https://en.wikipedia.org/wiki/Rec_Room_\(video_game\)](https://en.wikipedia.org/wiki/Rec_Room_(video_game)) (accessed Nov. 12, 2020).
- [347] D. Kushner. “Beyond Second Life: Philip Rosedale’s Gutsy Plan for a New Virtual-Reality Empire.” IEEE Spectrum. <https://spectrum.ieee.org/telecom/internet/beyond-second-life-philip-rose-ales-gutsy-plan-for-a-new-virtualreality-empire> (accessed Nov. 12, 2020).
- [348] Facebook. <https://www.facebook.com/spaces> (accessed Nov. 12, 2020).
- [349] Oculus. www.oculus.com/facebookhorizon (accessed Nov. 12, 2020).
- [350] VR Chat. <https://hello.vrchat.com> (accessed Nov. 12, 2020).
- [351] AltspaceVR. <https://altvr.com> (accessed Nov. 12, 2020).
- [352] “As Social VR Grows, Users Are the Ones Building Its Worlds.” WIRED. www.wired.com/story/social-vr-worldbuilding (accessed Nov. 12, 2020).
- [353] J. Wood. “These are the top 10 emerging technologies of 2019.” World Economic Forum. <https://www.weforum.org/agenda/2019/07/these-are-the-top-10-emerging-technologies-of-2019/> (accessed Nov. 12, 2020).
- [354] “Top 10 Emerging Technologies 2019.” World Economic Forum. http://www3.weforum.org/docs/WEF_Top_10_Emerging_Technologies_2019_Report.pdf (accessed Nov. 12, 2020).
- [355] M. Salimian, S. Brooks, D. Reilly, “IMRCE: a Unity toolkit for virtual co-presence”, in *Proc. of the Symposium on Spatial User Interaction (SUI '18)*, Berlin, Germany, 2018.
- [356] S. Gunkel, H. Stokking, M. Prins, O. Niamut, E. Siahaan, P. Cesar, “Experiencing virtual reality together: social VR use case study”, in *Proc. of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video(TVX '18)*, SEOUL, Republic of Korea, 2018.
- [357] S. Orts-Escolano et al., “Holoportation: Virtual 3D teleportation in real-time”, in *Proc. of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*, Tokyo, Japan, 2016.
- [358] K. Rapoza. “Business Travel Market To Surpass \$1 Trillion This Year.” Forbes. <https://www.forbes.com/sites/kenrapoza/2013/08/06/business-travel-market-to-surpass-1-trillion-this-year/> (accessed Nov. 12, 2020).
- [359] D. Biello. “Can Videoconferencing Replace Travel?” Scientific American. <https://www.scientificamerican.com/article/can-videoconferencing-replace-travel/> (accessed Nov. 12, 2020).

- [360] M. Sra, A. Mottelson, Pattie Maes, “Your place and mine: Designing a shared VR experience for remotely located users”, in *Proc. of the 2018 Designing Interactive Systems Conference (DIS ’18)*, pp. 85-97, 2018. DOI: <https://doi.org/10.1145/3196709.3196788>.
- [361] Mel Slater. 2009. “Place Illusion and Plausibility can Lead to Realistic Behaviour in Immersive Virtual Environments”, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 364, no. 1535, pp. 3549–3557, 2009, doi: <http://dx.doi.org/10.1098/rstb.2009.0138>
- [362] F. Biocca and C. Harms, “Defining and Measuring Social Presence: Contribution to the Networked Minds Theory and Measure”, in *Proc. of PRESENCE 2002*, pp. 7–36, 2012.
- [363] J. Short, E. Williams and B. Christie, *The Social Psychology of Telecommunications*, London, UK: Wiley, 1976.
- [364] N. Durlach and M. Slater, “Presence in Shared Virtual Environments and Virtual Togetherness”, in *Presence*, vol. 9, no. 2, pp. 214-217, April 2000, doi: 10.1162/105474600566736.
- [365] R. Schroeder, “Copresence and Interaction in Virtual Environments: An Overview of the Range of Issues”, in *Presence 2002: Fifth international workshop*, pp. 274--295.
- [366] Erving Goffman, *Behavior in Public Places*, Free Press. 2008.
- [367] VR Together. <https://vrtogether.eu/> (accessed Nov. 12, 2020).

5.9 Travel and Tourism

The application of XR technologies has demonstrated a lot of value for the travel and tourism sector. Many traveller practitioners, Destination Marketing Organisations, and start-ups have been experimenting with the use of VR technology for tourism. The main tourism-related uses for VR can be seen within six principal areas of tourism:

1. Planning and management;
2. Marketing;
3. Entertainment;
4. Education;
5. Accessibility;
6. Heritage preservation.

5.9.1 Planning and management

VR’s attributes render it exceptionally apt for the visualisation of spatial environments, which is why VR is commonly utilised for urban, environmental, and architectural planning. It permits the creation of realistic, navigable models that tourism planners can evaluate from an unlimited number of perspectives when considering possible developments [368]. VR has also been used as a tool for communicating tourism plans to members of community, and to invite input from stakeholders [369].

5.9.2 Marketing

VR’s unparalleled strength as a marketing tool lies in its ability to provide a sensory experience of a product, service, or destination to a prospective tourist. We have seen many innovative ways of this application. International brands such as Airbus, Qantas, British Airways, as well as Destination Marketing Organisations (“DMOs”) have started implementing VR advertising in their communication strategies,

both online and offline. One of the most notable examples of VR travel experiences was the “Marriot Teleporter” (see Figure 61). The user could visit destinations without packing a bag or boarding a plane. Using the Oculus Rift and 4D sensory elements, they created the “Teleporter” to virtually send users to several locations around the world in an immersive experience. As a tool for marketing, this VR experience proved to have increased Marriott’s customer demand for these destinations by 51%.



Figure 61: Marriot Teleporter.

Various studies have argued the benefits of integrating VR technologies into travel marketing. For example, virtual experiences provided more effective advertising than brochures for both theme parks and natural parks [370]. Researchers have found that a ‘virtual tour’ of panoramic photos on a hotel website may offer psychological relief to travellers experiencing travel anxiety [371]. Similarly, projects such as ScotlandVR [372] and Virtual Helsinki [373] recreate the destinations using a mix of 360-degree video, and animated maps, menus and photos. The Chief Executive of Visit Scotland commented that “far from being a fad or gimmick, VR is revolutionising the way people choose the destinations they might visit, by allowing them to ‘try before they buy’ and learn more about the country in a unique and interactive way”.

5.9.3 Entertainment

In addition to being a marketing tool, VR tourism attractions and experiences can serve as entertainment. Some experiences are designed for use at home. For example, Rewind Rome 3D used stereoscopy and 3D digital designs based on exacting historical research that transported the viewer into the daily life or ancient Rome [374]. Another example are the interactive virtual tours designed by Globetrotter VR [375]. The company uses a combination of reality capture, panoramic images, and Web VR technology to recreate “edu-cament” tours around popular tourism locations (see Figure 62). The company offers live guided tours where a tour guide takes the guests around the virtual environment in an online session of up to 10 people, providing opportunity for questions and real-time interaction much like a classic walking tour.

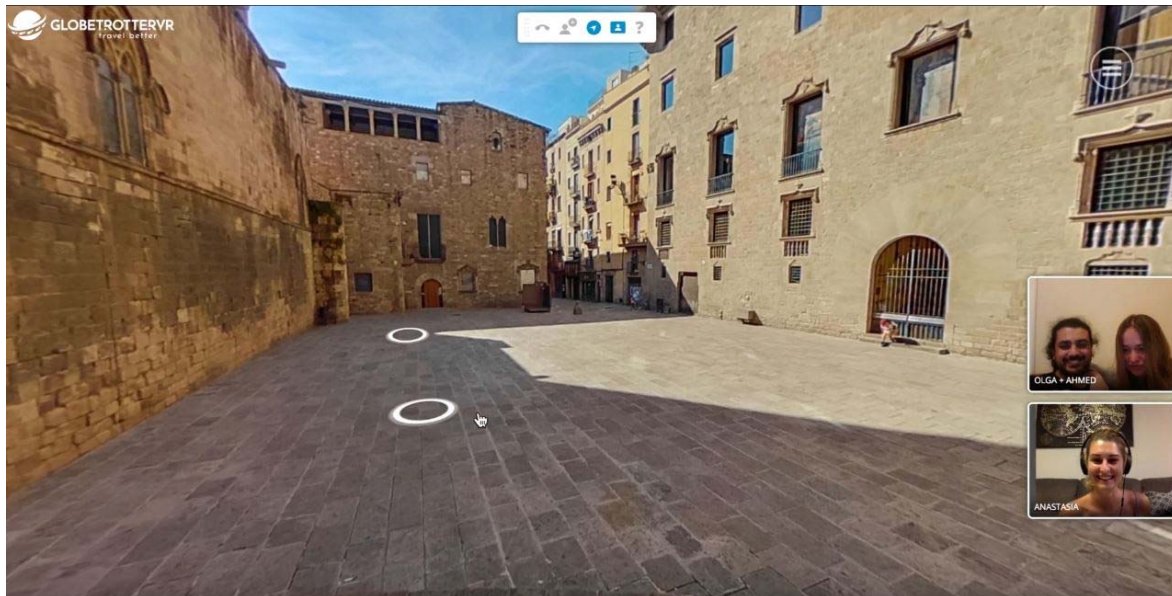


Figure 62: Globetrotter VR Live Virtual Tour.

VR can also be offered as entertainment in theme parks. Disney has used the technology to create the 'Aladdin's Magic Carpet Ride', where the user is wearing an HMD and uses a motorcycle-like machine to fly on a virtual magic carpet. France's Futuroscope is a theme park that leverages immersive technologies with several 3D and 4D cinema, and interactive installations [376].

5.9.4 Education and tourism

Aside from being highly entertaining, VR also has enormous potential as an educational tool. Firstly, VR allows great potential for interaction, has the possibility to add multimedia information into the experience allowing access to an array of valuable information through a single product. Moreover, the entertaining qualities of VR, which have been noted in some studies of VR and learning [377][378], are important to recognise because they can offer us solutions on how to keep the user engaged and focused on the learning material. VR's educational potential has been exploited in museums, heritage areas, and other tourist sites.

For example, the Foundation of the Hellenic World has created a VR installation that allowed users to journey through the ancient city of Miletus, become archaeologists who reassemble ancient vases from virtual shards of ceramic, conduct virtual experiments related to some of Archimedes' discoveries, and assist an ancient sculptor in creating a statue of Zeus [379]. The Foundation also launched an interactive 130-person virtual theatre "Tholos", where the show is interactive, and controlled by the spectator [380].

On the other hand, AR's capacity to superimpose educational material over the real world can also be useful for education. For example, several Portuguese heritage sites, including the Lisbon National Pantheon and the 12th century Pinhel Castle, have introduced fixed AR devices that look like traditional tourist binoculars but display images on a single, larger screen. Through these devices the traveller has access to a collection of illustrative information superimposed over the spots being viewed [381].

5.9.5 Accessibility

VR provides a unique opportunity to access historical sites and places of interest. While such access is limited only to the virtual world, it can be the desirable choice in cases where an actual visit may be impossible. For example, a tourist site may be too expensive, too far away, too dangerous, or simply no longer exist. In addition to providing a best possible alternative in such scenarios, virtual models permit unique interaction with historical objects or other fragile items that cannot be handled in the real world.

For instance, a Glasgow-based company Soluis has created a mobile app that uses VR technology optimised for a Google Cardboard headsets that allows the user to explore the famous rock art site of Game Pass Shelter in South Africa via an immersive 360° tour with embedded 3D models [382]. Another striking example of the use of VR and photogrammetry to recreate a world that no longer exists, is Memoria: Stories of La Garma. This is an interactive virtual reality journey that allows the audience to explore the memories, paintings and objects trapped inside the cave of La Garma in Cantabria, Spain for more than 16.000 years (see Figure 63).



Figure 63: Memoria: Stories of La Garma.

VR's capacity to facilitate access to sites can benefit everyone, but this function is especially helpful for disabled individuals. In situations where facilitating disabled access can be impossible due to conservation requirements or prohibitively large costs, VR can provide an alternative forms of access. For example, Shakespeare's Birthplace in Stratford-upon-Avon has installed a VR exhibit on the ground floor that offers visitors the opportunity to explore the various levels of the grand house [383]. Finally, many online virtual experiences can offer people with disabilities or serious illnesses the opportunity to visit remote places and take part in activities such as sky-diving or skiing in the Alps that they wouldn't be able to do in real life.

5.9.6 Preserve heritage from mass tourism

The world-wide mass tourism is considered as the most important reason of damage of cultural heritage sites. The Acropolis in Athens, the pyramids of Gizhee or even under-water cultural heritage sites need to take actions to preserve from daily tourism. Recently, the table mountain in South Africa has been closed for the public. Therefore, virtual visits will offer an important contribution to preserve cultural heritage sites.

The list of heritage sites and historical objects that can be accessed virtually is continuously growing and numerous heritage sites and objects from around the world already have been digitised as 3D virtual models. Notable examples include 3D models of Michelangelo's statues of David [384], 150 sculptures from the Parthenon [385], a virtual recreation of Cambodia's Angkor Wat temples [386], the Hawara pyramid complex from ancient Egypt [387].

Rendering such sites and objects as virtual 3D models serves as a valuable tool for heritage preservation because such virtual models can contain exceptionally accurate data sets that can be stored indefinitely. Furthermore, while a historical site or object may suffer from the impact of time, a virtual model can provide detailed information on its previous state that can be used both to monitor degradation and provide a blueprint for restorative works. Finally, large number of travellers overwhelm some of the world's most treasured sites, particularly those listed as UNESCO World Heritage Sites that tend to attract the largest number of tourists. Numerous researchers have suggested that VR potentially could help to preserve our global heritage by offering an alternative form of access to threatened sites [388].

5.9.7 References

- [368] R. Cheong, "The virtual threat to travel and tourism". *Tourism Management*, Vol. 16 (6), Elsevier Ltd., Sept. 1995, pp.417–422, [https://doi.org/10.1016/0261-5177\(95\)00049-T](https://doi.org/10.1016/0261-5177(95)00049-T).
- [369] D.A. Guttentag, "Virtual reality: Applications and implications for tourism", *Tourism Management*, Vol.31 (5), Elsevier Ltd., Oct. 2010, pp.637-651, <https://doi.org/10.1016/j.tourman.2009.07.003>.
- [370] C.-S. Wan, S.-H. Tsaur, Y.-L. Chiu, W.-B. Chiou, "Is the advertising effect of virtual experience always better or contingent on different travel destinations?", *Journal of Information Technology & Tourism*, Vol. 9(1), 2007, pp.45–54.
- [371] O. Lee, J.-E. Oh, "The impact of virtual reality functions of a hotel website on travel anxiety", *Cyberpsychology & Behavior*, Vol.10 (4), Sept. 2007, pp.584–586, DOI: 10.1089/cpb.2007.9987.
- [372] Scotland VR. <https://www.visitscotland.com/campaign/avis/app/> (accessed Nov. 14, 2020).
- [373] Virtual Helsinki. <https://www.virtualhelsinki.fi/> (accessed Nov. 14, 2020).
- [374] 3DRewindRome. <http://rome4u.com/museums/3drewind.html> (accessed Nov. 14, 2020).
- [375] GlobetrotterVR. <https://globetrotter-vr.com> (accessed Nov. 14, 2020).
- [376] Futuroscope. <https://www.futuroscope.com/en/attractions-and-shows> (accessed Nov. 14, 2020).
- [377] D. Allison, B. Wills, D. Bowman, J. Wineman, L. Hodges, "The Virtual Reality Gorilla Exhibit", *IEEE Computer Graphics and Applications*, Vol.17 (6). 1997, pp.30-38. DOI: 10.1109/38.626967.
- [378] M. Roussou, M. Oliver, M. Slater, "The virtual playground: An educational virtual reality environment for evaluating interactivity and conceptual learning", *Virtual Reality*, Vol.10 (6), 2006, pp.227-240, DOI: 10.1007/s10055-006-0035-5.
- [379] A. Gaitatzes, D. Christopoulos, M. Roussou, "Reviving the past: cultural heritage meets virtual reality", *Proc. of the 2001 Conference on Virtual Reality, Archaeology, and Cultural Heritage*, ACM Press, 2001, pp. 103–110.
- [380] "Tholos Theatre". <http://www.tholos254.gr/en/> (accessed Nov. 14, 2020).
- [381] "The promise of augmented reality". *The Economist*. <https://www.economist.com/science-and-technology/2017/02/04/the-promise-of-augmented-reality> (accessed Nov. 14, 2020).

- [382] <https://www.soluis.com/> (accessed Nov. 14, 2020).
- [383] <https://www.shakespeare.org.uk/visit/shakespeares-new-place/shakespeare-xr/> (accessed Nov. 14, 2020).
- [384] "Statue of David".
<https://sketchfab.com/3d-models/david-f18c62d53bf6470888465db52614c8a0> (accessed Nov. 14, 2020).
- [385] "Parthenon Gallery". <https://vgl.ict.usc.edu/Data/ParthenonGallery/> (accessed Nov. 14, 2020).
- [386] "Virtual Angkor". <https://www.virtualangkor.com/> (accessed Nov. 14, 2020).
- [387] N. Shiode, W. Grajetzki, "A virtual exploration of the lost labyrinth: developing a reconstructive model of Hawara Labyrinth pyramid complex." Centre for Advanced Spatial Analysis (CASA), University College London, paper 29, Dec. 2000.
- [388] S. T. Refsland, T. Ojika, A. C. Addison and R. Stone, "Virtual Heritage: Breathing new life into our ancient past," in IEEE MultiMedia, vol. 7, no. 2, pp. 20-21, April-June 2000, doi: 10.1109/MMUL.2000.848420.

5.10 Conclusion

The section about XR applications focused on the main domains where XR tends to be a promising technology with significant potential of growth. In this revised version of the report, the list of application domains has been completed.

6 Standards

Various Standards Developing Organisations (SDO) are directly addressing AR specific standards, and others are focusing on technology related to AR. We present in this section the standardisation activities in the XR domain.

6.1 XR specific standards

This section describes existing technical specifications published by various SDOs which directly address XR specific standards.

6.1.1 ETSI

ETSI has created an Industry Specification Group called Augmented Reality Framework (ISG ARF) [389] aiming at defining a framework for the interoperability of Augmented Reality components, systems and services, which identifies components and interfaces required for AR solutions. Augmented Reality (AR) is the ability to mix in real-time spatially-registered digital content with the real world surrounding the user. The development of a modular architecture will allow components from different providers to interoperate through the defined interfaces. Transparent and reliable interworking between different AR components is key to the successful roll-out and wide adoption of AR applications and services. This framework originally focusing on augmented reality is also well suited to XR applications. It covers all functions required for an XR system, from the capture of the real world, the analysis of the real world, the storage of a representation of the real world (related to ARCloud), the preparation of the assets which will be visualised in immersion, the authoring of XR applications, the real-time XR scene management, the user interactions, the rendering and the restitution to the user.

ISG ARF has published two Group Reports and a Group Specification:

- ETSI GR_ARF001 v1.1.1 published in April 2019 [390], provides an overview of the AR standards landscape and identifies the role of existing standards relevant to AR from various standards setting organisations. Some of the reviewed standards are directly addressing AR as a whole, and others are addressing key technological components that can be useful to increase interoperability of AR solutions;
- ETSI GR_ARF002 v1.1.1 published in August 2019 [391], outlines four categories of industrial use cases identified via an online survey - these are inspection/quality assurance, maintenance, training and manufacturing - and provides valuable information about the usage conditions of AR technologies. A description of real-life examples is provided for each category of use cases highlighting the benefits in using AR.
- ETSI GS_ARF003 v1.1.1 published in March 2020 [392] defines the architecture of a framework for augmented reality solutions. The specification introduces the characteristics of an AR system, defines a functional reference architecture and describes the functional building blocks and the relationships between these blocks. The generic nature of the architecture was validated by mapping the workflow of several use cases to the components of this framework architecture. The scope of the ISG is AR but the AR interoperability framework should overall be applicable to XR components and systems.

6.1.2 Khronos

OpenXR™ [393] defines two levels of API interfaces that a VR platform's runtime can use to access the OpenXR™ ecosystem. Applications and engines use standardised interfaces to interrogate and drive devices. Devices can self-integrate to a standardised driver interface. Standardised hardware/software interfaces reduce fragmentation while leaving implementation details open to encourage industry innovation. For areas that are still under active development, OpenXR™ also supports extensions to allow for the ecosystem to grow to fulfil the evolution happening in the industry.

The OpenXR™ working group aims to provide the industry with a cross-platform standard for the creation of VR/AR applications. This standard would abstract the VR/AR device capabilities (display, haptics, motion, buttons, poses, etc.) in order to let developers access them without worrying about which current hardware is used. In that way, an application developed with OpenXR™ would be compatible with several hardware platforms. OpenXR™ aims to integrate the critical performance concepts to enable developers to optimise for a single and predictable target instead of multiple proprietary platforms. OpenXR™ focuses on the software and hardware currently available and does not try to predict the future innovation of AR and VR technologies. However, its architecture is flexible enough to support such innovations in a close future.

6.1.3 Open ARCloud

The Open ARCloud [394] is an association created in 2019 intending to build reference implementations of the core pieces of an open and interoperable spatial computing platform for the real world to achieve the vision of what many refer to as the “Mirror World” or the “Spatial Web”. The association has started a reference open Spatial Computing platform (OSCP) with three core functions: Geopose to provide the capability to obtain, record, share and communicate geospatial position and orientation of any real or virtual objects; a locally shared machine readable world which provides users and machines with a powerful new way to interact with reality through the standardised encoding of geometry, semantics, properties, and relationships; and finally an access to everything in the digital world nearby through a local listing of references in a “Spatial Discovery Service”.

6.1.4 MPEG

MPEG is a Standard Developing Organisation (SDO) addressing media compression and transmission. MPEG is well known for its sets of standards addressing video and audio content, but other standards are now available and are more specifically addressing XR technologies.

Firstly, the Mixed and Augmented Reality Reference Model international standard (ISO/IEC 18039) [395] is a technical report defining the scope and key concepts of mixed and augmented reality, the relevant terms and their definitions, and a generalised system architecture that together serve as a reference model for Mixed and Augmented Reality (MAR) applications, components, systems, services, and specifications. This reference model establishes the set of required modules and their minimum functions, the associated information content, and the information models that have to be provided and/or supported to claim compliance with MAR systems.

Secondly, the Augmented Reality Application Format (ISO/IEC 23000-13) [396] focuses on the data format used to provide an augmented reality presentation and not on the client or server procedures. ARAF specifies scene description elements for representing AR content, mechanisms to connect to local and

remote sensors and actuators, mechanisms to integrate compressed media (image, audio, video, and graphics), mechanisms to connect to remote resources such as maps and compressed media.

6.1.5 Open Geospatial Consortium

OGC has published an “Augmented Reality Markup Language” (ARML 2.0) [397] which is an XML-based data format. Initially, ARML 1.0 was a working document extending a subset of KML (Keyhole Mark-up Language) to allow richer augmentation for location-based AR services. While ARML uses only a subset of KML, KARML (Keyhole Augmented Reality Mark-up Language) uses the complete KML format. KARML tried to extend even more KML, offering more control over the visualisation. By adding new AR-related elements, KARML deviated a lot from the original KML specifications. ARML 2.0 combined features from ARML 1.0 and KARML and has been released as an official OGC Candidate Standard in 2012 and approved as a public standard in 2015. While ARML 2.0 does not explicitly rule out audio or haptic AR, its defined purpose is to deal only with mobile visual AR.

6.1.6 W3C

The W3C has published the WebXR Device API [398] which provides access to input and output capabilities commonly associated with Virtual Reality (VR) and Augmented Reality (AR) hardware, including sensors and head-mounted displays, on the Web. By using this API, it is possible to create Virtual Reality and Augmented Reality web sites that can be viewed with the appropriate hardware like a VR headset or AR-enabled phone. Use cases can be games, but also 360 and 3D videos and object, and data visualisation. A new revision of the working draft was published in July 2020.

6.1.7 References

- [389] ETSI. <https://www.etsi.org/committee/arf> (accessed Nov. 12, 2020).
- [390] “Augmented Reality Framework (ARF); AR standards landscape.” ETSI. https://www.etsi.org/deliver/etsi_gr/ARF/001_099/001/01.01.01_60/gr_ARF001v010101p.pdf (accessed Nov. 12, 2020).
- [391] “Augmented Reality Framework (ARF) Industrial use cases for AR applications and services.” ETSI. https://www.etsi.org/deliver/etsi_gr/ARF/001_099/002/01.01.01_60/gr_ARF002v010101p.pdf (accessed Nov. 12, 2020).
- [392] “Augmented Reality Framework (ARF) AR framework architecture.” ETSI. https://www.etsi.org/deliver/etsi_gs/ARF/001_099/003/01.01.01_60/gs_ARF003v010101p.pdf (accessed Nov. 12, 2020).
- [393] Khronos Group. <https://www.khronos.org/openxr> (accessed Nov. 12, 2020).
- [394] Open AR Cloud. <https://www.openarcloud.org/> (accessed Nov. 12, 2020).
- [395] “Information technology - Computer graphics, image processing and environmental data representation - Mixed and augmented reality (MAR) reference model.” Iso. <https://www.iso.org/standard/30824.html> (accessed Nov. 12, 2020).
- [396] “Information technology - Multimedia application format (MPEG-A) — Part 13: Augmented reality application format.” ISO. <https://www.iso.org/standard/69465.html> (accessed Nov. 12, 2020).

[397] “OGC® Augmented Reality Markup Language 2.0 (ARML 2.0).” OGC.
<https://www.ogc.org/standards/arml> (accessed Nov. 12, 2020).

[398] W3C. <https://www.w3.org/blog/tags/webxr/> (accessed Nov. 12, 2020).

6.2 XR related standards

6.2.1 Khronos

OpenVX™ [399] is an open-royalty-free standard for cross platform acceleration of computer vision applications. OpenVX™ enables performance and power-optimised computer vision processing, especially important in embedded and real-time use cases such as face, body and gesture tracking, smart video surveillance, advanced driver assistance systems (ADAS), object and scene reconstruction, augmented reality, visual inspection, robotics and more. OpenVX™ provides developers with a unique interface to design vision pipelines, whether they are embedded on desktop machines, on mobile terminals or distributed on servers. These pipelines are expressed thanks to an OpenVX™ graph connecting computer vision functions, called "Nodes", implementations of abstract representations called Kernel. These nodes can be coded in any language and optimised on any hardware as long as they are compliant with OpenVX™ interface. Also, OpenVX™ provides developers with more than 60 vision operations interfaces (Gaussian image pyramid, Histogram, Optical flow, Harris corners, etc.) as well as conditional node execution and neural network acceleration.

OpenGL™ specification [400] describes an abstract API for drawing 2D and 3D graphics. Although it is possible for the API to be implemented entirely in software, it is designed to be implemented mostly or entirely in hardware. OpenGL™ is the premier environment for developing portable, interactive 2D and 3D graphics applications. Since its introduction in 1992, OpenGL™ has become widely used in the industry and supports 2D and 3D graphics application programming interface (API), bringing thousands of applications to a wide variety of computer platforms. OpenGL™ fosters innovation and speeds application development by incorporating a broad set of rendering, texture mapping, special effects, and other powerful visualisation functions. Developers can leverage the power of OpenGL™ across all popular desktop and workstation platforms, ensuring wide application deployment.

WebGL™ [401] is a cross-platform, royalty-free web standard for a low-level 3D graphics API based on OpenGL™ ES, exposed to ECMAScript via the HTML5 Canvas element. Developers familiar with OpenGL™ ES 2.0 will recognise WebGL™ as a Shader-based API, with constructs that are semantically similar to those of the underlying OpenGL™ ES API. It stays very close to the OpenGL™ ES specification, with some concessions made for what developers expect out of memory-managed languages such as JavaScript. WebGL™ 1.0 exposes the OpenGL™ ES 2.0 feature set; WebGL™ 2.0 exposes the OpenGL ES 3.0 API.

glTF™ (GL Transmission Format) [402] is a royalty-free asset delivery format for the efficient transmission and loading of 3D scenes and models by applications using the JSON standard. The format targets maximum interoperability and efficiency by minimizing the size of the 3D assets and the runtime processing needed to unpack and use those assets. glTF™ defines a common publishing format for 3D content tools and is already supported by many open-source WebGL™ engines like Three.js [403]. glTF™ 2.0, published in 2017, defines an extensibility mechanism and supports extensions such as streaming compressed geometry (mesh) data.

6.2.2 MPEG

MPEG-I (ISO/IEC 23090) [404] is dedicated to the compression of immersive content. It is structured according to the following parts: Immersive Media Architectures, Omnidirectional Media Format, Versatile Video Coding, Immersive Audio Coding, Point Cloud Compression, Immersive Media Metrics, and Immersive Media Metadata.

MPEG-V (ISO/IEC 23005) [405] provides an architecture and specifies associated information representations to enable the interoperability between virtual worlds, e.g., digital content providers of a virtual world, (serious) gaming, simulation, and with the real world, e.g., sensors, actuators, vision and rendering, robotics. Thus, this standard address many components of an XR framework, such as the sensory information, the virtual world object characteristics, the data format for interaction, etc.

MPEG-4 part 25 (ISO/IEC 14496-25) [406] is related to the compression of 3D graphics primitives such as geometry, appearance models, animation parameters, as well as the representation, coding and spatial-temporal composition of synthetic objects.

MPEG-7 part 13 Compression Descriptors for Visual Search [407] is dedicated to high performance and low complexity compact descriptors very useful for spatial computing.

MPEG-U Advanced User Interaction (AUI) interface (ISO/IEC 23007) [408] aims to support various advanced user interaction devices. The AUI interface is part of the bridge between scene descriptions and system resources. A scene description is a self-contained living entity composed of video, audio, 2D graphics objects, and animations. Through the AUI interfaces or other existing interfaces such as DOM events, a scene description accesses system resources of interest to interact with users. In general, a scene composition is conducted by a third party and remotely deployed. Advanced user interaction devices such as motion sensors and multi touch interfaces generate the physical sensed information from user's environment.

6.2.3 3GPP

3GPP SA WG4 (SA4) addresses the media distribution and codec aspects such as streaming and conversational services.

Within Release 15, 3GPP SA WG4 (SA4) published a technical specification TS 26.118 [409] on streaming of VR content. TS 26.118 defines a set of operating points covering a large range of device capabilities and media profiles mapping operating points to Dynamic Adaptive Streaming over HTTP (DASH) delivery. TS 26.118 also defines an end-to-end architecture and reference client architectures for VR streaming services as well as system metadata that supports rendering of audiovisual VR content on HMDs and 2D screens.

Within Release 16, SA4 published a technical report TR 26.928 [410] that collects information on XR in the context of 5G radio and network services. TR 26.928 includes a classification of different XR use cases and device types, identifies client and network architectures that support XR use cases and describes the integration of XR applications into the 5G system architecture.

6.2.4 Open Geospatial Consortium

OGC GML [411] serves as a modelling language for geographic systems as well as an open interchange format for geographic transactions on the Internet. GML is mainly used for geographical data interchange,

for example by Web Feature Service (WFS). WFS is a standard interface that allow exchanging geographical features between servers or between clients and servers. WFS helps to query geographical features, whereas Web Map Service is used to query map images from portals.

OGC CityGML [412] is data model and exchange format to store digital 3D models of cities and landscapes. It defines ways to describe most of the common 3D features and objects found in cities (such as buildings, roads, rivers, bridges, vegetation and city furniture) and the relationships between them. It also defines different standard levels of detail (LoDs) for the 3D objects. LoD 4 aims to represent building interior spaces.

OGC IndoorGML [413] specifies an open data model and XML schema for indoor spatial information. It represents and allows for exchange of geo-information that is required to build and operate indoor navigation systems. The targeted applications are indoor robots, indoor localisation, indoor m-Commerce, emergency control, etc. IndoorGML does not provide spaces geometry but it can refer to data described in other formats like CityGML, KML or IFC.

OGC KML [414] is an XML language focused on geographic visualisation, including annotation of maps and images. Geographic visualisation includes not only the presentation of graphical data on the globe, but also the control of the user's navigation in the sense of where to go and where to look. KML became an OGC standard in 2015 and some functionalities are duplicated between KML and traditional OGC standards.

6.2.5 W3C

GeoLocation API [415] is a standardised interface to be used to retrieve the geographical location information from a client-side device. The location accuracy depends of the best available location information source (global position systems, radio protocols, Mobile network location or IP address location). Web pages can use the Geolocation API directly if the web browser implements it. It is supported by most desktop and mobile operating systems and by most web browsers. The API returns 4 location properties; latitude and longitude (coordinates), altitude (height), and accuracy.

6.2.6 References

- [399] Khronos Group. <https://www.khronos.org/opencv/> (accessed Nov. 12, 2020).
- [400] Khronos Group. <https://www.khronos.org/opengl/> (accessed Nov. 12, 2020).
- [401] Khronos Group. <https://www.khronos.org/webgl/> (accessed Nov. 12, 2020).
- [402] Khronos Group. <https://www.khronos.org/gltf/> (accessed Nov. 12, 2020).
- [403] Threejs. <https://threejs.org/> (accessed Nov. 12, 2020).
- [404] MPEG-I. <https://mpeg.chiariglione.org/standards/mpeg-i> (accessed Nov. 12, 2020).
- [405] MPEG-V. <https://mpeg.chiariglione.org/standards/mpeg-v> (accessed Nov. 12, 2020).
- [406] MPEG Graphics Compression Model. <https://mpeg.chiariglione.org/standards/mpeg-4/3d-graphics-compression-model> (accessed Nov. 12, 2020).
- [407] MPEG Compact Descriptors for Visual Search. <https://mpeg.chiariglione.org/standards/mpeg-7/compact-descriptors-visual-search> (accessed Nov. 12, 2020).

- [408] MPEG-U Rich Media User Interface. <https://mpeg.chiariglione.org/standards/mpeg-u> (accessed Nov. 16, 2020).
- [409] 3GPP TS 26.118: "3GPP Virtual reality profiles for streaming applications".
- [410] 3GPP TR 26.928: "Extended Reality (XR) in 5G".
- [411] "Geography Markup Language." OGC. <https://www.opengeospatial.org/standards/gml> (accessed Nov. 12, 2020).
- [412] "CityGML." OGC. <https://www.opengeospatial.org/standards/citygml> (accessed Nov. 12, 2020).
- [413] "IndoorGML SWG." OGC. <https://www.opengeospatial.org/projects/groups/indoorgmlswg> (accessed Nov. 12, 2020).
- [414] "KML." OGC. <https://www.opengeospatial.org/standards/kml> (accessed Nov. 12, 2020).
- [415] "Geolocation API Specification 2nd Edition." W3C. <https://www.w3.org/TR/geolocation-API/> (accessed Nov. 12, 2020).

7 Review of current EC research

EC-funded research covers a wide range of areas within fundamental research on VR/AR/MR as well as applications and technology development. The analysis covers all the relevant projects funded by the EC, which have an end date not later than January 2016.

A large number of projects develop or use VR and AR tools for the cultural heritage sector (**4D-CH-WORLD**, **DigiArt**, **eHeritage**, **EMOTIVE**, **GIFT**, **GRAVITATE**, **i-MareCulture**, **INCEPTION**, **ITN-DCH**, **Scan4Reco**, **ViMM**), in part due to a dedicated call on virtual museums (**CULT-COOP-08-2016**). Among the projects funded under this programme, **i-MareCulture** aims to bring publicly unreachable underwater cultural heritage within digital reach by implementing virtual visits, and serious games with immersive technologies and underwater AR. The project **Scan4Reco** advanced methods to preserve cultural assets by performing analysis and aging predictions on their digital replicas. The project also launched a virtual museum that contains the cultural assets studied during the project. As a heritage-related project, **MEMEX** targets inclusion of socially fragile communities, with AR storytelling as part of the toolset. With the goal to foster mutual understanding between refugees and local communities, **SO-CLOSE** plans the development of a Memory Center interactive platform including immersive content.

VR/AR/MR technologies enrich media and content production in entertainment (**ACTION-TV**, **DBRLive**, **first.stage**, **ImmersiaTV**, **Immersify**, **INVICTUS**, **POPART**, **SAUCE**, **VISUALMEDIA**, **VRACE**, among others). As an example, **Immersify** aims, e.g. to develop advanced video compression technology for VR video, to provide media players and format conversion tools, and to create and promote new immersive content and tools.

In projects focusing on social and work-related interaction (**AlterEgo**, **CO3**, **CROSS DRIVE**, **I.MOVE.U**, **INTERACT**, **IRIS**, **PRESENT**, **REPLICATE**, **VRTogether**), research concentrates on the improvement of technologies or generation of platforms that facilitate usage of XR technologies. For example, **REPLICATE** employed emerging mobile devices for the development of an intuitive platform to create real-world-derived digital assets for enhancement of creative processes through the integration of Mixed Reality user experiences. **CROSS DRIVE** targeted the space sector in creating a shared VR workplace for collaborative data analysis as well as mission planning and operation.

Strong fields of research and application appear in education and training (**ARETE**, **AUGGMED**, **ASSISTANCE**, **CybSPEED**, **E2DRIVER**, **LAW-TRAIN**, **NEWTON**, **REVEAL**, **TARGET**, **WEKIT**, **WhoLoDancE**, among others). For example, in the **TARGET** project, a serious gaming platform was developed to expand the training options for security critical agents, and, in the **LAW-TRAIN** project, a mixed-reality platform was established to train law enforcement agents in criminal investigation. Several projects within the health sector also target education such as **CAPTAIN**, **HOLOBALANCE**, **SurgASSIST**, **UpSurgeOn Academy**. Furthermore, the above-listed projects related to heritage generally also have an educational component.

The health sector can be roughly divided in three categories:

1. A strong focus is placed on improving conditions for the aging population and those with impairments of any kind (**AbleGames**, **AlterEgo**, **CAPTAIN**, **HOLOBALANCE**, **KINOPTIM**, **MetAction**, **OACTIVE**, **PLUTO**, **PRIME-VR2**, **RAMCIP**, **See Far**, **Sound of Vision**, **WorkingAge**). **PRIME-VR2** for example aims at the development of an accessible collaborative VR environment for rehabilitation. Through the integration of AR in smart glasses, **See Far** targets the mitigation of age-related vision loss;

2. Several projects lie in the surgical field: **EndoMapper, RASimAs, SMARTsurg, SurgASSIST, UpSurgeOn Academy, VOSTARS**. In the **VOSTARS** project, a hybrid video optical see-through AR head-mounted display is being developed for surgical navigation;
3. Another focus is placed on mental health (**CCFIB, VIRTUALTIMES, VRMIND**). Within **VIRTUALTIMES** for example, a personalised and neuroadaptive VR tool is developed for diagnosis of psychopathological symptoms.

Outside of these categories, **PIDS** identifies nutrition interventions to improve population health and uses XR to study dietary choices based on social status. **SOCRATES** develops a platform for obesity treatment. Several projects target or relate to the design and engineering fields (**CARBODIN, DIMMER, EASY-IMP, FURNIT-SAVER, HyperCOG, MANUWORK, MINDSPACES, OPTINT, RECLAIM, SPARK, ToyLabs, TRINITY, V4Design**, among others). **ToyLabs** for example developed a platform for product improvement through various means, among them the use of AR technologies to include customer feedback.

In the sectors of maintenance, construction and renovation, projects predominantly use AR technologies: **Artwin, BIM4EEB, BugWright2, EDUSAFE, ENCORE, INSITER, PACMAN, PreCoM, PROPHECY**. With **INSITER**, AR with access to a digitised database is used in construction to enable the design and construction of energy-efficient buildings. In comparing what is built against the building information model (BIM), the mismatch of energy performance between the design and construction phases of a building can be reduced.

Projects such as **AEROGLOSS, ALADDIN, ALLEGRO, AssAssINN, CARBODIN, E2DRIVER, I-VISION, RETINA, SimuSafe, SUaaVE, ViAJeRo, VISTA, WrightBroS** can be classified to contribute to the transportation and vehicles sector. Within **AEROGLOSS**, AR was used to support pilots in aerial navigation using head-mounted displays. **E2DRIVER** will develop a training platform for the automotive industry targeting increase of energy efficiency. **VISTA** is part of the **ACCLAIM** cluster funded by the European **Clean Sky** programme. The project **ACCLAIM** targets improvement in the assembly of aircraft cabin and cargo elements by developing, e.g. VR for assembly planning and an AR process environment. **VISTA** handles post-assembly inspections using suitable AR interfaces for the human operator.

A number of projects are dedicated to fight crime (**ALADDIN, CONNEXIONS, CRIMETIME, Infinity, RISEN**). By combining AR/VR-technologies with artificial intelligence, **Infinity** strives to build a solution for data-driven investigations. XR is also applied to aid in disaster management and first responder support (**CENTAURO, HyResponder, INGENIOUS, RESPOND-A, TERRIFFIC, xR4DRAMA**).

Technology to support research questions is used in projects such as **COGNIBRAINS, EMERG-ANT, FLYVISUALCIRCUITS, IN-Fo-trace-DG, NEUROBAT, NeuroVisEco, NEWRON, SOCIAL LIFE, Vision-In-Flight**, which investigate animal-environment interactions, memory formation, insect navigation, and animal vision. The outcome of the latter focus can, in turn, be expected to provide insights to improve computer-vision and machine-vision technologies and algorithms. Fundamental research on brain response and behaviour is also expected to take another leap through the use of VR/AR technologies (**ActionContraThreat, eHonesty, EVENTS, HOMEOSTASIS, MESA, METAWARE, NEUROMEM, NewSense, PLATYPUS, RECONTEXT, SELF-UNITY, Set-to-change, SpaceCog, TRANSMEM**). As part of the **Human Brain Project**, the **Neurobotics** platform allows one to test simulated brain models with real or virtual robot bodies in a virtual environment.

Several projects specifically focus on technology progress in the areas of:

Wearables: **Digital Iris, EXTEND, HIDO, LIGHTFIELD, NGEAR3D, REALITY, See Far, WEAR3D;**
 Displays: **ETN-FPI, HoviTron, LOMID;**
 Sound: **BINCI, Sound of Vision, SoundParticles, SOUNDS, VRACE;**
 Haptics: **DyViTo, H-Reality, MULTITOUCH, ph-coding, TACTILITY, TouchDesign;**
 Camera development: **DBRLive, FASTFACEREC, PERCOSDECAM;**
 Computer graphics, animation, and related fields: **ANIMETRICS, FunGraph, RealHands, REALITY CG, VirtualGrasp.**

Another large area of development concerns human-robot interaction using XR (**CogIMon, CONBOTS, FACTORY-IN-A-DAY, LIAA, RAMCIP, SoftPro, SYMBIO-TIC, TRAVERSE**). As an example, **RAMCIP** developed a robot assistant supporting the elderly as well as patients with mild cognitive impairments and early Alzheimer's disease at home. Patient-robot communication technology includes an AR display and an underlying empathic communication channel together with touch-screen, speech, and gesture recognition.

Other notable projects not placed in one of the categories above include **iv4XR**, which, in combination with artificial intelligence methods aims to build a novel verification and validation technology for XR systems. Within **ImAc**, the focus is on the accessibility of services accompanying the design, production and delivery of immersive content.

7.1 References

4D-CH-WORLD: <https://cordis.europa.eu/project/rcn/106777/reporting/en>
 AbleGames: <https://cordis.europa.eu/project/id/958637>
 ACCLAIM: <https://www.projectacclaim.eu/>
 ActionContraThreat: <https://cordis.europa.eu/project/rcn/220842/factsheet/en>
 ACTION-TV: <https://cordis.europa.eu/project/rcn/191629/factsheet/en>
 AEROGLOSS: <https://glass.aero/>
 ALADDIN: <https://aladdin2020.eu/>
 ALLEGRO: <http://www.allegro-erc.nl/>
 AlterEgo: <https://cordis.europa.eu/project/id/600610>
 ANIMETRICS: <https://cordis.europa.eu/project/rcn/100399/reporting/en>
 ARETE: <https://www.areteproject.eu/>
 Artwin: <https://artwin-project.eu/>
 AssAssiNN: <https://cordis.europa.eu/project/id/886977>
 ASSISTANCE: <https://cordis.europa.eu/project/rcn/222583/factsheet/en>
 AUGGMED: <https://cordis.europa.eu/project/rcn/194875/factsheet/en>
 BIM4EEB: <https://www.bim4eeb-project.eu/the-project.html>
 BINCI: <https://binci.eu/>
 BugWright2: <https://cordis.europa.eu/project/id/871260>
 CAPTAIN: <https://cordis.europa.eu/project/rcn/211575/factsheet/en>
 CARBODIN: <https://carbodin.eu/>
 CCFIB: <https://cordis.europa.eu/project/rcn/108306/reporting/en>
 CENTAURO: <http://www.centauro-project.eu/>
 Clean Sky: <https://www.cleansky.eu/>
 CO3: <https://cordis.europa.eu/project/rcn/218757/factsheet/en>
 CogIMon: <https://www.cogimon.eu/>
 COGNIBRAINS: <https://cordis.europa.eu/project/id/835032>

CONBOTS: <https://www.conbots.eu/>
 CONNEXIONS: <https://www.connexions-project.eu/>
 CRIMETIME: <https://cordis.europa.eu/project/id/772911>
 CROSS DRIVE: <https://cordis.europa.eu/project/rcn/188842/factsheet/en>
 CULT-COOP-08-2016: <https://cordis.europa.eu/programme/rcn/700239/en>
 CybSPEED: <http://www.ehu.eus/ccwintco/cybSPEED/>
 DBRLive: <https://www.supponor.com/>
 DigiArt: <http://digiart-project.eu>
 Digital Iris: <https://viewpointssystem.com/en/eu-program/>
 DIMMER: <https://cordis.europa.eu/project/rcn/110900/factsheet/en>
 DyViTo: <https://dyvito.com/>
 E2DRIVER: <https://www.clustercollaboration.eu/profile-articles/e2driver-welcome-eu-training-platform-automotive-supply>
 EASY-IMP: <https://cordis.europa.eu/project/rcn/109126/reporting/en>
 EDUSAFE: <https://cordis.europa.eu/project/rcn/105266/reporting/en>
 eHeritage: <http://www.eheritage.org/>
 eHonesty: <https://cordis.europa.eu/project/rcn/218598/factsheet/en>
 EMERG-ANT: <https://cordis.europa.eu/project/rcn/212128/factsheet/en>
 EMOTIVE: <https://emotiveproject.eu/>
 ENCORE: <https://cordis.europa.eu/project/rcn/220934/factsheet/en>
 EndoMapper: <https://cordis.europa.eu/project/id/863146>
 ETN-FPI: <http://www.full-parallax-imaging.eu>
 EVENTS: <https://cordis.europa.eu/project/rcn/220910/factsheet/en>
 EXTEND: <https://varjo.com/>
 FACTORY-IN-A-DAY: <http://www.factory-in-a-day.eu/>
 FASTFACEREC: <https://cordis.europa.eu/project/rcn/218200/factsheet/en>
 first.stage: <http://www.first-stage.eu/>
 FLYVISUALCIRCUITS: <https://cordis.europa.eu/project/rcn/100624/reporting/en>
 FunGraph: <https://cordis.europa.eu/project/rcn/217966/factsheet/en>
 FURNIT-SAVER: <https://furnit-saver.eu/>
 GIFT: <https://gifting.digital/>
 GRAVITATE: <http://gravitate-project.eu>
 HIDO: <https://cordis.europa.eu/project/rcn/111165/factsheet/en>
 HOLOBALANCE: <https://holobalance.eu/>
 HOMEOSTASIS: <https://cordis.europa.eu/project/id/863732>
 HoviTron: <https://cordis.europa.eu/project/id/951989>
 H-Reality: <https://cordis.europa.eu/project/rcn/216340/factsheet/en>
 Human Brain Project: <https://www.humanbrainproject.eu/en/>
 HyperCOG: <https://www.hypercog.eu/>
 HyResponder: <https://hyresponder.eu/>
 ImAc: <http://www.imac-project.eu/>
 iMARECULTURE: <https://imareculture.eu>
 ImmersiaTV: <http://www.immersiatv.eu>
 Immersify: <https://immersify.eu/objectives/>
 I.MOVE.U: <https://cordis.europa.eu/project/rcn/109099/factsheet/en>
 INCEPTION: <https://www.inception-project.eu/en>
 Infinity: <https://cordis.europa.eu/project/id/883293>
 IN-Fo-trace-DG: <https://cordis.europa.eu/project/rcn/214885/factsheet/en>

INGENIOUS: <https://ingenious-first-responders.eu/>
 INSITER: <https://www.insiter-project.eu/Pages/VariationRoot.aspx>
 INTERACT: <https://cordis.europa.eu/project/rcn/106707/reporting/en>
 INVICTUS : <https://cordis.europa.eu/project/id/952147>
 IRIS: <https://cordis.europa.eu/project/id/610986>
 ITN-DCH: <https://www.itn-dch.net/>
 iv4XR: <https://cordis.europa.eu/project/rcn/223956/factsheet/en>
 I-VISION: <http://www.ivation-project.eu/>
 KINOPTIM: <https://cordis.europa.eu/project/rcn/106678/reporting/en>
 LAW-TRAIN: <http://www.law-train.eu/index.html>
 LIAA: <https://cordis.europa.eu/project/id/608604>
 LIGHTFIELD: <https://cordis.europa.eu/project/id/945983>
 LOMID: <http://www.lomid.eu/>
 MANUWORK: <http://www.manuwork.eu/>
 MEMEX: <https://memexproject.eu/en/home>
 MESA: <https://cordis.europa.eu/project/rcn/192349/factsheet/en>
 MetAction: <https://cordis.europa.eu/project/rcn/220806/factsheet/en>
 METAWARE: <https://cordis.europa.eu/project/rcn/199667/factsheet/en>
 MINDSPACES: <http://mindspaces.eu/>
 MULTITOUCH: <https://cordis.europa.eu/project/rcn/224423/factsheet/en>
 NEUROBAT: <https://cordis.europa.eu/project/rcn/100861/reporting/en>
 Neurobotics: <https://www.humanbrainproject.eu/en/robots/>, <https://neurobotics.net>
 NEUROMEM: <https://cordis.europa.eu/project/rcn/205205/factsheet/en>
 NeuroVisEco: <https://cordis.europa.eu/project/rcn/202546/factsheet/en>
 NEWRON: <https://cordis.europa.eu/project/rcn/204707/factsheet/en>
 NewSense: <https://cordis.europa.eu/project/rcn/221453/factsheet/en>
 NEWTON: <http://www.newtonproject.eu/>
 NGEAR 3D: <https://cordis.europa.eu/project/id/960828>
 OACTIVE: <https://www.oactive.eu/>
 OPTINT: <https://cordis.europa.eu/project/rcn/206993/factsheet/en>
 PACMAN: <https://cordis.europa.eu/project/rcn/205837/factsheet/en>
 PERCOSDECAM: <https://cordis.europa.eu/project/id/878873>
 ph-coding: <https://www.ph-coding.eu/>
 PIDS: <https://cordis.europa.eu/project/id/803194>
 PLATYPUS: <https://platypus-rise.eu/>
 PLUTO: <https://cordis.europa.eu/project/id/879510>
 POPART: <https://alicevision.org/popart/>
 PreCoM: <https://www.precom-project.eu/>
 PRESENT: <https://www.upf.edu/web/present/>
 PRIME-VR2: <http://www.prime-vr2.eu/>
 PROPHECY: <https://prophecy.eu/overview>
 RAMCIP: <https://ramcip-project.eu/>
 RASimAS: <https://cordis.europa.eu/project/rcn/111086/factsheet/en>
 RealHands: <https://cordis.europa.eu/project/rcn/223614/factsheet/en>
 REALITY: <https://www.brighterwave.com/>
 REALITY CG: <https://cordis.europa.eu/project/rcn/97093/reporting/en>
 RECLAIM: <https://www.reclaim-project.eu/>
 RECONTEXT: <https://cordis.europa.eu/project/rcn/98096/factsheet/en>

REPLICATE: <http://www.replicate3d.eu/>
 RESPOND-A: <https://respond-a-project.eu/>
 RETINA: <http://www.retina-atm.eu/>
 REVEAL: <https://cordis.europa.eu/project/id/732599>
 RISEN: <https://www.risen-h2020.eu/>
 SAUCE: <https://www.sauceproject.eu/>
 Scan4Reco: <https://scan4reco.it/it/gr/>
 See Far: <https://cordis.europa.eu/project/rcn/219052/factsheet/en>
 SELF-UNITY: <https://cordis.europa.eu/project/rcn/216574/factsheet/en>
 Set-to-change: <https://cordis.europa.eu/project/rcn/217958/factsheet/en>
 SimuSafe: <http://simusafe.eu/>
 SMARTsurg: <http://www.smartsurg-project.eu/>
 SO-CLOSE: <https://so-close.eu/>
 SOCIAL LIFE: <https://cordis.europa.eu/project/rcn/94363/factsheet/en>
 SOCRATES: <https://cordis.europa.eu/project/id/951930>
 SoftPro: <https://www.softpro.eu/>
 Sound of Vision: <http://www.soundofvision.net/>
 SoundParticles: <https://cordis.europa.eu/project/rcn/224745/factsheet/en>
 SOUNDS: <https://cordis.europa.eu/project/id/956369>
 SpaceCog: <https://cordis.europa.eu/project/id/600785>
 SPARK: <http://www.spark-project.net/>
 SUaAVE: <http://www.suaave.eu/>
 SurgASSIST: <https://www.incision.care/>
 SYMBIO-TIC: <http://www.symbio-tic.eu/>
 TACTILITY: <https://cordis.europa.eu/project/rcn/223946/factsheet/en>
 TARGET: <http://www.target-h2020.eu>
 TERRIFFIC: <https://www.terrific.eu/>
 ToyLabs: <http://www.toylabs.eu/>
 TRANSMEM: <https://cordis.europa.eu/project/rcn/111455/factsheet/en>
 TRAVERSE: <http://users.isr.ist.utl.pt/~aahmad/traverse/doku.php>
 TRINITY: <https://trinityrobotics.eu/>
 TouchDesign: <https://cordis.europa.eu/project/rcn/214715/factsheet/en>
 UpSurgeOn Academy: <https://cordis.europa.eu/project/id/880895>
 V4Design: <https://v4design.eu/>
 ViAJeRo: <https://viajero-project.org/>
 ViMM: <http://www.vi-mm.eu/>
 VirtualGrasp: <https://cordis.europa.eu/project/rcn/217198/factsheet/en>
 VIRTUALTIMES: <http://hci.uni-wuerzburg.de/projects/virtualtimes/>, <https://www.gleechi.com/>
 Vision-In-Flight: <https://cordis.europa.eu/project/rcn/218441/factsheet/en>
 VISTA: https://www.projectacclaim.eu/?page_id=514
 VISUALMEDIA: <http://www.visualmediaproject.com>
 VOSTARS: <https://www.vostars.eu/>
 VRACE: <https://cordis.europa.eu/project/rcn/220824/factsheet/en>
 VRMIND: <https://cordis.europa.eu/project/id/733901>
 VRTogether: <https://vrtogether.eu>
 WEAR3D: <https://cordis.europa.eu/project/rcn/111025/factsheet/en>
 WEKIT: <https://wekit.eu>
 WhoLoDance: <http://www.wholodance.eu/>

WorkingAge:<https://www.workingage.eu/>

WrightBroS:<https://cordis.europa.eu/project/rcn/218206/factsheet/en>

xR4DRAMA: <https://cordis.europa.eu/project/id/952133/de>

8 Conclusion

This landscape report provides a recent and quite complete overview of the XR technologies landscape. The market analysis is based on the latest figures available as of autumn 2020. From these figures, one can foresee the huge worldwide economic potential of this technology. However, the position of Europe is quite different in several aspects, such as in investment, main players, and technology leadership. Hence, the report illustrates, where the potential of future investment lies in.

The description of XR technologies contains not only the current state-of-the-art in research and development, but it also provides terms and definitions in each area covered. Hence, this report also acts as a guide or handbook for immersive/XR and interactive technologies.

Based on thorough analysis of the XR market, the major applications are presented showing the potential of this technology. The report shows that the industry and healthcare sectors constitute a huge potential for XR. In addition, social VR or, equivalently, collaborative tele-presence, also holds a tremendous potential, including for Europe, because of its strong reliance on software and algorithms.