



***MOVING THE EUROPEAN
XR TECH INDUSTRY FORWARD***

www.xr4all.eu

b com

 **Europe
Unlimited**

 **Fraunhofer**
Heinrich Hertz Institute

stere@psia

 **LUCIDWEB**

XR4ALL

(Grant Agreement 825545)

“eXtended Reality for All”

Coordination and Support Action

D4.1: Landscape Report

Issued by: Fraunhofer HHI
 Issue date: 30/11/2019
 Due date: 30/11/2019
 Work Package Leader: Fraunhofer HHI

Start date of project: 01 December 2018
 months

Duration: 30

Document History

Version	Date	Changes
0.1	31/07/2019	Draft structure
0.2	31/08/2019	Revised structure
0.3	05/11/2019	Comments from 1 st review included
1.0	30/11/2019	Final version

Dissemination Level

PU	Public	X
PP	Restricted to other programme participants (including the EC Services)	
RE	Restricted to a group specified by the consortium (including the EC Services)	
CO	Confidential, only for members of the consortium (including the EC)	

Main authors	
Name	Organisation
Oliver Schreer	HHI
Ivanka Pelivan	HHI
Peter Kauff	HHI
Ralf Schäfer	HHI
Anna Hilsmann	HHI
Paul Chojecki	HHI
Thomas Koch	HHI
Ramon Wiegratz	HHI
Jérôme Royan	BCOM
Muriel Deschanel	BCOM
Albert Murienne	BCOM
Laurent Launay	BCOM
Jacques Verly	I3D

Quality reviewers	
Name	Organisation
Jacques Verly	I3D
Muriel Deschanel	BCOM

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.

© XR4ALL Consortium, 2019



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

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

1	INTRODUCTION	8
2	THE SCOPE OF EXTENDED REALITY.....	9
2.1	References.....	10
3	XR MARKET WATCH	11
3.1	Market development and forecast	11
3.2	Areas of application	13
3.3	Investments.....	14
3.4	Shipment of devices	15
3.5	Main players.....	18
3.6	International, European and regional associations in XR	20
3.7	Patents	23
3.8	References.....	24
4	XR TECHNOLOGIES.....	26
4.1	Video capture for XR	26
4.2	3D sound capture	29
4.3	Scene analysis and computer vision	32
4.4	3D sound processing algorithms	40
4.5	Input and output devices	43
4.6	Cloud services	52
4.7	Conclusion	53
5	XR APPLICATIONS.....	54
5.1	Advertising and commerce	54
5.2	Design, engineering, and manufacturing	57
5.3	Health and medicine	65
5.4	Journalism & weather	75
5.5	Social VR.....	75
5.6	Conclusion	82
6	STANDARDS	83
6.1	XR specific standards.....	83
6.2	XR related standards	85
7	REVIEW OF CURRENT EC RESEARCH	89
7.1	References.....	91
8	CONCLUSION.....	95

List of figures

Figure 1: Extended reality scheme.	9
Figure 2: VR/AR market forecast by Gartner and Credit Suisse [7].	11
Figure 3: Market growth rates by worldwide regions [4].	12
Figure 4: AR/VR regional revenue between 2017 and 2022 [10].	12
Figure 5: Distribution of VR/AR companies analysed in survey by Capgemini Research [13].	13
Figure 6: Separated AR and VR sector revenue from 2017 to 2022 [8].	14
Figure 7: XR4ALL analysis of European investors for start-ups [15].	15
Figure 8: VR unit shipments in the last three years [17].	16
Figure 9: Forecast of AR unit shipments from 2016 to 2022 [19].	17
Figure 10: Forecast of VR and AR shipments [20].	17
Figure 11: AR Industry Landscape by Venture Reality Fund [22].	18
Figure 12: VR Industry Landscape by Venture Reality Fund [22].	19
Figure 13: Comparison chart of VR headset resolutions [134].	46
Figure 14: Examples for different AR applications in advertising and commerce	54
Figure 15: Sample views of ZREALITY virtual show room.	55
Figure 16: The fashion eco-system.	55
Figure 17: AR application for assembly in industrial environment.	57
Figure 18: AR application for quality control in industrial environment.	59
Figure 19: AR application for maintenance in industrial environment.	60
Figure 20: AR application for planning in industrial environment.	62
Figure 21: AR application for logistics.	63
Figure 22: AR application for training in industrial environment.	64
Figure 23: Example of training and learning use of AR in the medical domain.	66
Figure 24: Example of pre-operative use of AR/VR.	67
Figure 25: Example of intra-operative use of AR/VR.	70
Figure 26: Example of intra-operative use of AR.	71
Figure 27: Mindmotion VR [175] (left) and Nirvana [176] (right).	72

Figure 28: Example for post-operative use of VR/AR.	73
Figure 29: Illustration of interaction in a virtual space, here based upon the vTime platform [184].....	77
Figure 30: Illustration of interaction in a virtual space, here based upon the Rec Room platform [199].	77

1 Introduction

This report provides a thorough analysis of the landscape of immersive interactive XR technologies carried out during summer and autumn 2019 by the members of the XR4ALL consortium. It is based on a desk research by a large number of researchers from Fraunhofer HHI, B<>com and Image & 3D Europe.

The document is organized 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.4. It consists on the development and forecast of XR technologies based on an in-depth analysis of most recent surveys and report 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 on all the relevant technologies is given that are necessary for the successful development of future immersive and interactive technologies. Latest research results and the current state-of-the-art is described completed 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 on 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 in 1994 the well-known Reality-Virtuality Continuum [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.

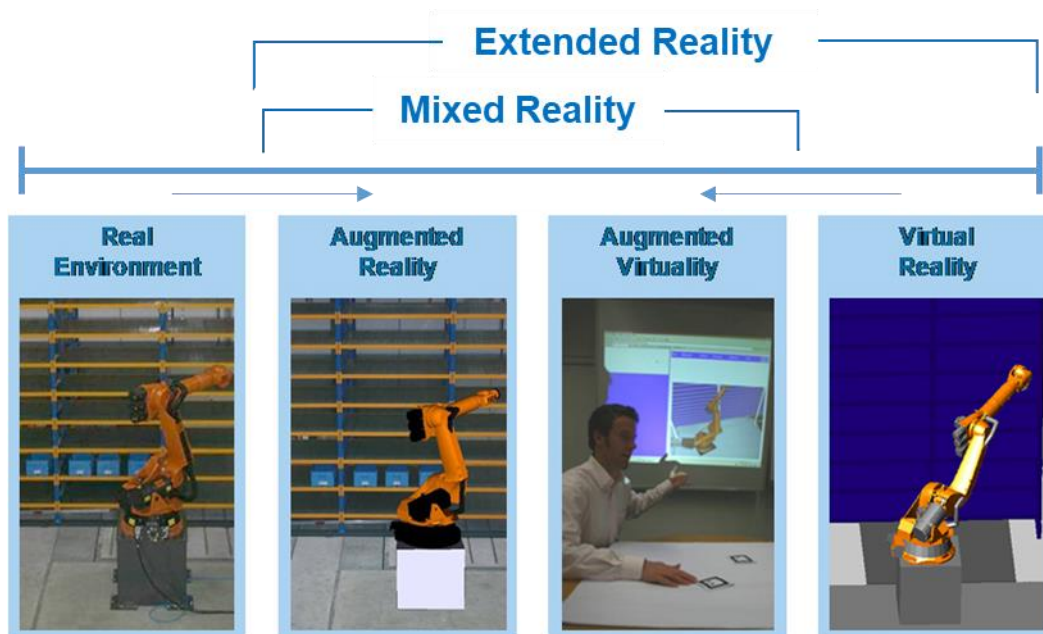


Figure 1: Extended reality scheme.

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

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.

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.

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 spans across technologies and application domains. XR could be considered a fusion of AR, AV, and VR technologies, but in fact it involves much 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. The complete set of technologies and applications will be described in the following chapters.

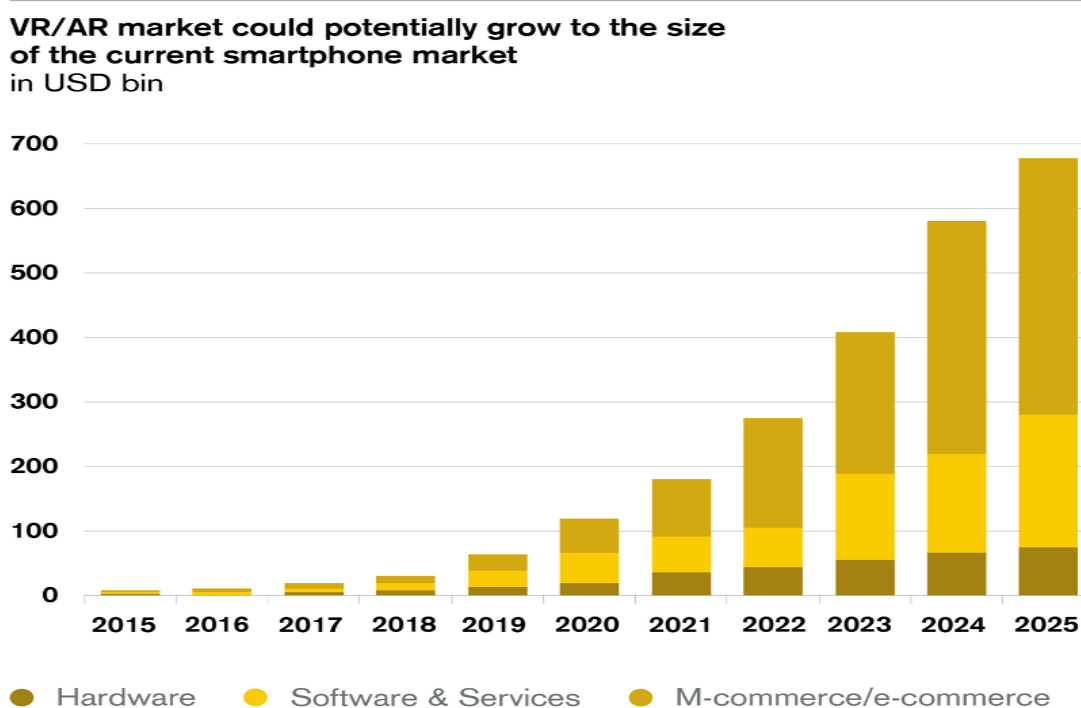
2.1 References

- [1] P. Milgram, H. Takemura, A. Utsumi, F. Kishino: "Augmented Reality: A class of displays on the reality-virtuality continuum". SPIE Vol. 2351, Proceedings of Telemanipulator and Telepresence Technologies. pp. 2351–34, 1994.
- [2] R. 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 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 85.0 billion (AR, valued at USD 4.2 billion in 2017) and USD 53.6 billion (VR, valued at USD 7.9 billion in 2018) 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 2). 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 2: VR/AR market forecast by Gartner and Credit Suisse [7].

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]. MarketsandMarkets find 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.

Extended Reality (XR) Market - Growth Rate by Region (2019-2024)

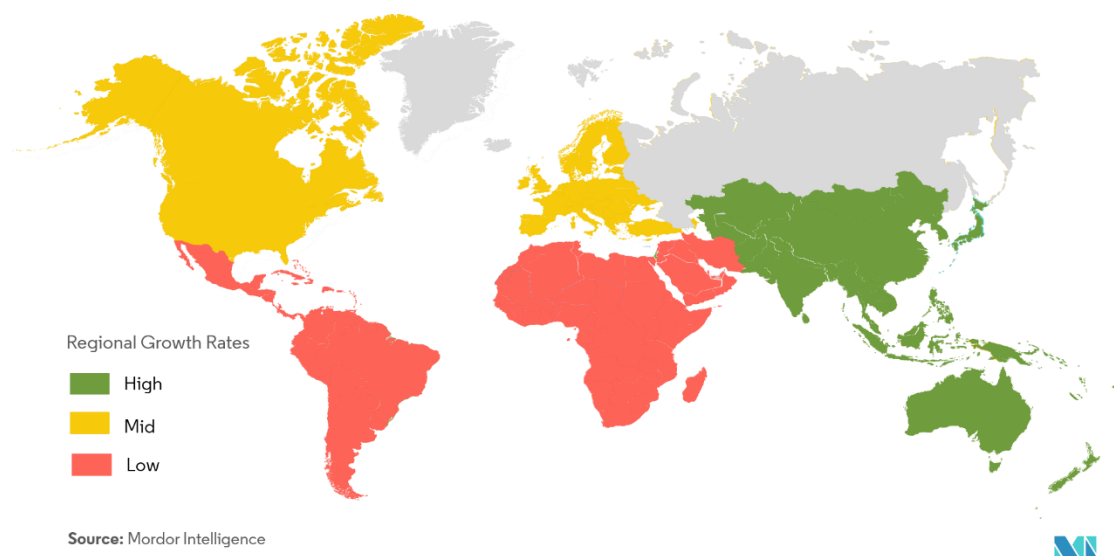
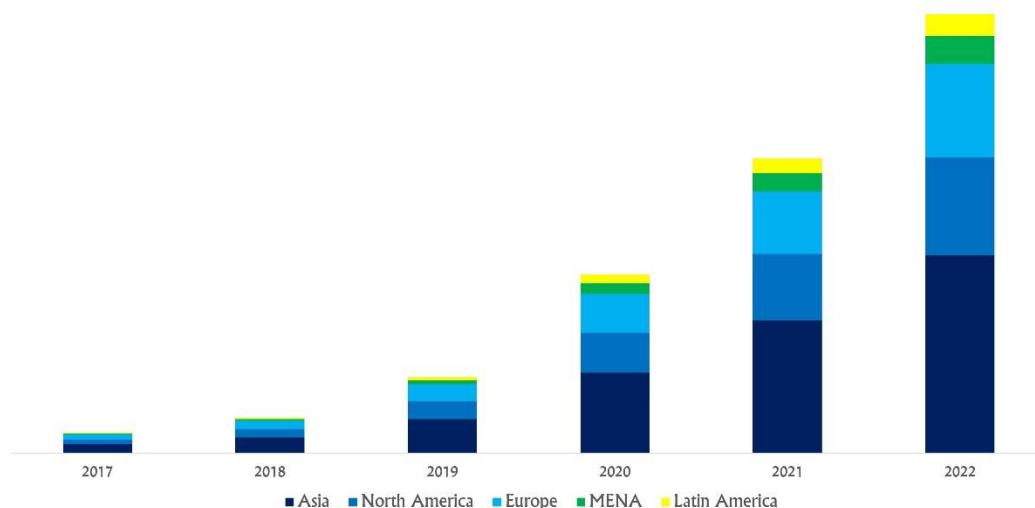


Figure 3: Market growth rates by worldwide regions [4].

AR/VR regional revenue



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

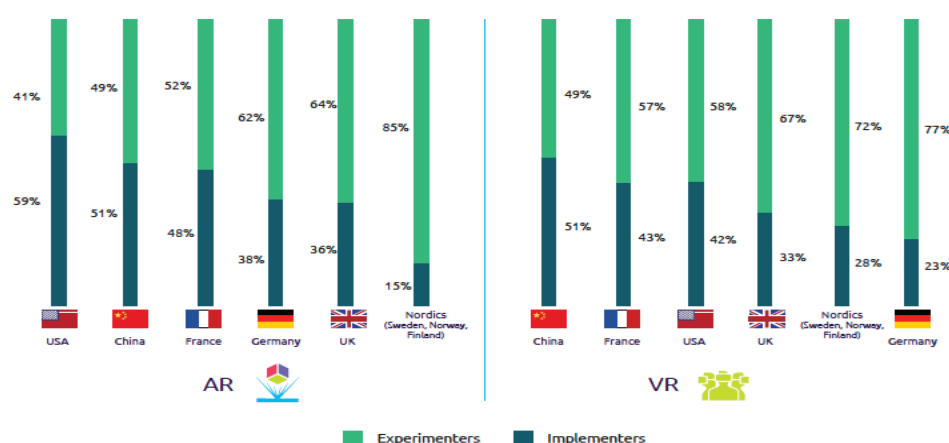
Figure 4: AR/VR regional revenue between 2017 and 2022 [10].

With the XR market growing exponentially, Europe accounts for about one fifth of the market in 2022 [10], with Asia as the leading region (mainly China, Japan, and South Korea) followed by North America and Europe at almost the same level. The enquiry in [11] 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 [12], 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.

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 has the strongest focus, closely followed by design and assembly. Other popular areas of implementation cover immersive training, and inspection and quality assurance [13]. Benefits from implementing AR/VR technologies include substantial increases in efficiency, safety, productivity, and reduction in complexity.

In a survey conducted in 2018 [13], 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 expect that AR/VR will become mainstream in their organization 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 organizations 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 5).



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 5: Distribution of VR/AR companies analysed in survey by Capgemini Research [13].

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 [13]. Highly specialized 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 6). 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].

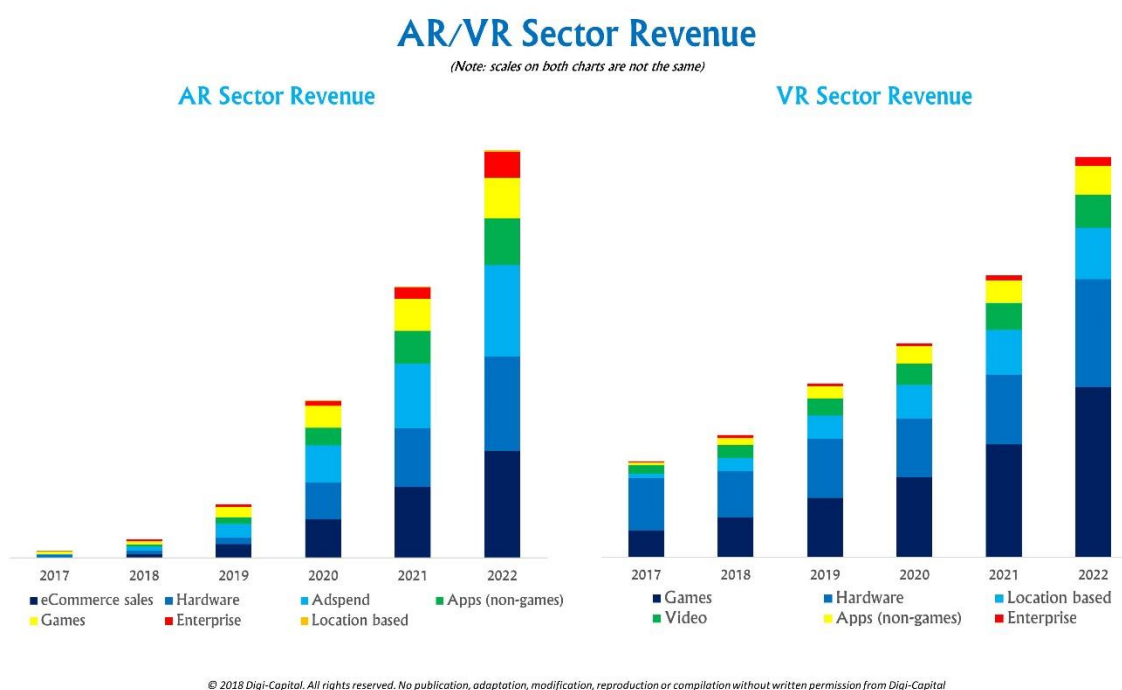


Figure 6: Separated AR and VR sector revenue from 2017 to 2022 [8].

3.3 Investments

While XR industries are characterized 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 [14]. 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 [15]. 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 7).

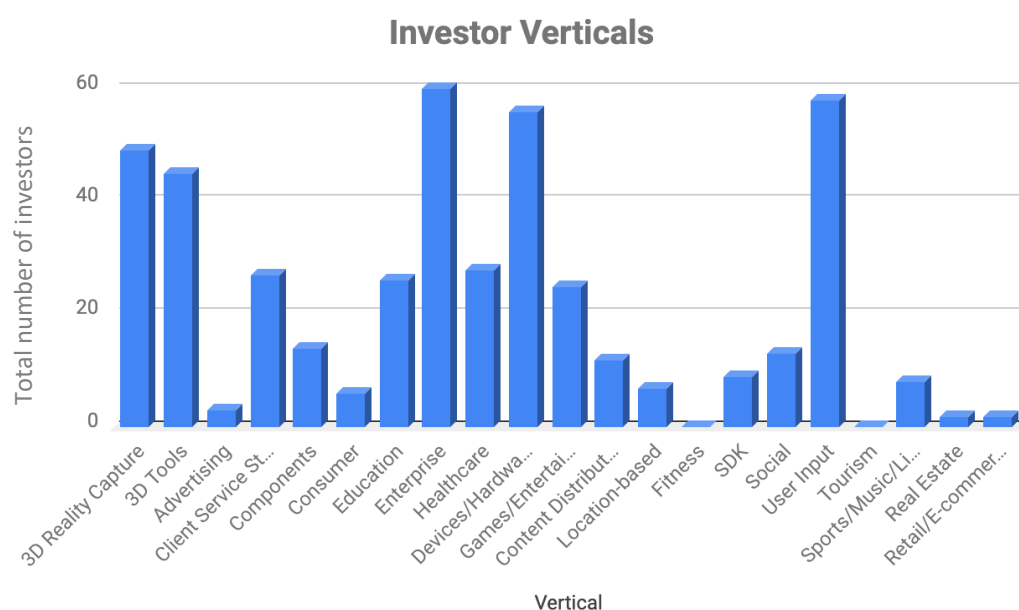


Figure 7: XR4ALL analysis of European investors for start-ups [15].

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) [16]. 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 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 [17]. It is expected to raise up to 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 8). 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 [18].

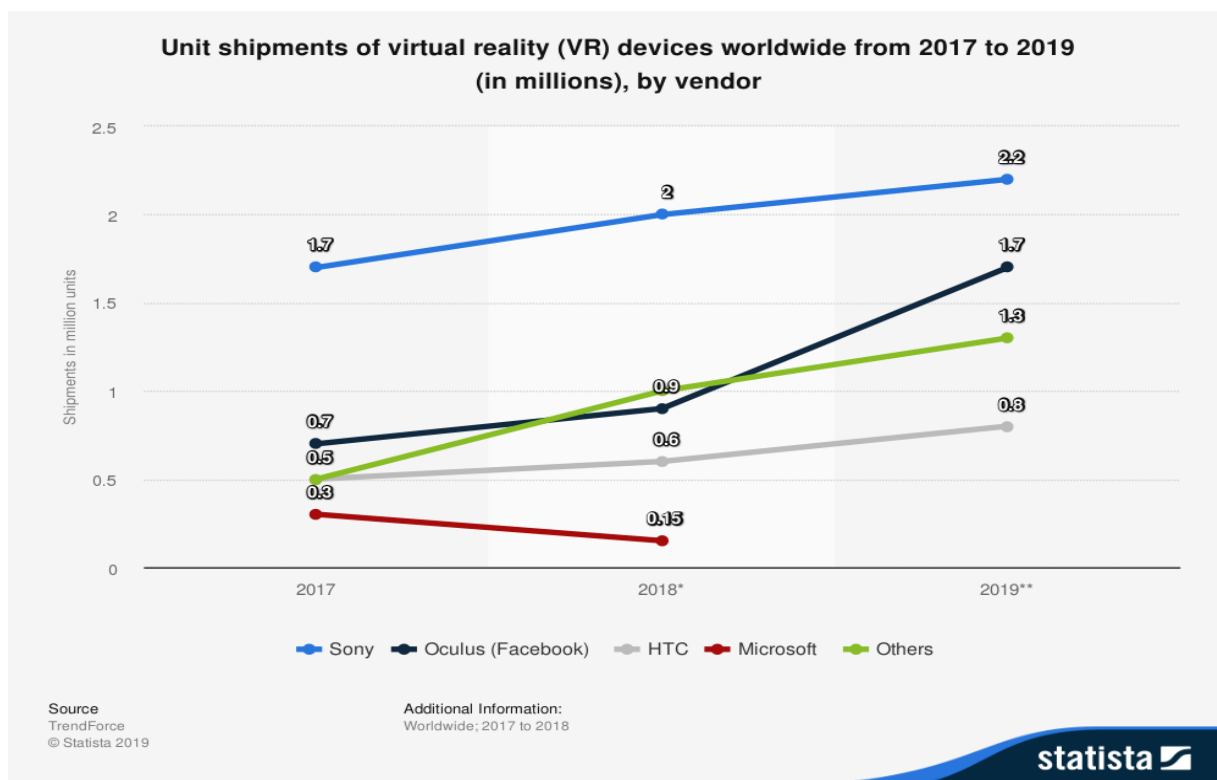


Figure 8: VR unit shipments in the last three years [17].

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 [19]. In 2019, the number of unit shipments will almost be at the same level for AR and VR headsets (about 6 millions), 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).

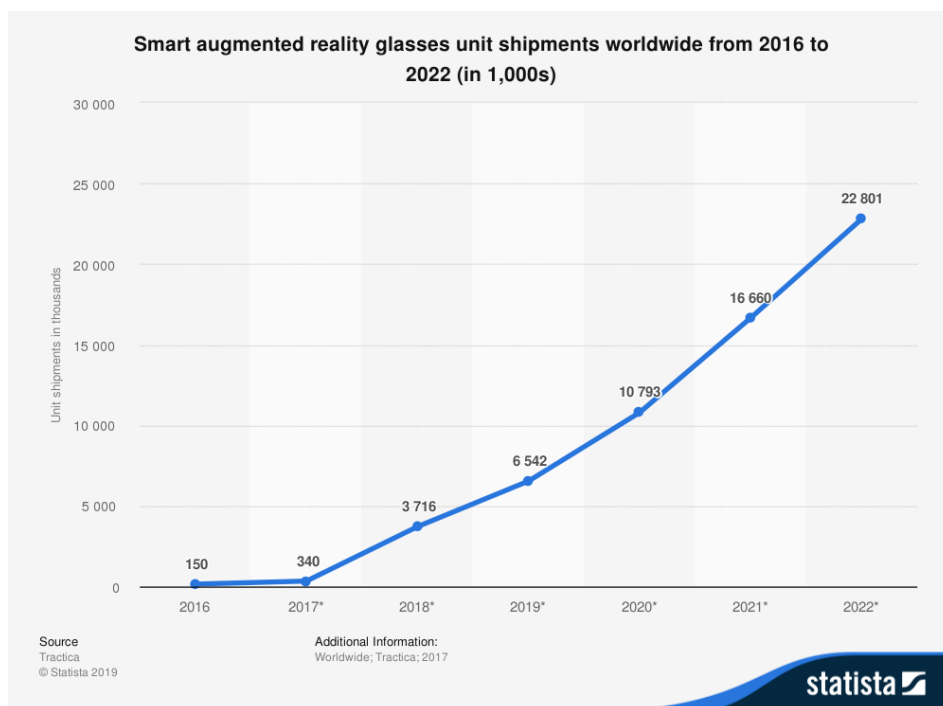


Figure 9: Forecast of AR unit shipments from 2016 to 2022 [19].

A CCS Insight forecast predicts that the shipment of VR and AR devices will continue to grow considerably and will reach over 25 million devices in 2020 and even over 50 million devices beyond 2022 [20]. However, the shipments of smartphone 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.

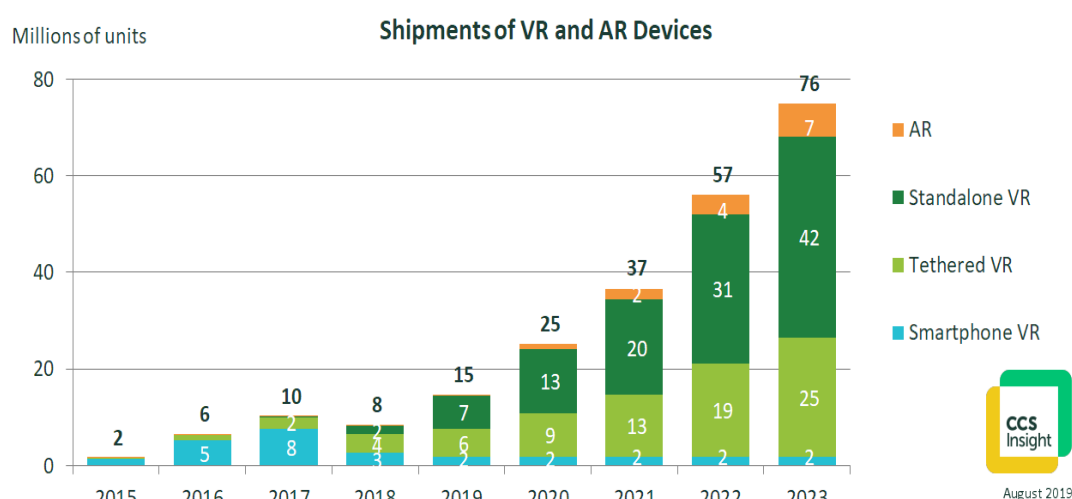


Figure 10: Forecast of VR and AR shipments [20].

3.5 Main players

With a multitude of players from start-ups and SMEs to very large enterprises, the VR/AR market is fragmented [21], 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 [12].

Main XR players [8][12] 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 11 gives a good overview of the AR industry landscape, while in Figure 12, the current VR industry landscape is depicted.

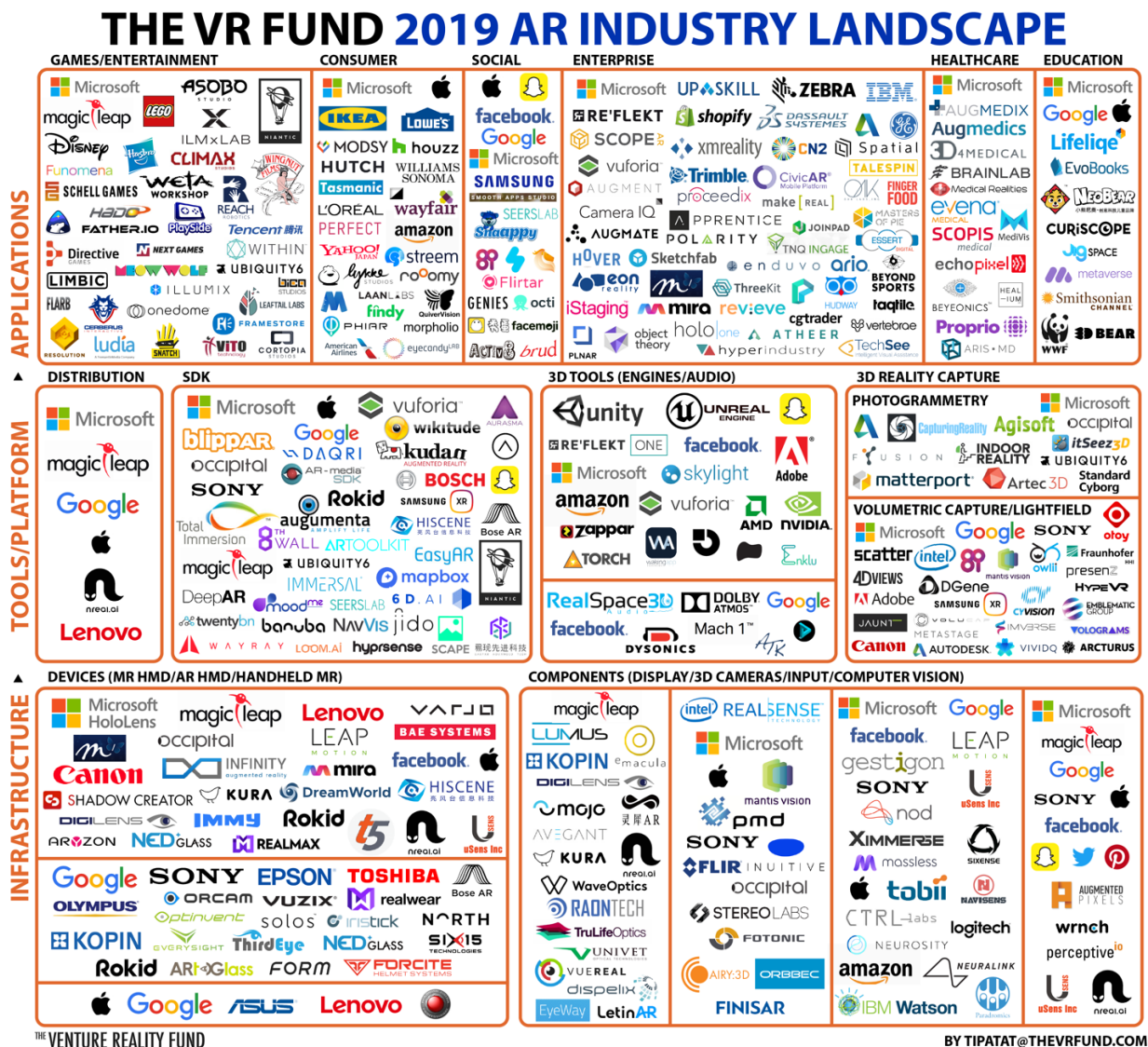


Figure 11: AR Industry Landscape by Venture Reality Fund [22].

THE VR FUND 2019 VR INDUSTRY LANDSCAPE

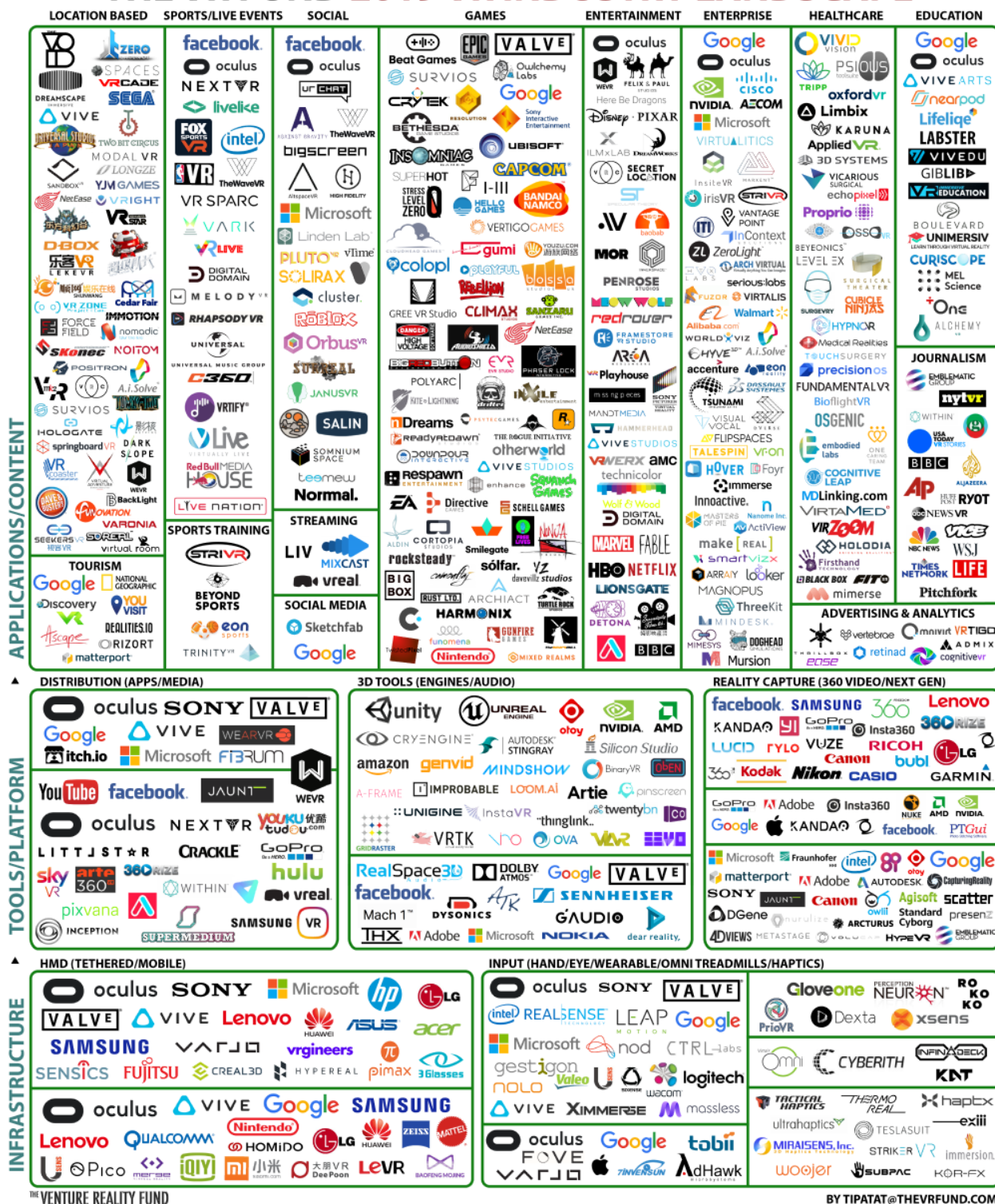


Figure 12: VR Industry Landscape by Venture Reality Fund [22].

In addition to the above corporate activities, Europe also has a long-standing tradition in research [12]. 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<>com (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 EuroVR, EUVR.org, 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 a Trello board for real-time updates on the VR/AR industry landscapes [22]. 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 listed below:

EuroVR

EuroVR is an International non-profit Association [23], 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. 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 organization of EuroVR 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 organizational members are part of EuroVR such as AVRLab, Barco, List CEA Tech, AFVR, GoTouchVR Haption, catapult, Laval Virtual, VTT, Fraunhofer FIT and Fraunhofer IAO and some European universities.

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 [24]. 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 membership of GoogleVR VIVE, Oculus, Microsoft, Samsung and PlayStationVR.

VR/AR Association (VRARA)

The VR/AR Association is an international organization 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 organizations and promotes the services of member companies [25]. The association states over 400 organizations registered as members.

VR Industry Forum (VRIF)

The Virtual Reality Industry Forum [26] is composed of a broad range of participants from sectors including, but not limited to, the movie, 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 organization, but will rely on, and liaise with, standards development organizations 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 organization focused on reducing barriers to and accelerating the smooth introduction and widespread adoption of Augmented Reality by and for professionals [27]. 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 [28]. 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 consists currently of more than 500 members, either individuals or companies spread among the whole world.

ERSTER DEUTSCHER FACHVERBAND FÜR VIRTUAL REALITY (EDFVR)

The EDFVR is the first German business association for immersive media [29]. 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 [30]. The association was founded in 2016. Our members are High-tech companies, established Media Companies, Research Institutes and Universities, Start-Ups, Freelancers and plain VR enthusiasts.

The VRBB organizes a yearly event named VRNowCon since 2016, which has an international reach of participants.

Virtual and Augmented Reality Association Austria (VARAA)

VARAA is the independent association of professional VR/AR users and companies in Austria [31]. 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 to the international VR/AR scene and the global VR/AR Association (VRARA Global).

Association Francaise de Realité Virtuelle, Augmentée, Mixte et D'Interaction 3D

The association has been founded in 2005 and comprises currently 1900 members from French public, academia and industry [32]. Its mission is to promote the development of virtual reality, augmented reality, mixed reality and 3D interaction in all their aspects: teaching, research, studies, developments and applications; to provide a means of communication between those interested in this field; and to have this community recognized by French, European and international institutions.

Virtual Reality Finland

The goal of the association is to help Finland become a leading country in VR and AR technologies [33]. The association is open for 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 [34]. 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 vigour of the public and private sectors.

VIRTUAL SWITZERLAND

This Swiss association has more than 60 members from academia and industry [35]. 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.

3.7 Patents

A study carried out by the XR4ALL consortium using the database available at the European Patent Office [36] showed that, among the 500 most recently published patents, 25 patents were filed by European companies. The publication dates in the leftmost column of the table below range between April 4th, 2019 and June 28th, 2019.

#	Applicant	Country	Title	Publication Date
1	Accenture Global Services Ltd.	IE	Augmented Reality Based Component Replacement and Maintenance	02/05/2019
2	Accenture Global Services Ltd.	IE	Virtual Reality Based Hotel Services Analysis and Procurement	15/04/2019
3	Aldin Dynamics Ehf.	IS	Methods and Systems for Path-Based Locomotion in Virtual Reality	25/04/2019
4	Arkio Ehf.	IS	Virtual/Augmented Reality Modelling Application for Architecture	16/05/2019
5	Atos Integration	FR	System for Composing or Modifying Virtual Reality Sequences, Method of Composing and System for Reading Said Sequences	06/06/2019
6	Bavastro Frederic	MC	Augmented Reality Method and System for Design	30/05/2019
7	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
8	Curious Lab Tech Ltd.	GB	Method and System for Generating Virtual or Augmented Reality	28/06/2019
9	Eaton Intelligent Power Ltd.	IE	Lighting and Internet of Things Design Using Augmented Reality	20/06/2019
10	Kitron Asa	NO	Method And System for Augmented Reality Assembly Guidance	08/04/2019
11	Medical Realities Ltd.	GB	Virtual Reality System for Surgical Training	23/05/2019
12	Metatellus Oue	EE	Augmented Reality Based Social Platform	23/05/2019
13	Nokia Technologies	FI	Virtual Reality Causal Summary Content	02/05/2019
14	Nokia Technologies	FI	Provision of Virtual Reality Content	02/05/2019
15	Nokia Technologies	FI	Provision of Virtual Reality Content	09/05/2019
16	Nokia Technologies	FI	Apparatus and Associated Methods for	13/06/2019

			Presentation of First and Second Virtual-or-Augmented Reality Content	
17	Nokia Technologies	FI	Apparatus and Associated Methods for Presentation of Augmented Reality Content	13/06/2019
18	Nokia Technologies	FI	Virtual Reality Device and a Virtual Reality Server	27/06/2019
19	Nousis Georgios Dimitriou; Kourtis Vasileios Vasileiou; Tsihli Eleanna Anastasiou	GR, GR, GR	A Method for the Production and Support Of Virtual-Reality Theatrical Performances - Installation for the Application of Said Method	04/04/2019
20	Roto Vr Ltd.	GB	Virtual Reality Apparatus	13/06/2019
21	Siemens AG.	DE	Display of Three-Dimensional Model Information in Virtual Reality	13/06/2019
22	Somo Innovations Ltd.	GB	Augmented Reality with Graphics Rendering Controlled by Mobile Device Position	23/05/2019
23	Stoecker Carsten; Innogy Innovation GmbH.	DE, DE	Augmented Reality System	25/04/2019
24	Tsapakis Stylianos Georgios	GR	Virtual Reality Set	24/05/2019
25	Unity Ipr Ap.	DK	Method and System for Synchronizing a Plurality of Augmented Reality Devices to a Virtual Reality Device	27/06/2019

3.8 References

- [3] <https://www.zionmarketresearch.com/report/augmented-and-virtual-reality-market>
- [4] <https://www.mordorintelligence.com/industry-reports/extended-reality-xr-market>
- [5] <https://www.marketsandmarkets.com/PressReleases/augmented-reality.asp>
- [6] <https://www.marketsandmarkets.com/PressReleases/ar-market.asp>
- [7] <https://www.credit-suisse.com/ch/en/articles/private-banking/virtual-und-augmented-reality-201706.html>
- [8] <https://www.credit-suisse.com/ch/en/articles/private-banking/zunehmende-einbindung-von-Virtual-und-augmented-reality-in-allen-branchen-201906.html>
- [9] <https://www.researchandmarkets.com/reports/4746768/virtual-reality-market-by-offering-technology>
- [10] <https://techcrunch.com/2018/01/25/ubiquitous-ar-to-dominate-focused-vr-by-2022/>

- [11] <https://optics.org/news/10/10/18>
- [12] https://ec.europa.eu/futurium/en/system/files/ged/vr_ecosystem_eu_report_0.pdf
- [13] <https://www.capgemini.com/research-old/augmented-and-virtual-reality-in-operations/>
- [14] <https://www.digi-capital.com/news/2019/07/ar-vr-investment-and-ma-opportunities-as-early-stage-valuations-soften/>
- [15] XR4ALL Project, Deliverable D5.1 Map of funding sources for XR technologies
- [16] <https://www.idc.com/getdoc.jsp?containerId=prUS45123819>
- [17] <https://www.statista.com/statistics/671403/global-virtual-reality-device-shipments-by-vendor/>
- [18] <https://www.roadtovr.com/monthly-connected-vr-headsets-steam-1-million-milestone/>
- [19] <https://www.statista.com/statistics/610496/smart-ar-glasses-shipments-worldwide/>
- [20] <https://www.ccsinsight.com/press/company-news/virtual-and-augmented-reality-headset-shipments-ready-to-soar/>
- [21] <https://ec.europa.eu/growth/tools-databases/dem/monitor/category/augmented-and-virtual-reality>
- [22] <http://www.thevrfund.com/resources/industry-landscape/>
- [23] <https://www.eurovr-association.org/>
- [24] <https://xra.org/>
- [25] <https://www.thevrara.com/>
- [26] <https://www.vr-if.org/>
- [27] <https://thearea.org/>
- [28] <https://ivrpa.org/>
- [29] <http://edfvr.org/>
- [30] <https://virtualrealitybb.org/>
- [31] <https://www.gensummit.org/sponsor/varaa/>
- [32] <https://www.af-rv.fr/>
- [33] <https://vrfinland.fi>
- [34] <https://fivr.fi/>
- [35] <http://virtualswitzerland.org/>
- [36] <https://worldwide.espacenet.com>

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 organized in three different major technologies: (1) 360-degree video (3-DoF), (2) 3D data from real scenes, and (3) 3-degrees of freedom (3DoF) + head motion parallax (3-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.

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 [37][38].

Finally, 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.

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 [39], 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 [40].

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 4.5.2). 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 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 [41]. 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 [42].

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 [43]. 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 [44].

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 cameras 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 this large amount of initial data is then merged to a dynamic 3D point cloud 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. To attain this goal, the 3D point cloud is usually converted to dynamic and often also simplified meshes with a related video organized as texture atlas for later texture mapping. 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.

In recent years, a number of volumetric studios have been created [45][46][47][48][49] that are able to produce high quality volumetric videos. 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 [46], Volucap [47], Uncorporeal [48], and 4DViews [49]) providing input for volumetric video similar to frame-by-frame photogrammetric reconstruction of the actors, while Microsoft's Mixed reality Capture Studios [45] 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 [47] 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 [50].

4.1.5 References

- [37] 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.
- [38] <http://stateofvr.com/>
- [39] <https://filmora.wondershare.com/virtual-reality/top-10-professional-360-degree-cameras.html>

- [40] <https://www.hhi.fraunhofer.de/en/departments/vit/technologies-and-solutions/capture/panoramic-uhd-video/omnicam-360.html>
- [41] <https://packet39.com/blog/2018/02/25/3dof-6dof-roomscale-vr-360-video-and-everything-in-between/>
- [42] <https://www.sonymusic.co.uk/news/2018-07-27/experience-tom-grennan-vr-on-playstation>
- [43] F. Fadli, H. Barki, P. Boguslawski, L. Mahdjoubi (2015), “3D Scene Capture: A Comprehensive Review of Techniques and Tools for Efficient Life Cycle Analysis (LCA) and Emergency Preparedness (EP) Applications”. 10.2495/BIM150081.
- [44] <https://www.aniwaa.com/best-3d-body-scanners/>
- [45] <http://www.microsoft.com/en-us/mixed-reality/capture-studios>
- [46] <http://8i.com>
- [47] <http://www.volucap.de>
- [48] <http://uncorporeal.com>
- [49] <http://www.4dviews.com>
- [50] O. Schreer, I. Feldmann, S. Renault, M. Zepp, P. Eisert, P. Kauff, “Capture and 3D Video Processing of Volumetric Video”, IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, September 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. Probably most well-known is 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 [51]. 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 [51]. 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. With these stimuli, the human sound perception also 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 localize sound sources in three-dimensional space. Again see [51] 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, Neve 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 [52]. 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 [53].

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 [54]. 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 [55].

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 [56].

4.2.7 References

- [51] J. Schnupp, I. Nelken, A. King, (2011). Auditory neuroscience: Making sense of sound. MIT press.
- [52] <https://www.iis.fraunhofer.de/en/ff/amm/consumer-electronics/uphear-microphone.html>
- [53] R. K. Furness, (1990). Ambisonics-an overview. In Audio Engineering Society Conference: 8th International Conference: The Sound of Audio. Audio Engineering Society.
- [54] https://mhacoustics.com/sites/default/files/Eigenbeam%20Datasheet_R01A.pdf
- [55] <http://www.fascinate-project.eu/index.php/tech-section/audio/>
- [56] <https://www.zylia.co/zylia-6dof.html>

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. Except some very special uses 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 [57]. In contrast, SLAM (Simultaneous Localization 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 [58]. 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 [59][60], 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 [61], significantly outperforming previous model-based approaches [62][63].

4.3.3 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 [64][65][66][67][68][69][70]. The best performing methods use body-markers. In a realistic marker less 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 [71][72][73]. General-case techniques for deformation and scene capture are far less

developed [75]. 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 [77][78]. First approaches have addressed general non-rigid deformation modelling by incorporating geometric constraints into deep learning.

4.3.4 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 [75]. 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 [76]. Recent progress in deep learning also enables the reconstruction of highly accurate human body models even from single RGB images [68]. Similarly, Pavlakos et al. [69] 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 [70] go one step further and enable real-time capture of humans including surface deformations due to clothes.

4.3.5 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) [71] or user intervention [75]. 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 [79], to analyse the relationships between 3D-shape, reflectance and natural illumination [80], and to estimate the reflectance maps of specular materials in natural lighting conditions [81]. For specific objects, like human faces, image statistics from sets of examples can be exploited for generic appearance modelling [82], and recent approaches have achieved realistic results using deep neural networks to model human faces in still images [83][84]. 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 [85][86][87]. 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 [88][89] 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.

4.3.6 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 [90] and [91], who synthesize novel video sequences of facial animations and other dynamic scenes by video resampling. Malleson et al. [92] 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. [93] and Serra et al. [94], 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. [95] introduced a flexible approach to synthesize 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 [96]. Similarly, Paier et al. [97] 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 [98][99], yielding high visual quality. However, these methods are limited to replaying segments of the captured motions. In [100] 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. [101] combined concatenation of captured 3D sequences with view dependent texturing for real-time interactive animation. Similarly, Volino et al. [102] presented a parametric motion graph-based character animation for web applications. Only recently, Boukhayma and Boyer [103][104] proposed an animation synthesis structure for the recomposition of textured 4D video capture, accounting for geometry and appearance.

They propose a graph structure that enables interpolation and traversal between precaptured 4D video sequences. Finally, Regateiro et al. [105] 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 synthesize images and video. Chan et al. [106], for example, use 2D skeleton data to transfer body motion from one person to another and synthesize new videos with a Generative Adversarial Network. The skeleton motion data can also be estimated from video by neural networks [107]. Liu et al. [108] 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 synthesizing facial video as shown, e.g., in [109].

4.3.7 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 exists, 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 are 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 localization 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 relocalization, the tracking and the mapping. The relocalization is used when the XR device has no idea about its pose (initialization 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 relocalization 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 relocalization task. The loop that tracks the XR device and that maps the real environment is called SLAM (Simultaneous localization And Mapping) [110][111]. 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 optimization based on vision processing unit. Recent implementation of SLAM pipelines are more and more using low-level components based on machine learning approaches [112][113][114][115]. Finally, future 5G network offering low latency and high bandwidth will allow to distribute into edge cloud and into centralized cloud efficient pipelines to improve AR devices localization accuracy, even on low-resources AR devices, and address large scale AR applications.

4.3.8 References

- [57] 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, pp. 11-20, 1996.
- [58] R. Mur-Artal, J. Motiel, J. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system", Trans. on Robot, 2015.
- [59] M. Poggi et al, "Learning monocular depth estimation with unsupervised trinocular assumptions", Int. Conf. on 3D Vision (3DV), Verona, 2018, pp. 324-333. Proc. 3DV, 2018.
- [60] H. Zhou, B. Ummenhofer, T. Brox, "DeepTAM: Deep tracking and mapping", European Conference on Computer Vision (ECCV), 2018.
- [61] 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.
- [62] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, A. Geiger, "Occupancy Networks: Learning 3D reconstruction in function space", arXiv:1812.03828, 2018.
- [63] J. Park, P. Florence, J. Straub, R. Newcombe, S. Lovegrove, "DeepSDF: Learning continuous SDFs for shape representation", arXiv:1901.05103, 2019
- [64] P.-L. Hsieh et al. "Unconstrained real-time performance capture", In Proc. Computer Vision and Pattern Recognition (CVPR), 2015.
- [65] M. Zollhöfer, J. Thies, P. Garrido, D. Bradley, T. Beeler, P. Pérez, M. Stamminger, M. Nießner, C. Theobalt, (2018). „State of the Art on Monocular 3D Face Reconstruction, Tracking, and Applications", Comput. Graph. Forum, 37, 523-550.

- [66] A. Tewari, M. Zollhöfer, F. Bernard, P. Garrido, H. Kim, P. Pérez, C. Theobalt, „High-Fidelity Monocular Face Reconstruction based on an Unsupervised Model-based Face Autoencoder”, IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 2018.
- [67] A. Tkach, A. Tagliasacchi, E. Remelli, M. Pauly, A. Fitzgibbon, “Online generative model personalization for hand tracking”, In ACM Trans. on Graphics 36(6), 2017.
- [68] 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.
- [69] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A. Osman, D. Tzionas, and M. Black, “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.
- [70] 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.
- [71] T. Alldieck et al. “Detailed human avatars from monocular video”, in Proc. Int. Conf. on 3D Vision (3DV), 2018.
- [72] D. Mehta et al. “VNect: Real-time 3D human pose estimation with a single RGB camera”, In ACM Transactions on Graphics (TOG), 36(4), 2017.
- [73] A. Kanazawa et al. “End-to-End recovery of human shape and pose”, In Proc. Computer Vision and Pattern Recognition (CVPR), 2018.
- [74] J. T. Barron et al. “Shape, illumination, and reflectance from shading”, in Trans. on Pattern Analysis and Machine Intelligence (PAMI), 2015.
- [75] V. F. Abrevaya, S. Wuhler, 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.
- [76] 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.
- [77] M. Habermann et al. “NRST: Non-rigid surface tracking from monocular video”, in Proc. GCPR, 2018.
- [78] C. Vondrick et al. “Generating videos with scene dynamics”, in Proc. Int. Conf. on Neural Information Processing Systems (NIPS) 2016.
- [79] T. Zhou et al. “Learning data-driven reflectance priors for intrinsic image decomposition”, in Proc. Int. Conf. on Computer Vision (ICCV), 2015.
- [80] 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.
- [81] R. Konstantinos et al. “Deep reflectance maps”, in Proc. Proc. Computer Vision and Pattern Recognition (CVPR), 2016.
- [82] T. F. Cootes et al., “Active appearance models”, in Trans. on Pattern Analysis and Machine Intelligence (PAMI), 23(6), 2001.

- [83] L. Hu et al. "Avatar digitization from a single image for real-time rendering", in ACM Transactions on Graphics (TOG), 36(6), 2017.
- [84] S. Lombardi et al. "Deep appearance models for face rendering", in ACM Transactions on Graphics (TOG), 37(4), 2018.
- [85] K. Bousmalis et al. "Unsupervised pixel-level DA with generative adversarial networks", in Proc. Proc. Computer Vision and Pattern Recognition (CVPR), 2017.
- [86] T.-C. Wang et al. "Video-to-video synthesis", in Proc. 32nd Int. Conf. on Neural Information Processing Systems (NIPS), pages 1152-1164, 2018.
- [87] C. Finn et al. "Unsupervised learning for physical interaction through video prediction", in Proc. Int. Conf. on Neural Information Processing Systems (NIPS), 2016.
- [88] S.M. Ali Eslami et al. "Neural scene representation and rendering", In Science 360 (6394), 2018.
- [89] Z. Zang et al. "Deep generative modeling for scene synthesis via hybrid representations", in arXiv:1808.02084, 2018.
- [90] C. Bregler, M. Covell, and M. Slaney, "Video Rewrite: Driving Visual Speech with Audio." in ACM SIGGRAPH, 1997.
- [91] A. Schodl, R. Szeliski, D. Salesin, and I. Essa, "Video Textures." in ACM SIGGRAPH, 2000.
- [92] C. Malleson, J. Bazin, O. Wang, D. Bradley, T. Beeler, A. Hilton, and A. Sorkine-Hornung, "Facedirector: Continuous control of facial performance in video", Proc. Int. Conf. on Computer Vision (ICCV), Santiago, Chile, Dec. 2015.
- [93] 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.
- [94] J. Serra, O. Cetinaslan, S. Ravikumar, V. Orvalho, and D. Cosker, "Easy Generation of Facial Animation Using Motion Graphs", Computer Graphics Forum, 2018.
- [95] F. Xu, Y. Liu, C. Stoll, J. Tompkin, G. Bharaj, Q. Dai, H.-P. Seidel, J. Kautz, and C. Theobald, "Video-based Characters - Creating New Human Performances from a Multiview Video Database", in ACM SIGGRAPH, 2011.
- [96] 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.
- [97] 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.
- [98] C. Bregler, M. Covell, and M. Slaney, "Video-based Character Animation", In ACM Symp. on Computer Animation, 2005.
- [99] P. Hilton, A. Hilton, and J. Starck, "Human Motion Synthesis from 3D Video", In Proc. Computer Vision and Pattern Recognition (CVPR), 2009.

- [100] 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.
- [101] 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.
- [102] 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.
- [103] 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.
- [104] 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.
- [105] 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.
- [106] C. Chan, S. Ginosar, T. Zhou, and A. Efros, “Everybody dance now”, in *Proc. Int. Conf. on Computer Vision (ICCV)*, Seoul, Korea, Oct. 2019.
- [107] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H.-P. Seidel, W. Xu, D. Casas, and C. Theobalt, “Vnect: Real-time 3d human pose estimation with a single RGB camera”, *Proc. Computer Graphics (SIGGRAPH)*, vol. 36, no. 4, July 2017.
- [108] L. Liu, W. Xu, M. Zollhoefer, H. Kim, F. Bernard, M. Habermann¹, W. Wang, and C. Theobalt, “Neural rendering and re-enactment of human actor videos”, *ACM Trans. of Graphics*, 2019.
- [109] H. Kim, P. Garrido, A. Tewari, W. Xu, J. Thies, M. Nießner, P. Pérez, C. Richardt, M. Zollöfer, and C. Theobalt, “Deep video portraits”, *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, p. 163, 2018.
- [110] Andrew J. Davidson, “Real-Time Simultaneous Localisation and Mapping with a Single Camera”, *IEEE Int. Conf. on Computer Vision (ICCV)*, 2003.
- [111] Georg Klein and David Murray, “Parallel Tracking and Mapping for Small AR Workspaces”, in *Proc. International Symposium on Mixed and Augmented Reality (ISMAR’07)*.
- [112] N. Radwan, A. Valada, W. Burgard, “VLocNet++: Deep Multitask Learning for Semantic Visual Localization and Odometry”, in *IEEE Robotics and Automation Letters*, Vol 3, issue 4, Oct. 2018
- [113] L. Sheng, D. Xu, W. Ouyang, X. Wang, “Unsupervised Collaborative Learning of Keyframe Detection and Visual Odometry Towards Monocular Deep SLAM”, *ICCV 2019*.
- [114] M. Bloesch, T. Laidlow, R. Clark, S. Leutenegger, A. J. Davison, “Learning Meshes for Dense Visual SLAM”, *ICCV 2019*
- [115] 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, January 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 [116] and [117].

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 predefined 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 [118]. 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) utilizes the Euclidean distance from a source to the different speakers and makes no assumption about the position of the listener [119]. 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 auralization. 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 auralization 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 [120]. 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 [120]. 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 [120] 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 [120]. 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 [121][122].

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 standardized format specialized 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 [123]. 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 [124]. 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 [125].

Besides being used in cinema and TV, 3D auralization can also be used for VR. In particular, the VR players used in game engines are suitable tools for the creation of 3D auralization. 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 spatialization 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 auralization. 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 auralization.

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

- [116] <https://www.vrtonung.de/en/virtual-reality-audio-formats/>
- [117] <https://www.vrtonung.de/en/spatial-audio-support-360-videos/>
- [118] V. Pulkki, (2001), “Spatial sound generation and perception by amplitude panning techniques”, PhD thesis, Helsinki University of Technology, 2001.
- [119] T. Lossius, P. Baltazar, T.de la Hogue, “DBAP—distance-based amplitude panning”, in Proc. Of Int. Computer Music Conf. (ICMC), 2009.
- [120] 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.
- [121] <https://facebookincubator.github.io/facebook-360-spatial-workstation/>
- [122] <https://support.google.com/youtube/answer/6395969?co=GENIE.Platform%3DDesktop&hl=en>
- [123] T. Ziemer, (2018) Wave Field Synthesis. In: Bader R. (eds) Springer Handbook of Systematic Musicology. Springer Handbooks. Springer, Berlin, Heidelberg
- [124] R. Rabenstein, S. Spors, (2006, May). Spatial aliasing artefacts produced by linear and circular loudspeaker arrays used for wave field synthesis. In Audio Engineering Society Convention 120. Audio Engineering Society.
- [125] <https://www.iis.fraunhofer.de/en/ff/amm/broadcast-streaming/mpeg.html>

4.5 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 4.5.1).

- 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 4.5.2).
- 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 4.5.3).
- 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 4.5.4).

4.5.1 Stereoscopic 3D displays and projections

Stereoscopic 3D (S3D) has been used for decades for the visualization 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 [126][127]. 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 polarization 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, autostereoscopic 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-dimensionnal 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” visualization 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. [128].

Since the 1950s, S3D viewing has seen several phases of popularity and a correspond 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 visualization. Each such wave came with extreme and unwarranted expectation. During, the last wave, TV manufacturers succeeded for w while in convincing every consumer to replaced 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 S3D 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 [129]), called Simulated Reality Display and demonstrated successfully at CES 2019 [130]. Similar to the earlier tracked auto-stereoscopic 3D displays, as for example Fraunhofer HHI’s Free3C Display [131] 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.

4.5.2 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 have been steadily growing [132]. At the gaming platform Steam, the yearly growing rate of monthly-connected headsets is even up to 80% [133].

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.

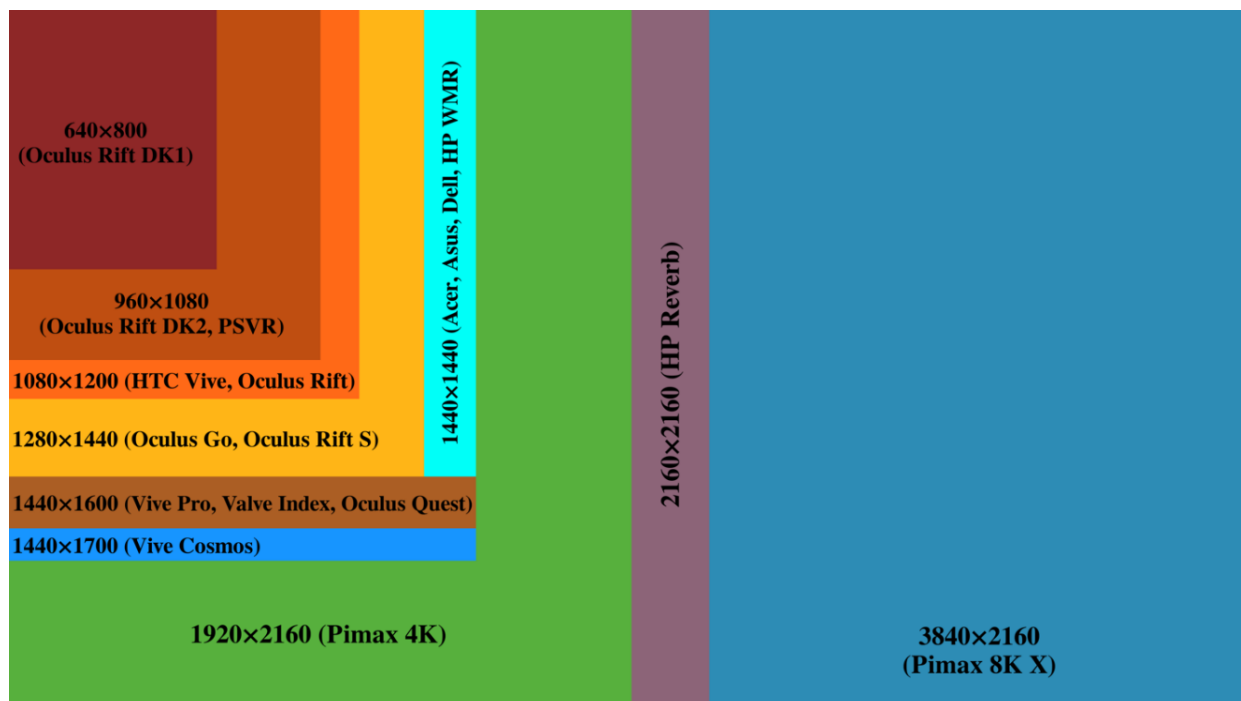


Figure 13: Comparison chart of VR headset resolutions [134].

Figure 13 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, and 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.

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.

4.5.3 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 micro screen, which is placed around the optical system and outside the field of view, in such a way that the image on the micro screen 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 micro screen. A system using beam-splitter is generally more bulky 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 projects 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 commercialization 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 [135].

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 Leap1 glasses seem to confirm that Microsoft currently dominates the AR field. Magic Leap itself admits that they have been leapfrogged by HoloLens 2 [136]. Rumours indicate that Microsoft purposely delayed the commercialization 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 need power supply have to be integrated in smart devices with extremely small form factors. One solution might be the combination with upcoming wireless technology, i.e. the 5G standard, and its capability to outsource complex computations to the cloud while keeping the low latency and fast response needed for interactivity (see section 4.9).

Very recent, VR headsets integrate two cameras, with one in front of each user's eye. Each such camera capture 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 Localization 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 [137]. 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.

4.5.4 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 4.5.1). 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 onboard cameras that analyze the real world environment, often in combination with additional depth sensors. The user position is then calculated in relation to previously analyzed spatial anchor points in the real 3D world. By contrast, location-based VR entertainment systems (e.g. The Void [138]) 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 onboard cameras on the headset determine its position relatively to the markers (e.g. Illusion Walk [139]).

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 [140]). 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 cameras to track and visualize 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 be integrated in the new HTC Vive Pro Eye [141]. 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-specialized haptic devices like the Phantom Premium from 3D Systems allow an extremely accurate force feedback [142]. Other highly specialized 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 [143], which has now merged with the above mentioned hand-tracking company Leap Motion [140], with the resulting company now being called Ultraleap [144]. 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. AvatarVR [145], Cyberith Virtualizer [146] or KAT VR [147]).

4.5.5 References

- [126] C. Cruz-Neira, D. J. Sandin, T.A. DeFanti, R. V. Kenyon, J. C. Hart, (1 June 1992). "The CAVE: Audio Visual Experience Automatic Virtual Environment". Commun. ACM. 35 (6): 64–72.
- [127] S. Manjrekar, S. Sandilya, D. Bhosale, S. Kanchi, A.Pitkar, M. Gondhalekar, "CAVE: An Emerging Immersive Technology - A Review", 2014 UK Sim-AMSS 16th International Conference on Computer Modelling and Simulation
- [128] H. Urey, K. V. Chellappan, E. Erden, P. Surman, „State of the Art in Stereoscopic and Autostereoscopic Displays“, Proceedings of the IEEE, Bd. 99, No. 4, S. 540–555, Apr. 2011.
- [129] See <https://bits-chips.nl/artikel/dimenco-back-in-dutch-hands/>
- [130] <https://www.dimenco.eu/newsroom/simulated-reality-development-kit-shipping-november-11th>
- [131] K. Hopf, P. Chojecki, F. Neumann, und D. Przewozny, „Novel Autostereoscopic Single-User Displays with User Interaction“, in SPIE Three-dimensional TV, Video, and Display V, Boston, MA, USA, 2006.
- [132] TrendForce Global VR Device Shipments Report, Sep. 2019, Source: <https://www.statista.com/statistics/671403/global-virtual-reality-device-shipments-by-vendor/>
- [133] Analysis: Monthly-connected VR Headsets on Steam Pass 1 Million Milestone <https://www.roadtovr.com/monthly-connected-vr-headsets-steam-1-million-milestone/>
- [134] Comparison of virtual reality headsets. Wikipedia, 2019. Source: https://en.wikipedia.org/wiki/Comparison_of_virtual_reality_headsets
- [135] Whitepaper, "Head Mounted Displays & Data Glasses". Applications and Systems, Virtual Dimension Center (VDC). 2016
- [136] <https://mspoweruser.com/magic-leap-admits-they-have-been-leapfrogged-by-hololens-2/>
- [137] „Augmented Reality: Innovations and Lifecycle“, Parks Associates, 2019. Source <https://www.parksassociates.com/report/augmented-reality>
- [138] <https://www.thevoid.com/>
- [139] <https://www.illusion-walk.com/>
- [140] <https://www.leapmotion.com/>
- [141] <https://vr.tobii.com/>
- [142] <https://www.3dsystems.com/scanners-haptics#haptic-devices>
- [143] <https://www.ultrahaptics.com/>
- [144] <https://ultraleap.com/>
- [145] <https://avatarvr.es/>
- [146] <https://www.cyberith.com/>
- [147] <https://katvr.com/>

4.6 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 centralized cloud or the edge the cloud.

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 [148][149]. Indeed, the 1 to 3 milliseconds of latency induced by the distribution of calculations on the edge cloud does not impact the user experience and preserves a “motion-to-photon” latency under 20ms. Thus, XR remote rendering allow the users to immerse themselves in CAD models of several hundred million polygons using autonomous AR or VR devices that can hardly display more than 100,000 polygons in real time [150]. 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 [151], 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 [152].

Also, AR is announced as a breakthrough poised to revolutionize 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 [153]. 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 [154] as well as standardization bodies such as the Industry Specification Group “Augmented Reality Framework” at ETSI [155] are already working on specifications and frameworks to ensure ARCloud interoperability.

4.6.1 References

- [148] S. Shi, V. Gupta, M. Hwang, R. Jana, “Mobile VR on edge cloud: a latency-driven design”, Proc. Of the 10th ACM Multimedia Systems Conference, pp. 222-231, June 2019.

- [149] “Cloud AR/VR Whitepaper”, GSMA: <https://www.gsma.com/futurenetworks/wiki/cloud-ar-vr-whitepaper>
- [150] Holo-Light AR edge computing: <https://www.holo-light.com/edge-computing-iss.html>
- [151] “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
- [152] Podcast Terry Schussler (Deutsche Telekom) on the importance of 5G and edge computer for AR: <https://www.thearshow.com/podcast/043-terry-schussler>
- [153] Charlie Fink’s, “Metaverse. An AR Enabled Guide to VR & AR”
- [154] <https://www.openarcloud.org/>
- [155] <https://www.etsi.org/committee/arf>

4.7 Conclusion

The presented technology domains are considered as the most relevant ones for interactive XR technologies and applications. However, some areas have not yet been documented such as synthetic content authoring, details on rendering of voxels, meshes, and mesh processing. This will be provided in the next version of this deliverable.

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. Several apps can be found for this application here [156][157][158][159][160][161]. All these applications use the integrated registration and tracking technology that is available for IOs devices (ARKit) and Android devices (ARCore).

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: [162][163][164][165].

In the food & beverage industry, AR is used to allow users to preview their potential order [166].



Figure 14: Examples for different AR applications in advertising and commerce
(l.t.r.) IKEA app, real estate app by Onirix, food preview app by Jarit.

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, Avametric, 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 & anytime [167]. Different styles can be combined and jointly discussed. Clothing can be presented in a photo-realistic way.



Figure 15: Sample views of ZREALITY virtual show room.

In Figure 16, the eco-system for the fashion industry is depicted. The major players for the development of AR & VR applications are listed.



Figure 16: The fashion eco-system.

5.1.1 References

- [156] <http://www.augmentedfurniture.com/>
- [157] <https://apps.apple.com/ca/app/build-com-home-improvement/id1053662668>
- [158] <https://apps.apple.com/ca/app/magicplan/id427424432>
- [159] <https://apps.apple.com/app/plnar/id1282049921>
- [160] <https://apps.apple.com/ca/app/housecraft/id1261483849>
- [161] <https://apps.apple.com/ca/app/ikea-place/id1279244498>
- [162] <https://www.onirix.com/learn-about-ar/augmented-reality-in-real-estate/>
- [163] <https://www.obsessar.com/>
- [164] <https://www.avametric.com/>
- [165] <https://www.virtusize.com/site/>
- [166] <https://jarit.app>
- [167] <https://www.zreality.com/vr-mode/>

5.2 Design, engineering, and manufacturing

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.

In this section, we'll attempt to identify and characterize the main uses for AR in industry 4.0 and construction.

5.2.1 Industry 4.0 - assembly



Figure 17: AR application for assembly in industrial environment.

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. 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.).

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 digitized, 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.

5.2.2 Industry 4.0 - 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. 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.



Figure 18: AR application for quality control in industrial environment.

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.

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.

5.2.3 Industry 4.0 - field servicing (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 contextualized manner. This includes maintenance procedures, technical documentation, map of inspection rounds, data from sensors, etc...



Figure 19: AR application for maintenance in industrial environment.

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.

Finally, the most advanced AR may include remote assistance systems. These systems enable the operator in the field to share what they're 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 mobilization 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.

5.2.4 Industry 4.0 - factory planning (designing a plant, workstation, or productive resources)

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 digitize 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 visualization 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.

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.

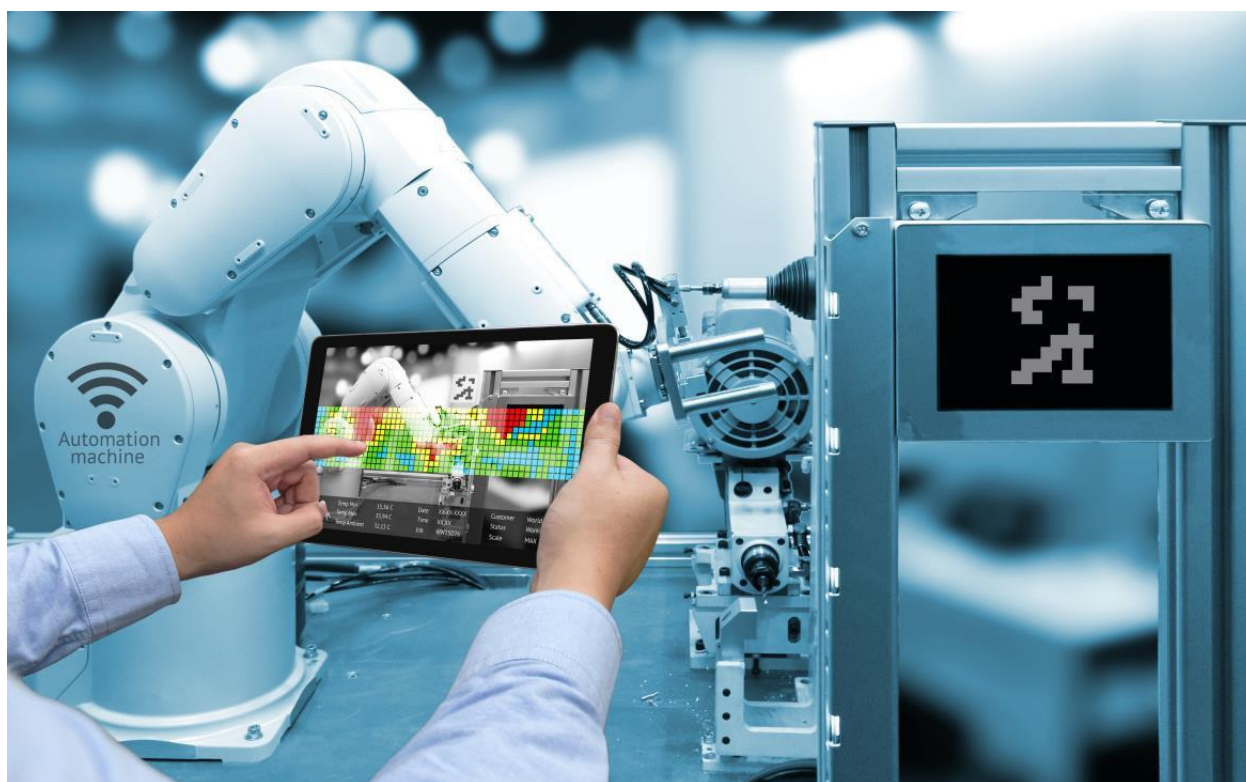


Figure 20: AR application for planning in industrial environment.

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.

5.2.5 Industry 4.0 - logistics

Although the use of AR is still emerging in the field of logistics, it does appear to be a promising source of time savings, particularly out of a desire to optimize the operations of order picking and preparation.

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.

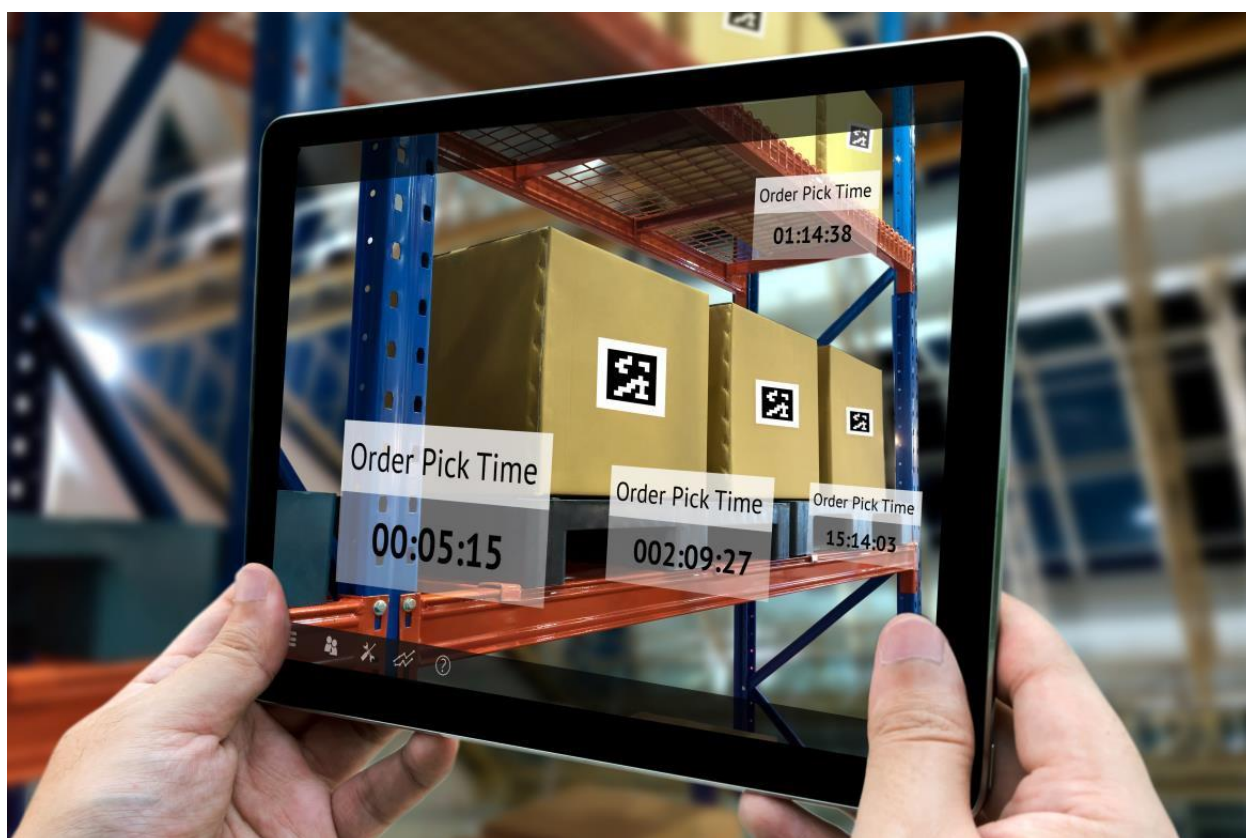


Figure 21: AR application for logistics.

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.

5.2.6 Industry 4.0 - training

VR and AR applied to the field of training is of great benefit and interest for learners, thanks to its visualization and natural interaction capabilities. The use of VR & AR for training enables one to understand phenomena and procedures. For instance, this use offer 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).



Figure 22: AR application for training in industrial environment.

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 quality and of memorization of knowledge acquired.

5.3 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-years period of 1995-1997 to the 2-year period of 2013-2015, thus separated by 18 years [168]. 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 visualization of imagery from X-ray computed tomography (CT) and magnetic-resonance imaging (MRI) via VR or AR systems could revolutionize 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.

5.3.1 Training and learning

Using VR and AR 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 visualization 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.

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.

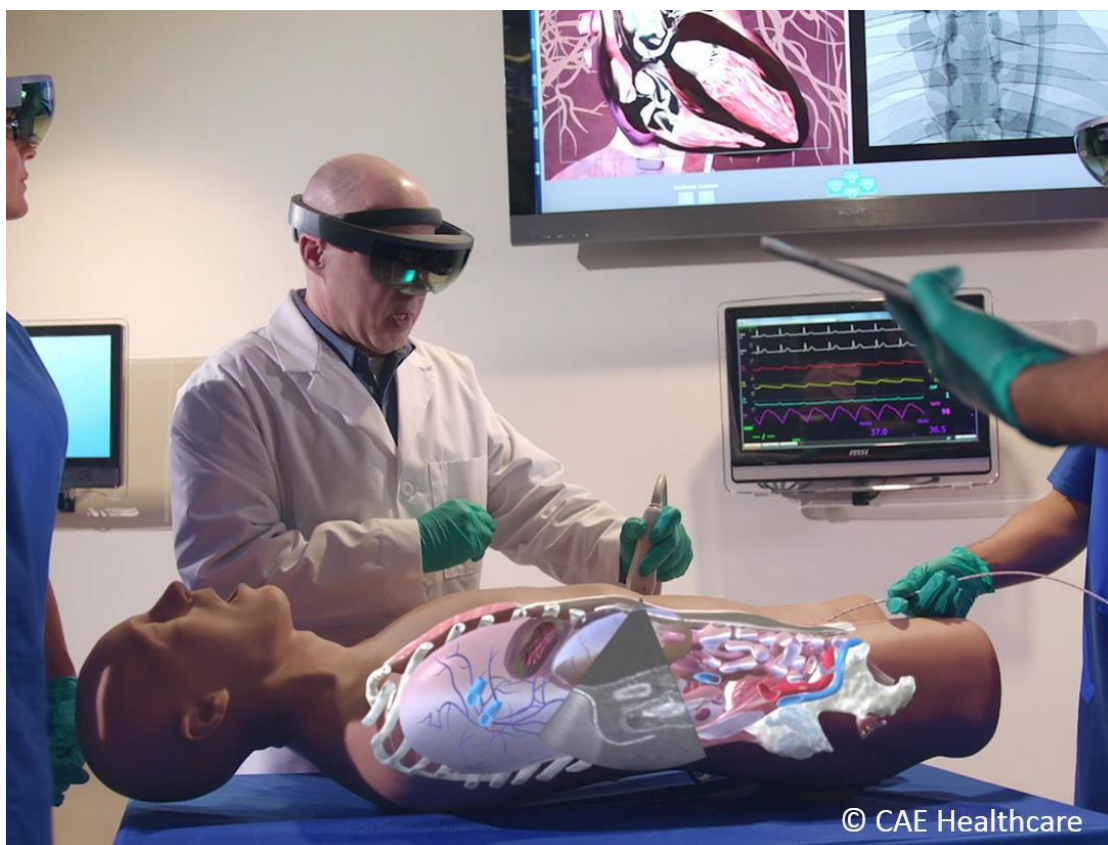


Figure 23: Example of training and learning use of AR in the medical domain.

To date, many proofs-of-concept 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 visualization 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.

5.3.2 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.

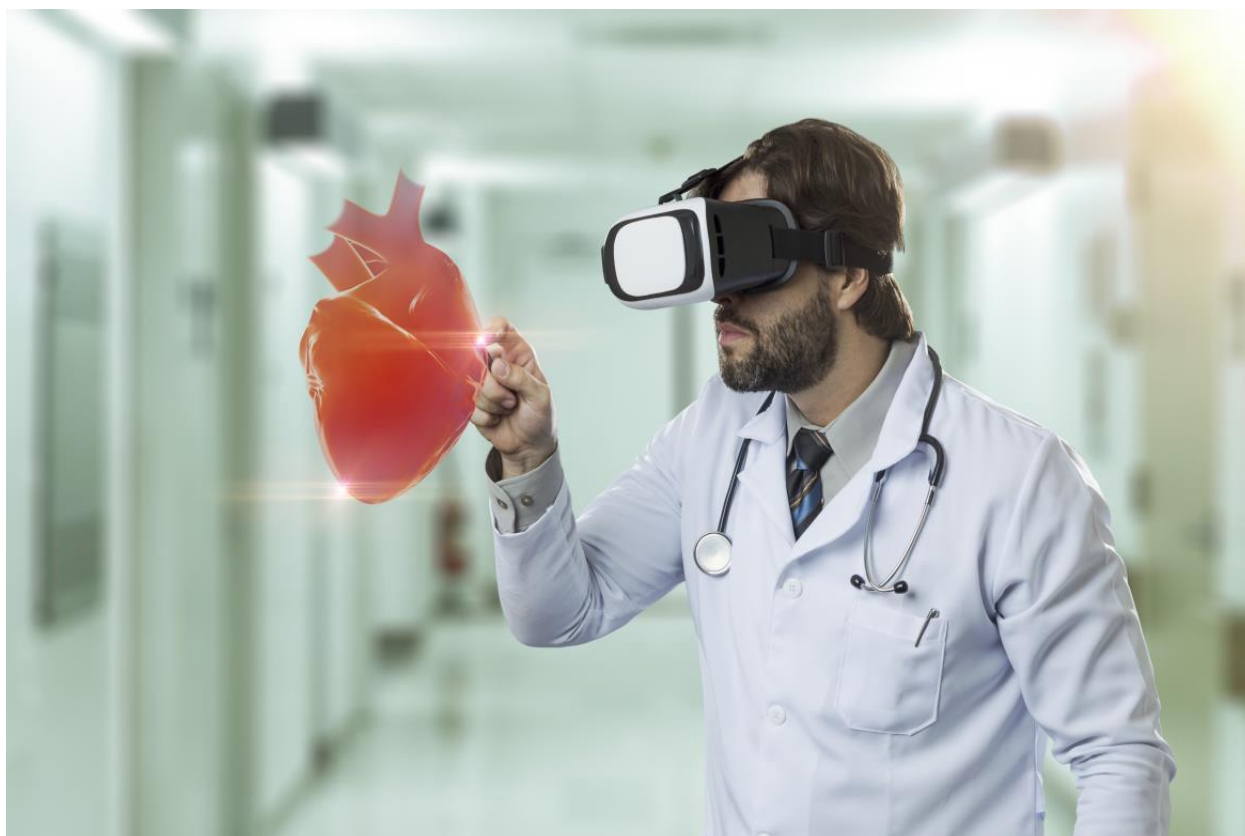


Figure 24: Example of pre-operative use of AR/VR.

The visualization 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. 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.

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 visualization 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.

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 a team specialists during treatment planning, and a better communication with the patient before a complex procedure.

Pre-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 synchronization 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.

5.3.3 Intra-operative usage

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.



Figure 25: Example of intra-operative use of AR/VR.

Nonetheless, a simple analog 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 AR (AR)/ mixed reality (MR) applications can be applied. The reasons are now listed.

First, a beam splitter is obligatory to digitize the analog 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, digitization 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 visualization 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 visualization, augmentation data can be distributed to any external 2D/3D display or remote XR visualization unit, whether VR headsets or AR glasses.



Figure 26: Example of intra-operative use of AR.

Thus, consulting external experts using XR communication during surgery becomes feasible. Both, digitization and XR technology will also allow for new image-based assistance functionalities, such as (1) 3D reconstruction and visualization of surgical areas, (2) multispectral image capture to analyse, visualize, segment, and/or classify tissue, (3) on-site visualization of blood flow and other critical surgery areas, (4) differentiation between soft tissues by blood flow visualization, (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 [169][170][171][172].

5.3.4 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. This follow-up can take place in different places, according to the needs:

- within the hospital;
- in specialized 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 here are some companies that provide VR systems for physiotherapists as well as rehabilitation structures: www.karunalabs.com, www.kinequantum.com or www.virtualisvr.com.

These type 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 [173] www.ezygain.com, that 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 [174][175]. 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 [176].

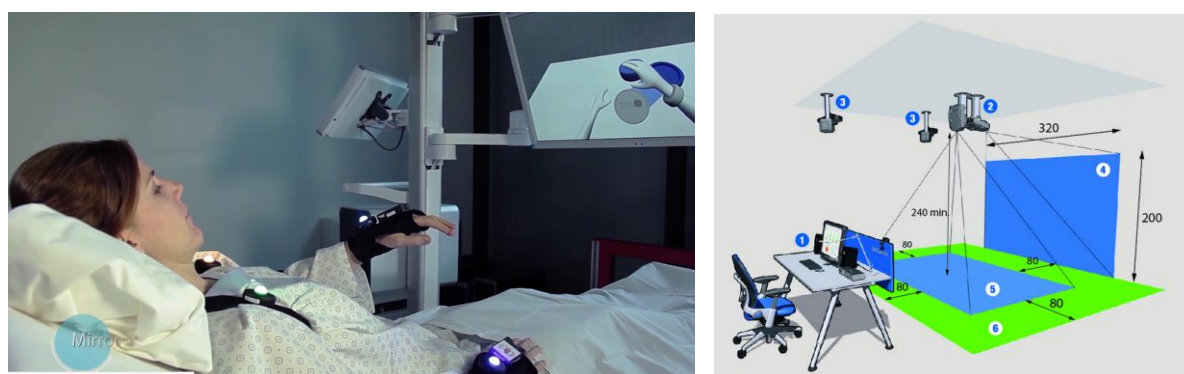


Figure 27: Mindmotion VR [175] (left) and Nirvana [176] (right).

The European research project VR4Rehab specifically focuses on enabling the co-creation of VR-based rehabilitation tools [177]. 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, the needs of patients and therapists. The project is partly funded by Interreg, a transnational funding scheme to bring European regions together.

The national project VRReha in Germany develops concepts and applications for therapy and rehabilitation [178]. 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.

Concerning the use of AR for rehabilitation, some studies have led to real AR applications, like HoloMed [179], which has been led by the Artanim motion capture centre in Switzerland. It features a solution coupling HoloLens with professional MoCap system, enabling augmented visualization of bone movements. They have developed an anatomical see-through tool to visualize 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 visualize 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 [180], which develops an Android/iOS app than 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 hospitalization. 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.

This follow-up can take place in different places, according to the needs:

- within the hospital;
- in specialized centres (rehabilitation centres);
- at home (home care).

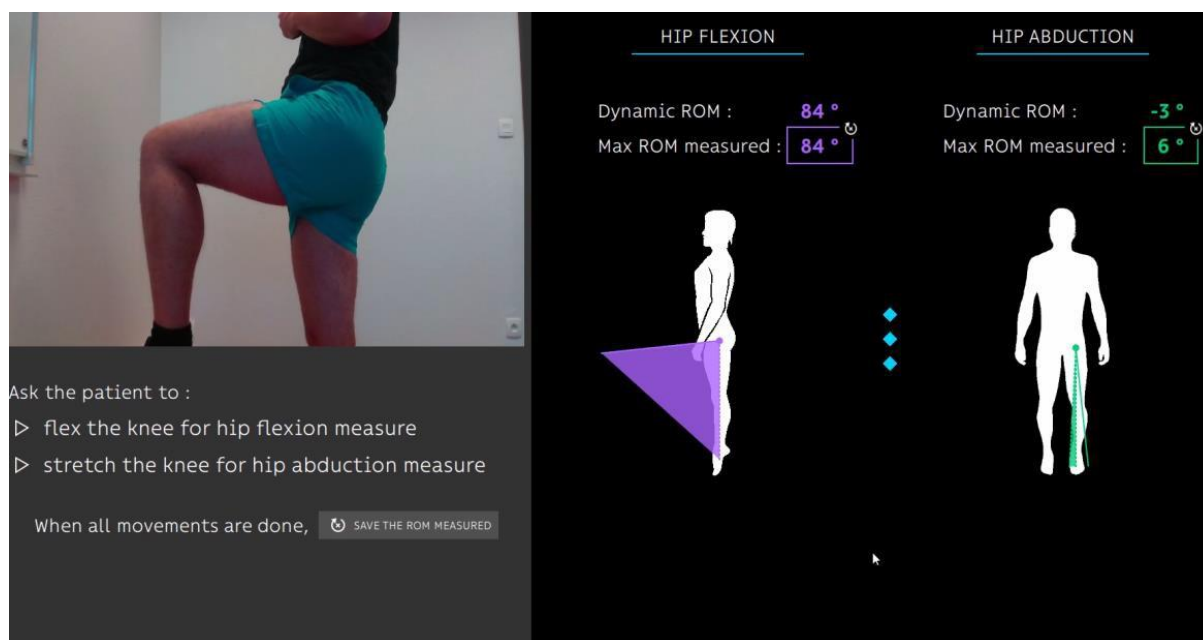


Figure 28: Example for post-operative use of VR/AR.

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.

AR can also help a nurse working on an in-home hospitalization. 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.

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.

5.3.5 References

- [168] L Chen, T Day, W Tang, NW John The 16th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2017
- [169] E. L. Wisotzky, J.-C. Rosenthal, P. Eisert, A. Hilsmann, F. Schmid, M. Bauer, A. Schneider, F. C. Uecker, 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.
- [170] 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, 5(1):373-376, 2019.
- [171] B. Kossack, E. L. Wisotzky, A. Hilsmann, P. Eisert, Local Remote Photoplethysmography Signal Analysis for Application in Presentation Attack Detection, Proc. Vision, Modeling and Visualization, Rostock, Germany, 2019.
- [172] A. Schneider, M. Lanski, M. Bauer, E. L. Wisotzky, J.-C. Rosenthal, "An AR-Solution for Education and Consultation during Microscopic Surgery", Proc. Computer Assisted Radiology and Surgery (CARS), Rennes, France, 2019.
- [173] <http://www.ezygain.com>
- [174] <https://www.mindmaze.com>
- [175] <https://www.mindmotionweb.com/>
- [176] <https://www.btsbioengineering.com/nirvana/discover-nirvana/>
- [177] <https://www.nweurope.eu/projects/project-search/vr4rehab-virtual-reality-for-rehabilitation/>
- [178] <https://www.vreha-project.com/en-gb/home>
- [179] <http://artanim.ch/project/holomed/>
- [180] <http://www.altoida.com>

5.4 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 [181]. 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. Another application of a similar type is Shangoo [182].

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 [183]. 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.

5.4.1 References

[181] <https://business.weather.com/products/max-reality>

[182] <http://www.amedia.it/shangoo/>

[183] <https://time.com/5628880/time-immersive-app-ar-vr/>

5.5 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 Encyclopaedia gives the following definition [184]:

Definition 1 of "social VR":

"(social **V**irtual **R**eality) 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 [185], 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 & 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.

5.5.1 Platforms

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

- Second Life (2003) [186][187]
- High Fidelity (2013) [188][189]
- vTime (2015) [190][191]
- Rec Room (2016) [192][193]

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 [194].

This article explains clearly that the key difference is that Second Life features a centralized 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 “centralized” 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 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 [195], shut down by Facebook on 25 Oct 2019 to make way for Facebook Horizon
- Facebook Horizon [196]
- VRChat [197]
- AltspaceVR [198]

5.5.2 Illustrations

The Figure 29 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 29: Illustration of interaction in a virtual space, here based upon the vTime platform [184].

Figure 30 shows an example scene produced via the Rec Room platform.



Figure 30: Illustration of interaction in a virtual space, here based upon the Rec Room platform [199].

5.5.3 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” [200]. 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 [201].

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.

5.5.4 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 [202]:

- 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.

5.5.5 Key applications and success factors

Gunkel et al. give 4 key use cases for “social VR”: video conferencing, education, gaming, and watching movies [203]. Furthermore, they give 2 important factors for the success of “social VR” experiences: interacting with the experience, and enjoying the experience.

5.5.6 Benefit for the environment

Collaborative telepresence has the huge potential of reducing the impact of business on the environment. Orts-Escolano et al. [204] 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 [205]. This does not count the cost on the environment. Indeed telepresence has been cited as key in battling carbon emissions in the future [206].

5.5.7 Some terminology

Although we called this section “Interaction & social”, we indicated above that the conventional, historical term is “social VR”, which can be generalized to “social XR”. We also indicated that a good “synonymous” is “collaborative telepresence”. In some papers, such as by Misha Sra [207], 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” [208].
- **Copresence**, also called “social presence” is used to refer to the sense of being in a computer generated environment with others [209][210][211][212].
- **Togetherness** is a form of human co-location in which individuals become “accessible, available, and subject to one another” [213]-. 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.”

5.5.8 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 [214]; (3) one must synchronise the interaction between all avatars and their actions. This will likely require a mix of centralized and decentralized control. Of course, this synchronization 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”.

5.5.9 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.

5.5.10 References

- [184] www.pcmag.com/encyclopedia/term/69486/social-vr
- [185] <https://ryanschultz.com/2018/07/10/what-is-the-definition-of-social-vr>
- [186] <https://secondlife.com>
- [187] https://en.wikipedia.org/wiki/Second_Life
- [188] <https://www.highfidelity.com>
- [189] [https://en.wikipedia.org/wiki/High_Fidelity_\(company\)](https://en.wikipedia.org/wiki/High_Fidelity_(company))
- [190] <https://vtime.net>
- [191] https://en.wikipedia.org/wiki/VTime_XR

- [192] <https://recroom.com>
- [193] [https://en.wikipedia.org/wiki/Rec_Room_\(video_game\)](https://en.wikipedia.org/wiki/Rec_Room_(video_game))
- [194] <https://spectrum.ieee.org/telecom/internet/beyond-second-life-philip-rosedales-gutsy-plan-for-a-new-virtualreality-empire>
- [195] <https://www.facebook.com/spaces>
- [196] www.oculus.com/facebookhorizon
- [197] <https://www.vrchat.com>
- [198] <https://altvr.com>
- [199] www.wired.com/story/social-vr-worldbuilding
- [200] <https://www.weforum.org/agenda/2019/07/these-are-the-top-10-emerging-technologies-of-2019/>
- [201] http://www3.weforum.org/docs/WEF_Top_10_Emerging_Technologies_2019_Report.pdf
- [202] M. Salimian, S. Brooks, D. Reilly, IMRCE: a Unity toolkit for virtual co-presence. SUI '18, October 13–14, 2018, Berlin, Germany. ACM ISBN 978-1-4503-5708-1/18/10...\$15.00, <https://doi.org/10.1145/3267782.3267794>.
- [203] S. Gunkel, H. Stokking, M. Prins, O. Niamut, E. Siahaan, P. Cesar, “Experiencing virtual reality together: social VR use case study”, TVX '18, June 26–28, 2018, SEOUL, Republic of Korea ACM 978-1-4503-5115-7/18/06. <https://doi.org/10.1145/3210825.3213566>.
- [204] S. Orts-Escolano, Ch. Rhemann, et al, “Holoportation: Virtual 3D teleportation in real-time”, UIST 2016, October 16-19, 2016, Tokyo, Japan. ACM 978-1-4503-4189-9/16/10..\$15.00. DOI: <http://dx.doi.org/10.1145/2984511.2984517>.
- [205] <https://www.forbes.com/sites/kenrapoza/2013/08/06/business-travel-market-to-surpass-1-trillion-this-year/>
- [206] <https://www.scientificamerican.com/article/can-videoconferencing-replace-travel/>
- [207] M. Sra, A. Mottelson, P. Maes, “Your place and mine: Designing a shared VR experience for remotely located users”, DIS '18, June 9--13, 2018, Hong Kong. <ACM. ISBN 978-1-4503-5198-0/18/06. . . \$15.00. DOI: <https://doi.org/10.1145/3196709.3196788> .
- [208] M. 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 364, 1535 (2009), 3549--3557. DOI: <http://dx.doi.org/10.1098/rstb.2009.0138>
- [209] F. Biocca, Ch. Harms, (2002), “Defining and Measuring Social Presence: Contribution to the Networked Minds Theory and Measure”, Proceedings of PRESENCE 2002 (2002), 7--36.
- [210] J. Short, E. Williams, and B. Christie, (1976), “The Social Psychology of Telecommunications”, Wiley.
- [211] N. Durlach, M. Slater, (2000), “Presence in Shared Virtual Environments and Virtual Togetherness”, Presence, Vol. 9, 2 (April 2000), 214--217. DOI: <http://dx.doi.org/10.1162/105474600566736>

- [212] R. Schroeder, (2002), “Copresence and Interaction in Virtual Environments: An Overview of the Range of Issues”, In Presence 2002: Fifth international workshop. 274--295.
- [213] E. Goffman, (2008), “Behavior in Public Places”, Free Press.
- [214] <https://vrtogether.eu/>

5.6 Conclusion

The section about XR applications focused on the main domains where XR tends to be a promising technology with significant potential of growth. Some other application domains are not yet listed although their relevance is recognized as well. The missing areas are as follows:

- Agriculture and food;
- Art and heritage, with topics such as virtual museum;
- Education, especially for basic and higher education in any science domain;
- Entertainment and sports;
- Security and sensing, with topics like crime scene investigation, emergency management, environment surveillance and protection, firefighting, and humanitarian operations;
- Transportation, with focus on areas such as logistics, and the road, rail, water, air, and space transportation industry.

In the next update of this deliverable, these missing application domains will be added.

6 Standards

Various Standards Developing Organizations (SDO) are directly addressing AR specific standards, and others are focusing on technology related to AR. We present next the standardization 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) [215] aiming at developing 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 visualized 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:

- ETSI GR_ARF001 v1.1.1 published in April 2019 [216], provides an overview of the AR standards landscape and identifies the role of existing standards relevant to AR from various standards setting organizations. 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 [217], 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.

6.1.2 Khronos

OpenXR™ [218] defines two levels of API interfaces that a VR platform's runtime can use to access the OpenXR™ ecosystem. Applications and engines use standardized interfaces to interrogate and drive devices. Devices can self-integrate to a standardized driver interface. Standardized 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 optimize 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 [219] 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 standardized 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 Organization (SDO) addressing media compression and transmission. Of course, MPEG is well known for the various 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) [220] is a technical report defining the scope and key concepts of mixed and augmented reality, the relevant terms and their definitions, and a generalized 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) [221] 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 Mark-up Language” (ARML 2.0) [222] which is a 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 visualization. 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 [223] 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 visualization. A working draft has been published in October 2019.

6.1.7 References

- [215] <https://www.etsi.org/committee/arf>
- [216] https://www.etsi.org/deliver/etsi_gr/ARF/001_099/001/01.01.01_60/gr_ARF001v010101p.pdf
- [217] https://www.etsi.org/deliver/etsi_gr/ARF/001_099/002/01.01.01_60/gr_ARF002v010101p.pdf
- [218] <https://www.khronos.org/openxr>
- [219] <https://www.openarcloud.org/>
- [220] <https://www.iso.org/standard/30824.html>
- [221] <https://www.iso.org/standard/69465.html>
- [222] <https://www.opengeospatial.org/standards/armi>
- [223] <https://www.w3.org/blog/tags/webxr/>

6.2 XR related standards

6.2.1 Khronos

OpenVX™ [224] is an open-royalty-free standard for cross platform acceleration of computer vision applications. OpenVX™ enables performance and power-optimized 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 optimized 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 [225] 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 visualization functions. Developers can leverage the power of OpenGL™ across all popular desktop and workstation platforms, ensuring wide application deployment.

WebGL™ [226] 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 recognize 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.

6.2.2 MPEG

MPEG-I (ISO/IEC 23090) [227] 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) [228] 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 component of a 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) [229] 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 [230] 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) [231] 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 Open Geospatial Consortium

OGC GML [232] 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 [233] 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 [234] 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 localization, indoor m-Commerce, emergency control, etc. IndoorGML does not provide spaces geometry but it can refer to data described in other format like CityGML, KML or IFC.

OGC KML [235] is an XML language focused on geographic visualization, including annotation of maps and images. Geographic visualization 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.

6.2.4 W3C

GeoLocation API [236] is a standardized 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.

6.2.5 References

- [224] <https://www.khronos.org/opencv/>
- [225] <https://www.khronos.org/opengl/>
- [226] <https://www.khronos.org/webgl/>
- [227] <https://mpeg.chiariglione.org/standards/mpeg-i>
- [228] <https://mpeg.chiariglione.org/standards/mpeg-v>
- [229] <https://mpeg.chiariglione.org/standards/mpeg-4/3d-graphics-compression-model>
- [230] <https://mpeg.chiariglione.org/standards/mpeg-7/compact-descriptors-visual-search>
- [231] <https://mpeg.chiariglione.org/standards/mpeg-u>
- [232] <https://www.opengeospatial.org/standards/gml>
- [233] <https://www.opengeospatial.org/standards/citygml>
- [234] <https://www.opengeospatial.org/projects/groups/indoorgmlswg>
- [235] <https://www.opengeospatial.org/standards/kml>
- [236] <https://www.w3.org/TR/geolocation-API/>

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.

VR/AR/MR technologies enrich media and content production in entertainment (**ACTION-TV**, **DBRLive**, **ImmersiaTV**, **Immersify**, **POPART**, **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**, **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 (**AUGGMED**, **ASSISTANCE**, **CybsPEED**, **E2DRIVER**, **LAW-TRAIN**, **REVEAL**, **TARGET**, **WEKIT**, 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**. 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 (**AlterEgo, CAPTAIN, HOLOBALANCE, KINOPTIM, MetAction, OACTIVE, 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: **RASimAs, SMARTsurg, SurgASSIST, 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 personalized and neuroadaptive VR tool is developed for diagnosis of psychopathological symptoms.

Several projects target or relate to the design and engineering fields (**DIMMER, EASY-IMP, FURNIT-SAVER, MANUWORK, MINDSPACES, OPTINT, SPARK, ToyLabs, TRINITY, V4Design**). **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: **BIM4EEB, EDUSAFE, ENCORE, INSITER, PACMAN, PreCoM, PROPHECY**. With **INSITER**, AR with access to a digitized 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, E2DRIVER, I-VISION, RETINA, SimuSafe, 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.

Technology to support research questions is used in projects such as **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**, **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**, **REALITY**, **See Far**, **WEAR3D**

Displays: **ETN-FPI**, **LOMID**

Sound: **BINCI**, **Sound of Vision**, **SoundParticles**, **VRACE**

Haptics: **DyViTo**, **H-Reality**, **MULTITOUCH**, **TACTILITY**, **TouchDesign**

Camera development: **DBRLive**, **FASTFACEREC**

Computer graphics, animation, and related fields: **ANIMETRICS**, **FunGraph**, **RealHands**, **REALITY CG**, **VirtualGrasp**.

Another large area of development concerns human-robot interaction using XR (**CogIMon**, **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>

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: <http://www.euromov.eu/alterego/>

ANIMETRICS: <https://cordis.europa.eu/project/rcn/100399/reporting/en>

AUGGMED: <https://cordis.europa.eu/project/rcn/194875/factsheet/en>

ASSISTANCE: <https://cordis.europa.eu/project/rcn/222583/factsheet/en>

BIM4EEB: <https://www.bim4eeb-project.eu/the-project.html>

BINCI: <https://binci.eu/>
CAPTAIN: <https://cordis.europa.eu/project/rcn/211575/factsheet/en>
CCFIB: <https://cordis.europa.eu/project/rcn/108306/reporting/en>
Clean Sky: <https://www.cleansky.eu/>
CO3: <https://cordis.europa.eu/project/rcn/218757/factsheet/en>
CogIMon: <https://www.cogimon.eu/>
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>
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>
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/>
H-Reality: <https://cordis.europa.eu/project/rcn/216340/factsheet/en>
Human Brain Project: <https://www.humanbrainproject.eu/en/>
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>
IN-Fo-trace-DG: <https://cordis.europa.eu/project/rcn/214885/factsheet/en>
INSITER: <https://www.insiter-project.eu/Pages/VariationRoot.aspx>
INTERACT: <https://cordis.europa.eu/project/rcn/106707/reporting/en>

IRIS: <http://www.iris-interaction.eu/>
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: <http://www.project-leanautomation.eu/>
LOMID: <http://www.lomid.eu/>
MANUWORK: <http://www.manuwork.eu/>
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://neurorobotics.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>
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>
PLATYPUS: <https://platypus-rise.eu/>
POPART: <http://www.popartproject.eu/>
PreCoM: <https://www.precom-project.eu/>
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>
RECONTEXT: <https://cordis.europa.eu/project/rcn/98096/factsheet/en>
REPLICATE: <http://www.replicate3d.eu/>
REVEAL: <http://revealvr.eu/>
RETINA: <http://www.retina-atm.eu/>
Scan4Reco: <https://www.scan4reco.eu>
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/>
SOCIAL LIFE: <https://cordis.europa.eu/project/rcn/94363/factsheet/en>
SoftPro: <https://www.softpro.eu/>

Sound of Vision: <http://www.soundofvision.net/>
SoundParticles: <https://cordis.europa.eu/project/rcn/224745/factsheet/en>
SpaceCog: <http://www.spacecog.eu/project/index.html>
SPARK: <http://www.spark-project.net/>
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>
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://projects.tuni.fi/trinity/about/>
TouchDesign: <https://cordis.europa.eu/project/rcn/214715/factsheet/en>
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/en/>
VRACE: <https://cordis.europa.eu/project/rcn/220824/factsheet/en>
VRMIND: <http://www.vrmind.co/>
VRTTogether: <https://vrtogether.eu>
WEAR3D: <https://cordis.europa.eu/project/rcn/111025/factsheet/en>
WEKIT: <https://wekit.eu>
WorkingAge: <https://www.workingage.eu/>
WrightBroS: <https://cordis.europa.eu/project/rcn/218206/factsheet/en>

8 Conclusion

This landscape report provides a recent and quite complete overview of the area of XR technologies. The market analysis is based on the latest figures available as of autumn 2019. 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/

This document represents the first version of this landscape report. Therefore, some application domains such as entertainment and cultural heritage have not been considered in detail. They will be included in the next version of the report.