

# **Virtual Reality/Augmented Reality White Paper (2017)**

**China Academy of Information and  
Communications Technology (CAICT)**

**Huawei Technologies Co., Ltd**

**December 2017**

---

## **COPYRIGHT NOTICE**

---

**This white paper is copyright of CAICT and is protected by applicable laws. The reproduction of text or opinions contained in this white paper, by forwarding, excerpts, or any other means, requires you to specify "source: CAICT". Violation of this notice will render you liable for legal proceedings.**

## **Acknowledgement**

Under the guidance of the department of electronics and information, Ministry of Industry and Information Technology, this white paper was jointly prepared by China Academy of Information and Communications Technology (CAICT) and Huawei Technologies Co., Ltd. During the preparation, CAICT gained professional support from the following parties: the iLab and the 2012 Labs of Huawei Technologies Co., Ltd.; the Innovation Business Department and the virtual reality content creation center of BOE Technology Group Co., Ltd.; GoerTek; China Telecom Shanghai Research Institute; Whaley VR; air360china; and HiScene.

Due to factors such as limited time, knowledge, and the still developing industry, there might be mistakes in this white paper. Please feel free to correct us anytime.

## **Preface**

Virtual Reality (VR)/Augmented Reality (AR) is a key field of the new-generation information and communications technology, which features large application space, huge industry potential, and wide technical span. It has great significance to driving the transformation and upgrade of core components, extensive smart devices, network transmission devices, cloud computing devices, telecom services, and software and industry information services. VR/AR has a long technical history. However, it is an emerging industry and the development of the VR/AR industry and technologies has not been fully finalized. From the perspective of technology, near-eye display, rendering processing, perception and interaction, network transmission, and content creation are forming the primary system. From the perspective of industry structure, the VR industrial system depends on components/equipment, tools/platforms, and contents/applications. When compared with the relatively mature intelligent terminals, VR has a similar industrial system. However, significant differences also exist.

In this white paper, the following aspects are systematically described: VR concepts, levels of VR experience, technology architecture, segment systems, and technology development roadmaps. In addition, a detailed segment map of China's VR industry was produced based on China's local VR enterprises and their roles in the key points of the VR industry chain. The white paper also analyzes the industry ecosystem and development trends, based on which development suggestions are offered to telecom operators.

## Contents

<b>1 Background</b>	<b>2</b>
1.1 Evolution of the VR Concept and Its Form Divisions	2
1.2 VR Rises Because of Lowered Threshold, Focused Capital, and Policy Support	4
1.3 VR Becomes the Development Focus in the Mobile Internet Evolution Towards the Artificial Intelligence Era	6
<b>2 Key Technology Trend</b>	<b>7</b>
2.1 Preliminary Formation of the VR Technology Architecture	7
2.2 Immersive Experience Improvement and Dizziness Control: Trends of Near-Eye Display Technologies	9
2.2.1 High Angular Resolution and Wide FOV: Key to Improve Immersive VR Experience	9
2.2.2 Dizziness Control: a Challenge to VR Near-Eye Display	10
2.2.3 AMOLED, LCOS, and OLEDoS: Leading Near-Eye Screen Technologies	10
2.3 Tracking and Positioning, Environment Understanding, and Multi-Channel Interaction: Focuses of Perception and Interaction Technologies	12
2.3.1 Tracking and Positioning: a Core Technology in the Field of VR and AR Perception and Interaction	12
2.3.2 Environment Understanding Based on Machine Vision: Trend of AR Perception and Interaction	13
2.3.3 Multi-Channel Interaction: Trend of VR Perception and Interaction	13
2.4 High Bandwidth, Low Latency, Large Capacity, and Service Isolation: Trends of Network Transmission Technologies	15
2.4.1 New Wi-Fi, Large-Capacity PON, and 5G: Trends of Access Network Technologies for VR	15
2.4.2 Simple Architecture, Smart Pipe, Multicast On Demand, and Network Isolation Are Development Trends of Bearer Networks for VR Services	16
2.4.3 A Low-Latency Data Center Is Key to VR Computing Capability Cloudification	18
2.4.4 Projection, Coding, and Transmission Technologies Are Key to Optimizing the VR Network Performance	19
2.4.5 VR-oriented Network O&M and Evaluation Are Important for Improving User Experience	19
2.5 Combination of Enhanced Rendering Algorithms and Capabilities Are Trend of Rendering Processing Technologies	21
<b>3 Industry Ecosystem Trend</b>	<b>24</b>
3.1 Components/Devices, Tools/Platforms, and Content Applications Dominate in the VR Industry Ecosystem	24
3.2 Perception and Interaction and Content Creation Become Development Focuses of VR in the Next Phase	25
3.3 IPR Competition Represents the Industry Development Trend	27
3.3.1 Global IPR Development Entered a Rapid Growth Stage, and the Growth Rate in China Is Notable	27
3.4 Emergence of the VR/AR+ Era	28
3.4.1 In VR+ Industries, VR is the key in smart manufacturing	29
3.4.2 In VR+ Medicine, VR is applied in surgery training/guidance, psychiatry, and rehabilitation.	30
3.4.3 In VR+ Games/Social Networking, VR gaming will become the biggest drive of the current VR industry.	30
3.4.4 In VR+ Movie/Live Broadcast, VR becomes a new form of display.	31
<b>4 Recommendations on the Development of the VR Industry for Operators</b>	<b>31</b>

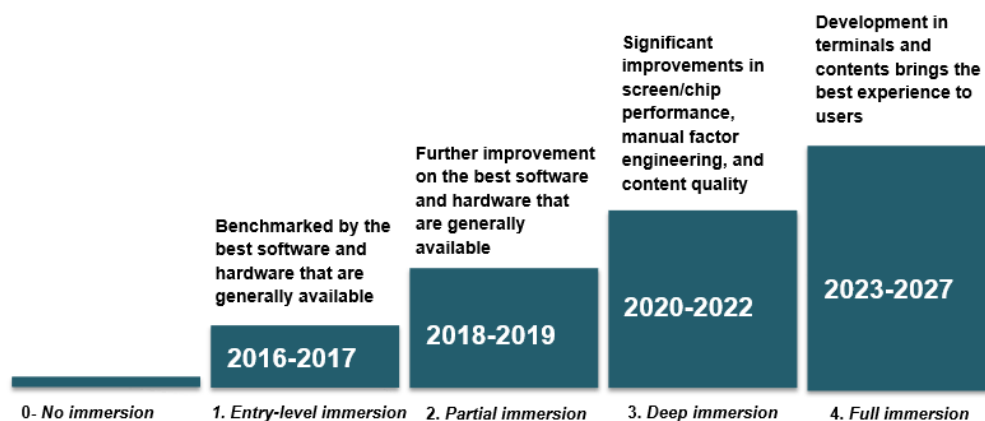
4.1 Operators Are One of the Core Links of the VR Industry Chain, Providing Assurance for VR Service Experience and Enabling Service Popularization	32
4.2 Operators Drive the Evolution of VR Service Forms: From Virtual Cinemas to Live VR, and Then to Cloud-based VR Services	32
4.3 VR Service Development Recommendations for Operators	33

# 1 Background

## 1.1 Evolution of the VR Concept and Its Form Divisions

The industry's definition for VR evolved from terminals to immersive experience. With the continuous development of technologies and the industry ecosystem, the VR concept will keep evolving. Therefore, the industry's discussion on VR is no longer limited to the form factor of terminals or how VR is realized. Instead, the focus now is experience, emphasizing key technologies, industry ecosystem, and integrated innovations in the application aspects. In this white paper, it is understood that VR/AR uses next-generation information and communications technologies in near-eye display, perception and interaction, rendering processing, network transmission, and content creation to build a new industry across terminals, channels, and the cloud, satisfy customers' requirements for an immersive experience, and drive the expansion and upgrade of information consumption and the integrated innovation in traditional industries. The improvement of an immersive experience depends on breakthroughs and progress of relevant technologies, which is a staged evolution process. Therefore, the VR service development is divided into the following stages, with different development stages corresponding to different experience requirements.

**Figure 1-1** Levels of VR immersive experience



Source: CAICT

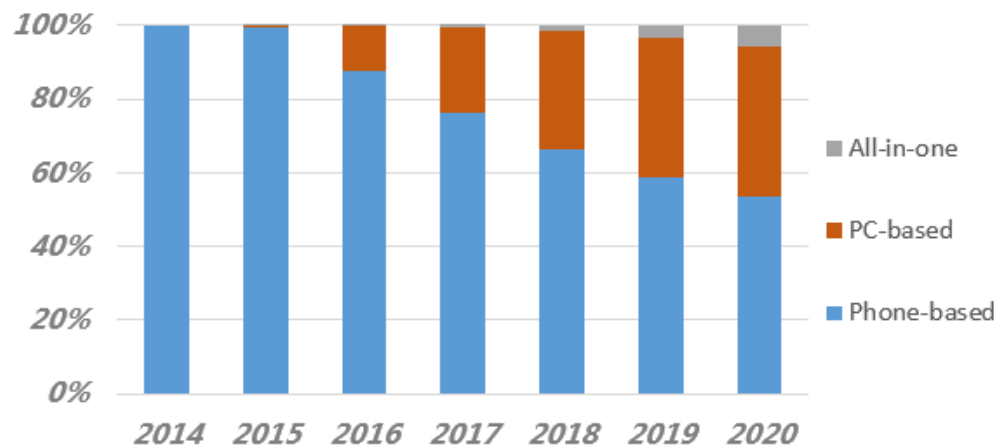
**Figure 1-2** VR immersive experience–network requirements

Level \ Specification	Video Resolution Threshold	Monocular Resolution Threshold	Typical Network Bandwidth Requirement	Typical Network RTT Requirement
<b>Entry-level immersion (EI)</b>	Full view 4K 2D video (Full screen resolution 3840*1920)	960*960 FOV 100° (~ 20 PPD)	20 to 50 Mbps	< 40 ms
<b>Partial immersion (PI)</b>	Full view 8K 2D video (Full screen resolution 7680*3840)	1920*1920 FOV 110° (~ 20 PPD)	50 to 200 Mbps	< 30 ms
<b>Deep immersion (DI)</b>	Full view 12K 2D video (Full screen resolution 11520*5760)	3840*3840 FOV 120° (~ 30 PPD)	200 Mbps to 1 Gbps	< 20 ms
<b>Fully Immersion (FI)</b>	Full view 24K 3D video (Full screen resolution 23040*11520)	7680*7680 FOV 120° (~ 60 PPD)	2 to 5 Gbps	< 10 ms

Source: Huawei iLab

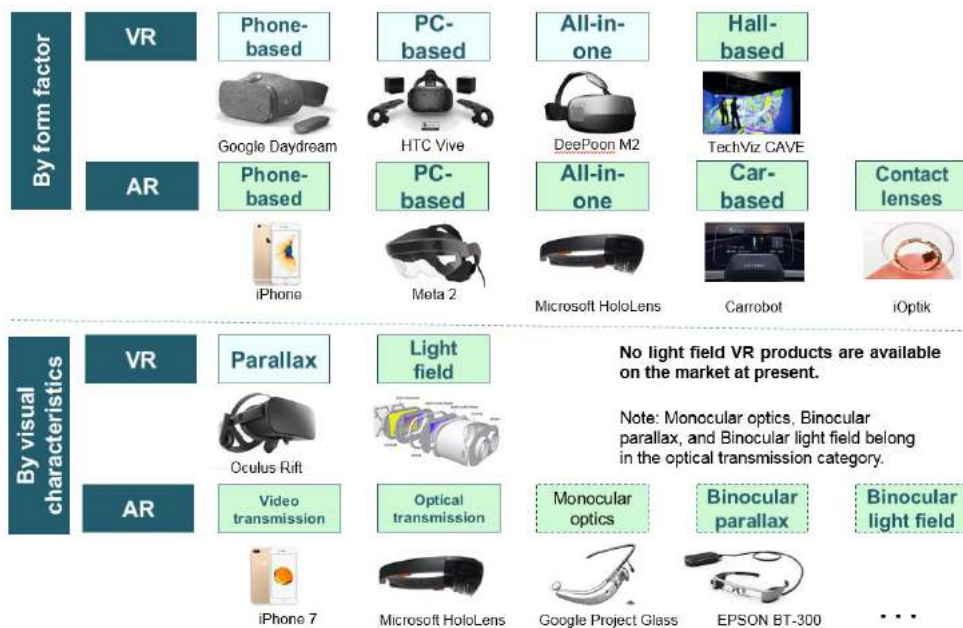
VR terminals evolve from one to many form factors, and from separated to integrated. In terms of terminal form factors, the mobile phone type has become the major terminal type at this stage. Globally, phone-based VR accounted for nearly 90% in 2016. By 2020, it is estimated that the penetration rate of PC-based and all-in-one VR terminals will rise to about 50% (Figure 1-3). As ICT giants such as Google, Facebook, and Apple held their 2017 global developer conferences (I/O, F8, and WWDC respectively), phone-based AR has become the mainstream for the mass market. Meanwhile, PC-based and all-in-one AR terminals such as Meta2 and Hololens dominate the enterprise market. In addition, under the influence of the self-driving cars and Internet of Vehicles (IoV) trends, built-in display-at-a-glance AR has become the emerging field. Ultimately, the forward-looking contact lenses form factor represents the industry's final expectation for AR design. In terms of terminal functions, VR generally includes AR. In the early stages, AR was usually discussed under the VR framework. However, with the continuous effort put into AR by the industry, AR and VR are gradually being separated. Specifically, VR and AR are mutually independent, with similarities in key components and terminal form factor, but differences in key technologies and application fields. Unless stated otherwise, this white paper discusses VR and AR in a general sense. VR uses isolated audio and video contents for an immersive experience, which poses high requirements on image quality. AR integrates virtual information to the real environment seamlessly, which poses high requirements on perception and interaction. In addition, VR focuses more on mass markets such as games, videos, live broadcast, and social media, whereas AR focuses more on vertical applications such as in different industries and for the military. As the technologies develop, VR and AR will be integrated.

**Figure 1-3** Global penetration rate of major VR terminal form factors



Source: HIS

**Figure 1-4** Categories of VR products



Source: CAICT

## 1.2 VR Rises Because of Lowered Threshold, Focused Capital, and Policy Support

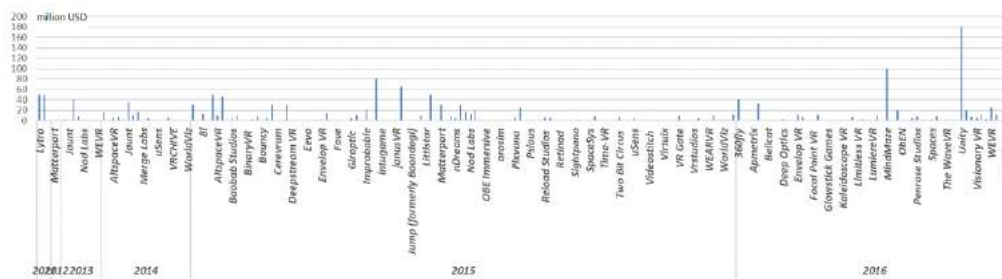
The hardware threshold is significantly reduced. Since the advent of VR devices in 1962 to today and with the popularization of smart phones, the hardware costs for VR are significantly reduced. The prices of VR devices dropped from tens of thousands of dollars to a few hundred dollars, which is mainly due to the development of optoelectronics and microelectronics. In terms of optoelectronics, the displays for VR transformed from CRT to TFT-LCD/AMOLED, with ever decreasing volume and weight of the screens, resolution improved to FHD+, and response time shortened to microseconds. In terms of microelectronics,



the popularization of low-cost SOC chips and visual processing units (VPUs) has become the development hotspot of VR in the integrated circuit field.

Capital investment direction is increasingly focused. After the Google Glass warm-up, Facebook acquired Oculus for USD 2 billion in 2014, sending a clear signal in the industry. Since then, global capital has been heavily invested in the VR field. The major ICT giants actively propose suggestions on development strategy, and many tech start-up companies are emerging. According to Digi-Capital statistics, the global investment on VR startups in 2016 reached USD 2.3 billion (excluding acquisition), with a year-on-year increase of over 200%. Geographically, China and the US have become the key development regions of the VR industry. VR startups headquartered in China and the US gained 20% and 60% respectively of the global investment on VR. In terms of investment fields, according to IHS statistics, development tools, games, and video contents gained 18%, 16%, and 11% investment respectively, ranking the top 3 in global VR investment. This reflects the changing factors that hinder VR popularization. That is, after the VR hardware threshold is lowered, the industry pays more and more attention to contents and applications such as games and videos, as well as exclusive development tools.

**Figure 1-5** Global investment on VR around 2014



Source: HIS

Developing the VR industry has become the national strategy of countries around the world. The US government listed VR as one of the key fields supported by the National Information Infrastructure (NII) in the 1990s. The US Department of Defense attaches great importance to R&D and applications of VR, giving VR key support in the aspects of performance evaluation of weapon systems, equipment operating training, and commanding of large-scale military exercises. The US Department of Energy developed the *Long-term Nuclear Technology Research and Development Plan* in 2000, in which the importance of developing, applying, and verifying VR technologies was clearly pointed out. Several US congress representatives jointly announced the VR guidance team in 2017 to ensure support and encouragement for the VR industry from the congress. In addition, the US has also established research projects on VR. For example, the Department of Health and Human Services and the Department of Education carried out pilots and demonstrations of VR on mental disease and primary and secondary school education respectively. The European Union began funding VR in the 1980s. In the *Horizon 2020* released in 2014, the VR funding reached tens of millions of Euros. Japan released the *Innovation 2025 Strategy* for technology development planning in 2007, and *Science, Technology, and Innovation Comprehensive Strategy 2014–Innovation Bridge to the Future* in 2014. Both documents defined VR as the key technology innovation direction. South Korea established a special fund of about CNY 240 million, funding nine emerging technologies as key national development fields, including VR, self-driving, and artificial intelligence. In addition, the Ministry of Science, ICT and Future Planning of South Korea plans to invest about CNY 2.4 billion from 2016 to 2020 to develop its national VR industry. The emphasis is to ensure the original technology R&D and industrial ecosystem improvements, and to reduce the current gap of two years between South Korea and the US in VR to half a year. Overall, VR development in the US is based mainly on enterprises and the government provides the platform. The US government attaches great importance to model applications of VR in various fields. To summarize, the European Union, South Korea, and Japan value top-level design and R&D of new technologies. They established special funds to guide the industry development of VR. In China, the government encourages the development of VR, listing it in multiple national strategy

documents, such as the *Thirteenth Five-Year Plan* of information, Made in China 2025, and Internet+. The Ministry of Industry and Information Technology, the National Development and Reform Commission, the Ministry of Science and Technology, the Ministry of Culture, and the Ministry of Commerce all released policies concerning VR. In addition, the provincial and municipal local governments proactively build industrial parks and labs to promote the development of local VR industries. By the end of 2016, nearly 20 provinces and municipalities in China started deploying the VR industry.

**Figure 1-6** VR industry development layout in some provinces and cities in China



### 1.3 VR Becomes the Development Focus in the Mobile Internet Evolution Towards the Artificial Intelligence Era

The development focus of the new-generation information and communication industry will shift from Mobile First to AI First, which has become commonly acknowledged by tech giants. Over the last two years, Google and Facebook both decided to make AI the focus of their future strategies at their global developer conferences.

The link between VR and mobile Internet mainly lies in terminal form factor and application software. VR terminals will continue to be developed in various form factors, such as phone-based, PC-based, and all-in-one, and will not be limited to a certain form factor. At present, the main form factor for VR development is phone-based. According to HIS, phone-based VR accounted for more than 80% of market share in the existing global VR market, and is expected to continue its dominance into 2020. Mobile phones are the primary VR platform medium for tech giants. For example, Facebook claimed mobile phones to be their primary AR experience platform at its 2017 developer conference, and Google's Daydream and Tango projects are both centered on mobile phones. Apple aims to build the world's largest AR platform using mobile phones, extending mobile phones instead of replacing them. In terms of applications, mobile apps are starting to become VR/AR-ready. According to Nielsen statistics, most of the top 10 apps downloaded in the US are VR/AR ready. In addition, VR applications bring challenges for mobile phone battery life and loads on the cloud. For example, Pokémon Go (a well-known AR game) attracted a huge number of users after it went online, and the load on the cloud server was ten times that expected in worst case scenarios.

The link between VR and AI lies in rendering processing, perception interaction, and universal AI. AI drives the key technology development of VR mainly from the perception and interaction and rendering processing aspects. In the perception interaction aspect, tech giants focus on scenario splitting, identification, locating, and reconstruction based on AI. For example, Google Lens uses AI to identify contents in the pictures, and Tango and the maps team use ambient 3D modeling to achieve indoor SLAM. In the rendering processing aspect, rendering requires a huge amount of computing resources, and is unsatisfactory in image noises and rendering time. Rendering technologies based on deep learning can

greatly improve image quality and shorten rendering time. On the other hand, VR drives AI development mainly from the universal AI aspect. AI tends to evolve from specific applications to universal applications, which depends on the accumulation of a vast amount of training data. However, it is difficult to collect such volumes of data with certain data structures in the real world. Therefore, well-known startups such as Improbable and OpenAI are dedicated to using virtual worlds instead of the real world for collection of training data. By building a complex and large-scale virtual world with a large population, various modeling experiments can be performed, such as the prevention and control of epidemic diseases and impacts of important policies including real estate policies. In this way, a large amount of training data can be acquired.

## 2 Key Technology Trend

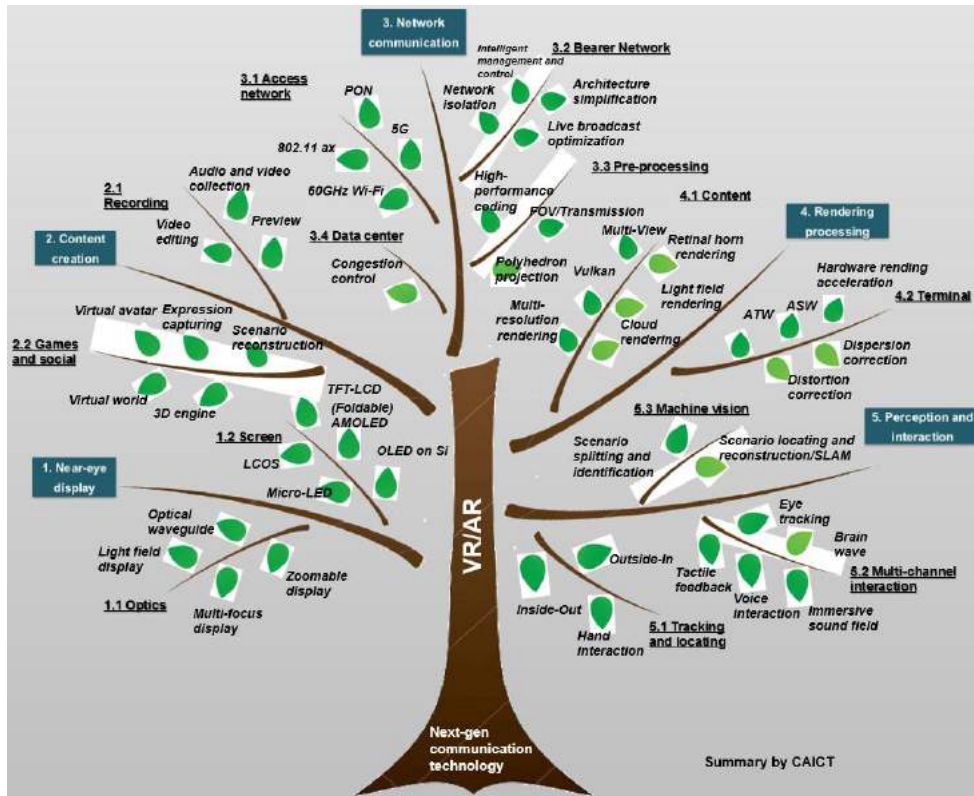
### 2.1 Preliminary Formation of the VR Technology Architecture

Multiple categories of technologies are involved in the VR technology architecture, as shown in the following figure. Because VR involves multiple fields and technologies, and is still in the early stage of development without a fixed development path, the definition and categories of VR technologies are still unclear. In this white paper, we proposed the VR technology system and the relevant references according to its development characteristics. As shown in Figure 2-1, the five rows indicate five technology fields: near-eye display, rendering processing, perception and interaction, network transmission, and content creation. The two columns indicate the key components/devices and content production tools and platforms. Both aspects support the development of VR. As shown in Figure 2-2, the technology fields in the rows can be divided into three hierarchies. The first hierarchy is the five rows correspondingly, and the second and third hierarchies are more detailed technologies.

Figure 2-1 VR technology architecture



**Figure 2-2** Hierarchy of VR technologies



Various references help identify VR key technologies. In terms of experience, human-machine interaction is the core characteristic of VR, whereas mobile phones are designed for communications. Without human-machine interaction, VR is nothing more than head-mounted TVs/mobile phones. In terms of applications, Goldman Sachs predicts that revenue from global VR games/social and videos/live broadcasting will account for 60% of the total revenue of the VR market in 2025. Therefore, content production technologies related to these two categories, such as development engines and video capture tools, have become the key. In terms of cost structure, for host-based VR headsets, the screen is the most expensive of all components, accounting for about one third of the total cost. In terms of innovation, integrated innovation has become the characteristic of VR development. Taking the Motion-to-Photons (MTP) latency control technology as an example, latency must be reduced in all the processes, including sensing and collection, computing and rendering, transmission and communication, and display and feedback, to satisfy the 20 ms latency threshold agreed by mainstream opinions in the industry. In terms of popularization time, refer to the annual technology maturity curve released by Gartner. In terms of intellectual property (IP), according to statistics published by CAICT IP center, display, interaction, modeling, locating, camera, headset, and application have become the hotspot aspects of patent application by 2017 H1 (index including but not limited to China, the US, Japan, and South Korea). In terms of capital market, content production and perception interaction technologies have become the investment hotspots. For example, the largest single investment in 2016 was in the field of development engines. In terms of strategy, ICT giants such as Facebook, Google, Apple, Microsoft, and Intel have all proactively proposed their VR development strategies. Mark Zuckerberg set VR as one of the three major technology development directions for Facebook in the next decade. Tim Cook announced that Apple will build the largest AR development platform in the world. In terms of supply chain security, centralization of key components affects the supply chain stability of VR enterprises. For example, in the AMOLED screen market, Samsung has a market share of more than 95%.



**Figure 2-3** Key fields of global VR intellectual properties

Source: CAICT

## 2.2 Immersive Experience Improvement and Dizziness Control: Trends of Near-Eye Display Technologies

### 2.2.1 High Angular Resolution and Wide FOV: Key to Improve Immersive VR Experience

High angular resolution display becomes a core technology to improve immersive experience of VR near-eye display. Focusing on improving a single performance indicator, instead of balancing all technical specifications, seems to be contradictory to the intrinsic characteristic of integrated innovation in the VR field. However, because VR head-mounted displays require higher definition, high screen resolution (and high aperture ratio) becomes key to reduce the screen-door effect. There is an urgent requirement in VR for 4K+ resolution, which is not required by smartphones. VR features 360-degree panoramic display, which makes pixel per degree (PPD) a core technology specification that is better suited for measuring the pixel density of VR near-eye display than pixel per inch (PPI). With the popularity of 4K screens in the future and a balanced design between Field of View (FOV) and resolution, the monocular PPD is expected to rise from 15 to over 30 by 2020.

Wide FOV display becomes a core technology to improve immersive experience of AR near-eye display. AR features man-machine interaction with realities and is often used to display suggestive, supplementary content based on realities. High interactivity (including a wide FOV) instead of high image quality, for example, high resolution, is a key trend of AR display technologies. However, the FOV of most model AR products ranges from 20 degrees to 40 degrees due to volume and weight restrictions. Improving AR visual interaction performance, including the FOV, has become an industry trend since screen technologies such

as organic light-emitting diode-on-silicon (OLEDoS) and microprojection technologies such as lead computing optical sight (LCOS) are available. Waveguide and light field display and other new optical system design technologies replace traditional ones such as expanding grating to become the focus of technology giants, including Google and Microsoft.

### 2.2.2 Dizziness Control: a Challenge to VR Near-Eye Display

To develop near-eye display technologies based on binocular visual characteristics is to capture the technological high ground of VR dizziness control. How VR makes users dizzy remains unknown. Universities in China, including Beijing Institute of Technology, have launched in-depth research into VR dizziness in terms of content design, individual differences, and VR software and hardware. From the perspective of binocular visual characteristics, the following sources of VR dizziness are accepted across the industry.

- **Image quality**  
Low-quality images, such as screen doors, streaking, and flickering, lead to visual fatigue, which can easily produce dizziness. Improving screen resolution, response time, refresh rate, and motion-to-photon (MTP) latency is the trend.
- **Conflicts between vision and other sensory channels**  
Strengthening the coordination between vision and auditory sense, tactile sense, vestibular system, and motion feedback is key to this issue. Besides non-mainstream solutions, such as vestibular stimulation and medicines, HTC Vive, Room Scale of Oculus, and the omnidirectional treadmill of Virtuix Omni are mainly used to reduce dizziness caused by such conflicts.
- **Vergence accommodation conflict (VAC)**  
Binocular parallax produces 3D effects but fails to adjust the binocular focus to the visual depth of field. Users cannot see objects become clearer or more blur as they move to or away from them in VR head-mounted display as in reality.

There are solutions to the first and second sources of dizziness, but no readily available products adopt any technical solutions to dizziness caused by VACs. Developing multifocal display, varifocal display, and light field display with an adjustable depth of focus are top priorities in the industry to control dizziness in near-eye display.

### 2.2.3 AMOLED, LCOS, and OLEDoS: Leading Near-Eye Screen Technologies

In VR display, active-matrix organic light emitting diode (AMOLED) is replacing liquid crystal. This trend matches the increasing AMOLED penetration rate in innovative small and mid-sized displays around the globe. Nearly 50% of all smartphones in the world are expected to use AMOLED screens by 2020. Mainstream vendors worldwide, including Oculus Rift, HTC Vive, Sony PSVR, and DP VR, choose AMOLED. AMOLED has many advantages in VR display over liquid crystal.

- The response time of AMOLED is lower than that of liquid crystal by an order of magnitude, avoiding streaking and blurring due to VR interaction.
- AMOLED VR head-mounted displays, without any backlight module, are lighter.
- Blue ray radiation, which causes retinopathy, is reduced with AMOLED.
- AMOLED consumes less power than liquid crystal in high-resolution or black background display.

Nonetheless, liquid crystal is still a vital technology in VR display. Display enterprises, such as JDI and BOE, have developed VR liquid crystal display panels featuring a response time within 6 ms, high resolution, and high refresh rate.

When it comes to AR display, LCOS- and OLEDoS-based optical see-through is the focus. LCOS is compact, light efficient, and cost-effective, and features a high resolution and high refresh rate. For these

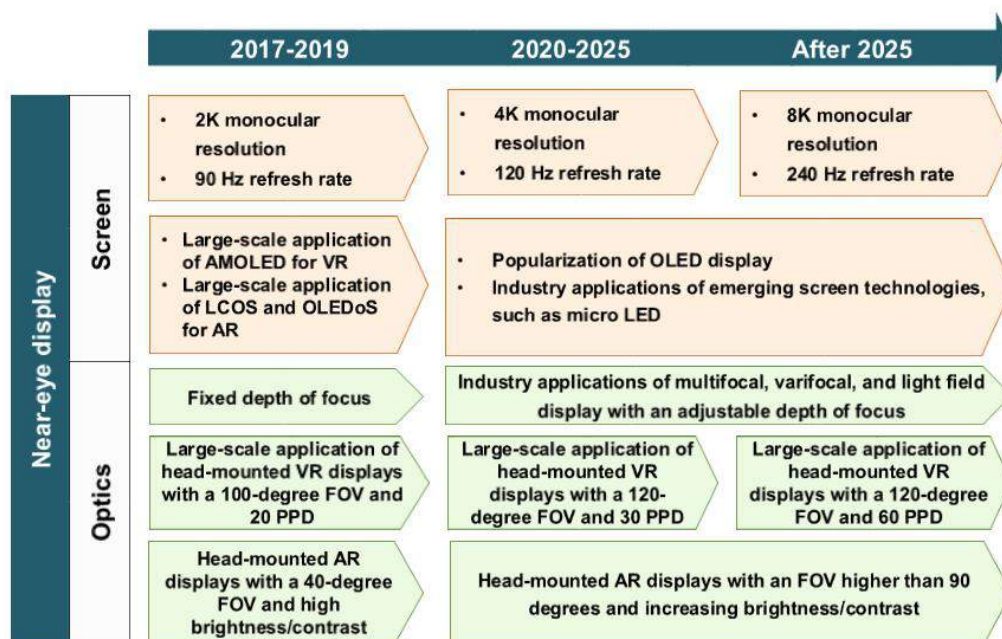
reasons, LCOS is used to make model AR terminal displays, such as Google Glass and Microsoft HoloLens. However, considering the effects of ambient light, AR terminals require higher brightness and contrast. OLEDDoS is an important alternative to LCOS to avoid the ghost effect. Model AR head-mounted displays using OLEDDoS include the R series of ODG and Moverio series of Epson.

A timetable of smooth VR technology evolution is formulated based on the relationships between goals at every stage and key technologies, including near-eye display, network transmission, rendering, perception and interaction, and content creation. The timetable is shown in the following figure.

During 2017 to 2019, it is estimated that large-scale application of screens with a 2K monocular resolution and 90 Hz refresh rate will be achieved, and the global penetration rate of AMOLED in PC-based and all-in-one VR headsets and that of LCOS and OLEDDoS in AR headsets will reach 95% (that in phone-based VR will reach 50%). As for optical system, large-scale application of head-mounted VR displays with a 90-degree FOV and 20 PPD and head-mounted AR displays with a 40-degree FOV will be achieved during 2017 to 2019. These displays will fulfill basic experience requirements in terms of brightness and contrast.

Screens with a 4K+ monocular resolution and 120 Hz+ refresh rate are expected to be widely used, and new screen technologies such as micro LED will be used in industry applications after 2020. In terms of optical system, multifocal, varifocal, and light field display with an adjustable depth of focus will be used in industry applications, and VR head-mounted displays with a 120-degree FOV and 30+ PPD will be widely used after 2020. With the limitations of power consumption control, black frame insertion frequency, and light efficiency, these displays need to increase brightness and contrast in ambient light.

**Figure 2-4** Roadmap of key VR near-eye display technologies





## 2.3 Tracking and Positioning, Environment Understanding, and Multi-Channel Interaction: Focuses of Perception and Interaction Technologies

### 2.3.1 Tracking and Positioning: a Core Technology in the Field of VR and AR Perception and Interaction

Tracking and positioning technologies are developing from outside-in to inside-out position tracking. Perception and interaction are core VR characteristics, without which VR will be reduced to head-mounted TVs or mobile phones in which communication is a fundamental characteristic. Tracking and positioning is a prerequisite of all forms of perception and interaction. Interactions are possible only when actual and virtual positions are mapped. Inside-out position tracking will be a popular tracking and positioning technology in the future. The following figure summarizes both outside-in and inside-out position tracking.

**Figure 2-5** Comparison and summary of two types of tracking and positioning technologies

Technology Type	Outside-In →	Inside-Out
Basic Principle	Tracking devices are deployed external to users to form a table- or room-scale interaction space. Installed at fixed positions, such devices can generate independent space and position information to determine the absolute position of users.	Tracking devices and users have the same origin of coordinates. These devices identify position changes based on principles such as vision and inertia, allowing users' virtual positions to change with real-world ones. Positions can be adjusted relatively (according to the origin of coordinates) or absolutely.
Model Product	<ul style="list-style-type: none"> <li>• PSVR based on visible light</li> <li>• "Constellation" tracking system of Oculus Rift</li> <li>• "Lighthouse" laser tracking system of HTC Vive</li> </ul>	<ul style="list-style-type: none"> <li>• Monocular fish-eye Qualcomm VR820/835</li> <li>• Microsoft HoloLens based on multiple cameras/ToF</li> <li>• Intel RealSense based on structured light</li> </ul>
Advantage	<ul style="list-style-type: none"> <li>• Simple algorithm</li> <li>• High robustness</li> <li>• Centimeter-level accuracy</li> <li>• Millimeter-level precision</li> <li>• Fulfilling VR/AR experience requirements</li> </ul>	Can be used in all conditions, including in mobility.
Disadvantage	<ul style="list-style-type: none"> <li>• Applicable only in stationary conditions, such as households or experience shops</li> <li>• Complex installation and commissioning of external devices</li> <li>• Independent power supply</li> <li>• Used with hosted VR terminals</li> </ul>	<ul style="list-style-type: none"> <li>• Complex algorithm</li> <li>• Consumption of a large number of system resources</li> <li>• Accuracy and precision poorer than those of outside-in tracking</li> <li>• Not robustness</li> </ul>
Example		

Somatosensory interaction using hands is evolving from gesture identification towards hand posture prediction/tracking. Gesture identification maps static hand shapes or dynamic gestures to control commands and triggers corresponding commands. These interactions require users to learn and adapt to certain gestures, impeding improvements of interaction experience. Gesture identification is widely used in AR head-mounted displays represented by HoloLens of Microsoft. This is because gesture identification is mature and has low hardware performance requirements. Gesture identification can be implemented easily using a monocular or RGB camera, but has a steep learning curve and accommodates only specific control commands, which makes it inefficient and unnatural.

Instead of identifying the meaning of a hand shape, hand posture prediction/tracking measures 26 degrees of freedoms of hand joints to rebuild the entire hand skeleton and outline, and makes virtual hands move with real hands, allowing users to move information like objects using their hands. Hand posture prediction/tracking is easier to learn and enriches more complex and natural interactions. The biggest challenge of hand posture prediction/tracking lies in its unsatisfactory performance when there are obstacles and a lack of necessary feedback. Overall user experience of such somatosensory interaction



technologies needs to be improved. Many newly established enterprises provide customized, modular hand posture prediction/tracking solutions based on dual cameras, structured light, and time of flight (ToF). Model products include LeapMotion, Intel RealSense, uSens, and Vidoo.

### **2.3.2 Environment Understanding Based on Machine Vision: Trend of AR Perception and Interaction**

Environment understanding is developing from marker-based to non-marker-based scenario segmentation and reconstruction. Unlike VR, most of what AR presents is real-world scenarios. The top priority of AR perception and interaction is to identify and understand real-world scenarios and objects, and add virtual objects to these scenarios in a more authentic and reliable way. Environment understanding based on machine vision is a focus of this field.

In early AR applications, most AR engines obtained feature information of image markers and matched this information to preset templates to identify the type and position of these markers. Markers evolved from regular geometrical shapes with clear edges, such as those in ARToolkit, to any images (in commercial engines such as Metaio and Vuforia). Marker-based identification technologies have many restrictions and cannot be widely used. With the development of deep learning and simultaneous localization and mapping (SLAM), VR/AR in the future will not be confined to marker-based identification and will evolve to semantic and geometrical understanding of real-world scenarios.

In terms of semantic understanding, the top priority is using a convolutional neural network (CNN) to identify and segment objects and scenarios in a single-frame image or multi-frame video. This process includes classification, detection, and semantic and object segmentation. The process identifies the type, position, and boundaries of an object, and further segments underlying components of the same type of objects. Deep learning-based semantic understanding usually requires a large training dataset for model training. However, its performance is much higher than that of marker-based identification technologies. Deep learning-based semantic understanding is now widely used to identify, segment, and track small objects for AR.

In terms of geometrical understanding, SLAM was used in early robot applications to repeatedly observe map features during robot movement with the starting point as the initial point, identify the robot position and posture, and map incrementally based on the robot position. In doing so, both simultaneous localization and mapping are achieved. In the VR/AR field, SLAM is widely used in inside-out tracking and positioning. The mapping process in SLAM can be used for 3D reconstruction, providing an ideal interface of presenting virtual information. ICT giants, including Google, Microsoft, and Apple, and all types of newly established science companies are working on this field.

### **2.3.3 Multi-Channel Interaction: Trend of VR Perception and Interaction**

Improving the consistency among all sensory channels and immersive experience of VR users is a key trend in the field of perception and interaction. The consistency of multi-channel interaction refers to that among sensory channels, including vision, auditory sense, and tactile sense, and that between proactive motions and motion feedback. Interaction technologies, including immersive sound field, eye tracking, tactile feedback, and voice interaction, are becoming rigid demands in the VR field to meet requirements for dizziness control and immersive experience.

As for the immersive sound field, great importance is attached to sound in the field of perception and interaction. Head reference transmission functions (HRTFs) are designed to enhance the consistency between vision and auditory sense, and produce authentic sound positions and near-far effects. These functions are also used to simulate sound transmission paths in conditions such as reflection, obstacles, isolation and enclosure, and reverberation and echoes. For example, NVIDIA used ray tracing and rendering technology to map VR audio interactions to objects in 3D scenarios, creating an immersive VR sound field that meets auditory and acoustic characteristics. Traditional acoustic vendors, such as Dolby and DTS, and Microsoft, Google, Qualcomm, Unity, Oculus, and OSSIC have launched hardware and software products for immersive sound fields.

Eye tracking is an industry focus because of several reasons. First, eye tracking can be used with fixation point rendering to adjust the fixation point of users and use different rendering resolutions in different areas. This reduces consumption of computing resources, power, and GPU costs. Eye tracking is poised to become a building block for delivering highly immersive VR experience on mobile terminals.

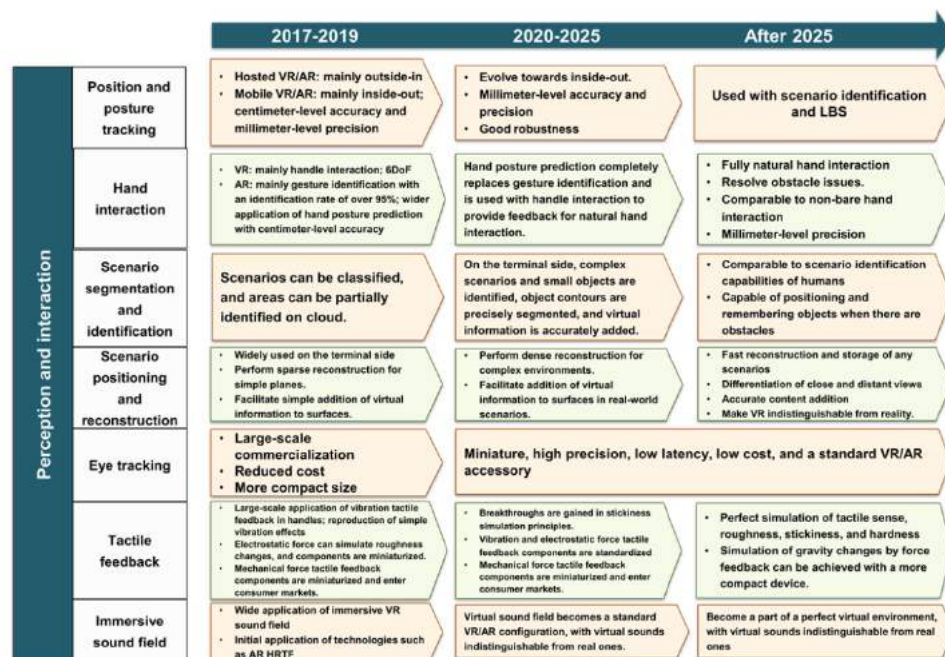
Second, eye tracking is expected to resolve VACs. Multifocal display achieves partial blurring through optical designs, while varifocal achieves partial blurring through GPU fixation point rendering. Varifocal display is dependent on eye tracking. For this reason, the combination of multifocal display, fixation point rendering, and eye tracking is expected to become a crucial, emerging technology combination in the VR field.

Third, eye tracking can be used to design innovative VR content, for example, VR games such as Interrogate of Fove and SOMA of Tobii. Representatives that use eye tracking for innovative VR content creation include SMI, which has been recently acquired by Apple, Tobii in Sweden, and 7Invensun in China.

In terms of tactile feedback, bare hand interaction is not widely accepted because the lack of feedback causes inaccurate operations. From low-cost vibration tactile feedback and high-cost mechanical force feedback to immature electrostatic resistance feedback, tactile feedback technologies are attracting more industry attention. Oculus Touch and Vive Controller are focusing on interactive handles, while newly established enterprises, such as Go Touch, are launching trials on finger cot feedback.

In conclusion, VR and AR have the same basic requirement, which is mapping real coordinates to virtual coordinates. This requirement is a prerequisite of smooth perception and interaction. Therefore, tracking and positioning technologies become basic capabilities of VR/AR engines or software development kits (SDKs). The availability of these engines and SDKs is dependent on positioning accuracy and precision. VR perception and interaction focus on multi-channel interaction. All that VR presents is virtual information, and therefore virtualization of reality interactive information is important. AR perception and interaction, in which a large part is real-world scenarios, focus on machine vision-based environment understanding. In general, perception and interaction are a popular field that major vendors prioritize and the most possible field for product differentiation. There are abundant requirements for perception and interaction.

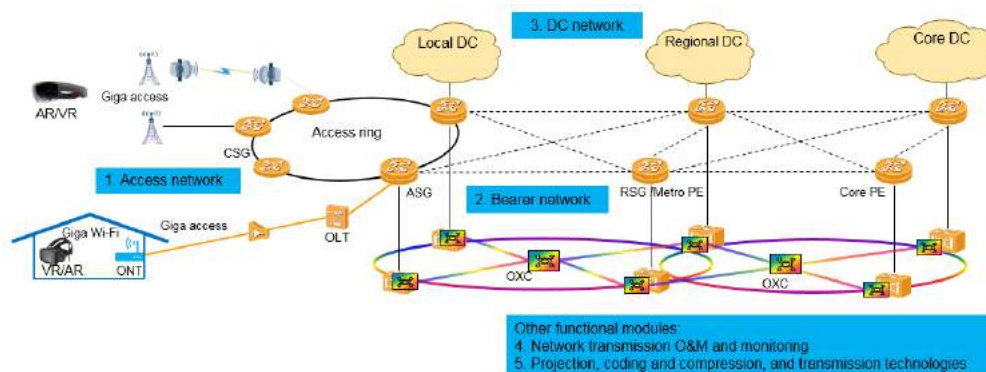
**Figure 2-6** Roadmap of key VR perception and interaction technologies



## 2.4 High Bandwidth, Low Latency, Large Capacity, and Service Isolation: Trends of Network Transmission Technologies

VR involves a wide range of network transmission technologies, including access network, bearer network, data center network, network transmission operation, maintenance, and monitoring, projection, and coding and compression. E2E network transmission for VR is experience-centric and aims to improve immersive experience. The following figure shows E2E network transmission for VR services.

**Figure 2-7** E2E network transmission for VR services



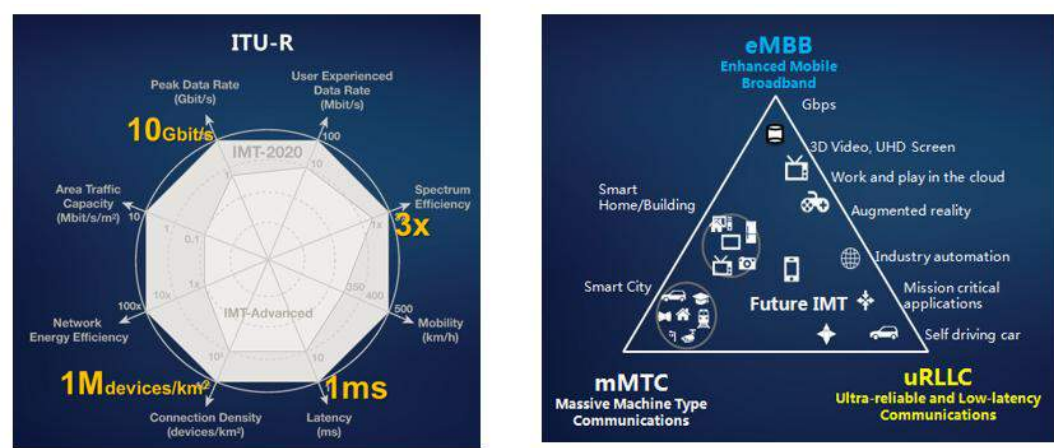
### 2.4.1 New Wi-Fi, Large-Capacity PON, and 5G: Trends of Access Network Technologies for VR

The next-generation Wi-Fi technology can provide household wireless coverage at a maximum rate of 10 Gbit/s over the air interface. Wireless VR head-mounted display is based on 60 GHz Wi-Fi technology. Household wireless coverage ensures mobility and convenience of VR services and fulfills service requirements in terms of bandwidth and latency. Wi-Fi devices based on 802.11n or 802.11ac are widely used to provide household wireless coverage. 802.11n supports both 2.4 GHz and 5 GHz bands, while 802.11ac supports 5 GHz band. 802.11ac-based Wi-Fi can implement 4x4 multiple-input multiple-output (MIMO) and beamforming on 80 MHz to reach a maximum rate of 1.7 Gbit/s over the air interface. 802.11ax is a next-generation Wi-Fi technology that introduces a wide range of new features, including 8x8 MIMO, OFDMA, and 1K QAM. 802.11ax can reach a maximum rate of 10 Gbit/s over the air interface and is highly resistant to interference, ensuring KPIs such as packet loss rate, latency, and bandwidth stability. Wireless VR head-mounted display uses wireless transmission technologies to losslessly transmit videos, improving user experience by eliminating wired connection. IEEE 802.11 is formulating a 60 GHz next-generation Wi-Fi standard 802.11ay, which uses channel bonding and MU-MIMO to provide 20 Gbit/s to 40 Gbit/s and transmit uncompressed video frame data.

With the bandwidth capability and latency performance of passive optical network (PON), light access technologies can be used to provide VR/AR services. Ethernet passive optical network (EPON) and Gigabit-capable passive optical network (GPON), which are fixed broadband access, serve as the access and aggregation network of household networks, and are closest to users among all operator networks. Fiber to the home (FTTH)-based EPON/GPON has been deployed on a large scale. GPON can provide actual bandwidths of 2.5 Gbit/s in the downlink and 1.25 Gbit/s in the uplink, while EPON can provide symmetrical actual bandwidth of 1 Gbit/s. The latency is about 1 ms to 1.5 ms, fulfilling the basic requirement of small-scale VR/AR services. Follow-up deployment of 10G PON technologies will increase the bandwidth tenfold. In addition, IEEE 802.3 has started to work on 25G and 100G PON standards (25G PON is used in FTTH scenarios), and the ITU-T has launched research into 10 Gbit/s+ next-generation PON requirements and technologies. Meanwhile, major telecom and optical module vendors are following up related technology research and standards formulation.

Ultra-high bandwidth, ultra-low latency, and ultra-high mobility of 5G ensure fully immersive VR experience. VR becomes a key field of early 5G commercial use. Besides improving traditional KPIs, including peak rate, mobility, latency, and spectral efficiency, 5G mobile communications technologies introduce four new key capability indicators: user experience rate, connection density, traffic density, and energy efficiency. 1 ms E2E latency, 10 Gbit/s throughput, and 1 million connections per square kilometer are three key 5G requirements to enable future-oriented communication. As for 5G application scenarios, with a user experience rate of 100 Mbit/s to 1 Gbit/s and millisecond-level transmission latency, 5G can provide mobile VR and other high-bandwidth, low-latency services, allowing operators to expand business boundaries. Countries that have early 5G commercialization plans, including Japan and South Korea, plan VR as an early, key 5G application field.

**Figure 2-8** 5G key capability indicators and three application scenarios



**Figure 2-9** Early 5G application fields in countries that have early 5G commercialization plans

5G Planning	Spectrum	Time	Early Application Field
South Korea	3.4 GHz to 3.7 GHz and 28 GHz	2018 (demonstration)	VR (2018 Pyeongchang Olympic Games)
Japan	4.5 GHz and 28 GHz	2018 (demonstration)	VR, cloud computing, self-driving and drone, and industrial devices
	3.6 GHz to 4.2 GHz and 4.4 GHz to 4.9 GHz	2020 (commercialization)	VR (2020 Tokyo Olympic Games)

### 2.4.2 Simple Architecture, Smart Pipe, Multicast On Demand, and Network Isolation Are Development Trends of Bearer Networks for VR Services

VR services bring new ideas for building bearer networks. Traditional bearer network construction seldom considers service experience. Network planning and capacity expansion are required only when the total bandwidth reaches the usage threshold. As VR services have high requirements for bandwidth and latency, the following ideas are put forward for building bearer networks to meet such requirements.

- Service experience must be guaranteed for each service interaction, such as a viewpoint switch using head motions, viewpoint switch in mobility, and other different interactions.
- The latency per interaction is a focus of network capabilities. For example, a viewpoint switch in immersive VR experience requires a latency within 10 ms.

- Bearer network bandwidths need to be planned to meet latency requirements of services. For example, a bearer network should be able to transmit a large volume of VR traffic within 10 ms.

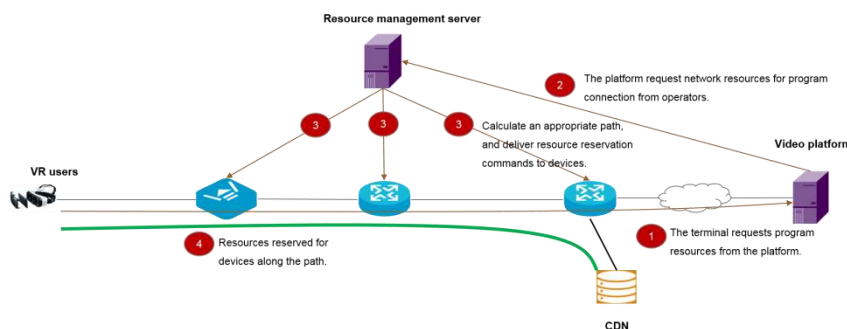
Bearer networks can support live video services in multicast mode. Traditional live video services are carried in a unicast mode and consume larger bandwidths, bringing great challenges to servers and bearer networks. Prevailing OTT solutions reduce the bit rate of the unicast streams, which negatively affects playback experience. Using multicast on demand (MOD) instead of the unicast mode to carry VR live video services can ensure user experience because network traffic and CDN server loads do not increase with the number of users.

Developing data-plane isolation technologies ensures a high priority of low-latency services. The bandwidths of IP networks are shared by different services, and latency-sensitive services may not have sufficient bandwidths. Network isolation technologies such as FlexE rise to VR, AR, and other latency-sensitive services. Using these technologies, large-granularity bandwidth services carried on the same port can be physically isolated and bound, and decouple rates at the MAC layer and the physical layer. This makes network design and capacity expansion more flexible while ensuring stable E2E low-latency networks.

Superior VR experience requires the development of customized smart pipes. If VR service flows are forwarded with other internet data in an undifferentiated manner, E2E service quality cannot be ensured. To solve this problem, operators provide differentiated customized services, adapting to on-demand, dynamic, open, and E2E pipe development.

- On demand: Network resources can be dynamically expanded or decreased based on VR service requirements.
- Dynamic: Network devices calculate the service QoS in an E2E way every second during the session instead of setting a fixed QoS. Resources are allocated and scheduled at each node, and released and reallocated to other services immediately after the service is terminated, improving the resource usage.
- Open: Operators provide user-friendly, clearly defined, and well-developed interfaces, which can be invoked and customized to ensure the network quality for VR services. These interfaces facilitate service application, adjustment, release, billing, reconciliation, and settlement.
- E2E: Central E2E management and calculation are required to ensure the service quality. The intelligent management system obtains the real-time status (including bandwidth and latency) of each device on a network providing VR services through a central management unit. The unit calculates an appropriate path based on the device status upon receiving a service request, and delivers resource reservation commands to devices along the path, preparing for service transmission.

**Figure 2-10** Customized network services based on VR services



A flattened network architecture improves the transmission efficiency of bearer networks. Big video services such as VR have high requirements on bandwidth, latency, and packet loss rate. Traditional bearer networks featuring high aggregation and convergence ratio encounter the following challenges:



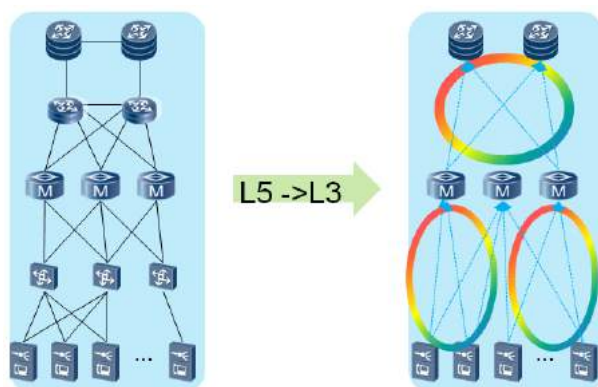
- Low network efficiency. More aggregated layers result in a lower convergence ratio and more E2E devices requiring capacity expansion. As the CDN is deployed at a higher network layer, service flows need to pass through a large number of network devices, causing a higher congestion probability and a longer E2E latency.
- Poor user experience. When multiple services are concurrent, the network utilization, packet loss rate, and latency increase. In lightly-loaded networks, 98.7% of burst packet losses occur at the aggregation nodes where high bandwidth changes to low bandwidth, increasing the packet loss ratio while deteriorating video service experience.

Therefore, the traditional network layers and network structure need to be simplified:

- Move the CDN downward to the broadband network gateway (BNG) or even to the CO.
- Eliminate the LAN switch (LSW) aggregation layer and metropolitan aggregation layer.
- Directly connect the BNG to the CR and move the BNG downward to the network edge.
- Directly connect the OLT to the BNG.
- Deploy the OTN to the CO.

This reconstruction improves the bearer network transmission efficiency by offering basic interconnected pipes with single-fiber ultra-high bandwidths, best adaptation distance, non-aggregated traffic, and fast on-demand bandwidth.

**Figure 2-11** Flattened bearer network architecture (decreasing the number of network layers from 5 to 3)



### 2.4.3 A Low-Latency Data Center Is Key to VR Computing Capability Cloudification

A low-latency data center is dependent on congestion control technologies, which are evolving from passive to proactive control. Currently, superior visual experience of VR services depends on costly graphics processing units (GPUs). In the future, VR computing capability will be cloudified, and complex calculation functions will be deployed in a data center. VR function cloudification imposes new requirements on data center networks, including larger data flow and lower service latency.

Congestion control technologies can be classified into passive and proactive types based on control points of congestion control technologies for the data center networks and detection mode of the network congestion status. Passive congestion control gives feedback slowly in high-speed networks. Proactive control points deployed at the network side or receiving end can accurately detect the congestion status and concurrency of multiple service flows at the receiving end without measurement. Using this detected information, proactive congestion control dynamically allocates the rate, speeds up convergence, requests the source end to control the data transmission accurately based on a specific rate, and schedules different

flows based on traffic requirements. Typical active congestion control technologies include ExpressPass and pHost. In the future, congestion control technologies will be more application-centric and meet application requirements regarding priority, bandwidth, latency, and flow completion time. The primary goal of congestion control technologies is to meet service requirements, controlling and avoiding congestion.

#### **2.4.4 Projection, Coding, and Transmission Technologies Are Key to Optimizing the VR Network Performance**

Projection technologies develop from equiangular projection (ERP) to polyhedral projection. In VR, the spherical information that users see needs to be transferred into the planar media format, requiring projection technologies. Traditional videos are not involved in the projection technology.

- An ERP is a mainstream format of VR 360-degree videos. However, image distortion may occur, and the compression efficiency is limited. Content providers YouTube, Oculus, Samsung Gear, Youku, and iQIYI adopt this projection technology to produce VR 360-degree media files.
- Polyhedral projection has been a buzzword of the telecom industry. It features few distortions and a high compression efficiency, and includes hexahedral, octahedral, icosahedral, or pyramid projection. In the Moving Picture Experts Group (MPEG) conference in 2016, Samsung submitted proposals related to polyhedron projection formats.

As the next-generation coding technologies develop steadily, the compression efficiency is improved significantly. Currently, the high efficiency video coding (HEVC) is used for VR coding. Latest research of MPEG and other standards organizations shows that the next-generation coding technology H.266 improves the compression efficiency by up to 30%. Prevailing coding tools are specific to 2D plane videos instead of spherical data. Therefore, a coding tool for VR 360 videos becomes a research focus.

VR transmission technologies change from full-view equi-quality mode to full-view non-equi-quality mode or field of view (FOV) mode. In full-view equi-quality transmission, terminals receive a data frame containing information about all spherical visions that users view. The interaction signals produced when users change viewpoints are processed by local terminals. Based on the viewpoint information, terminals detect the corresponding FOV information which has been cached locally, and correct the information in the player. In this way, users can view visual information from a normal visual angle. Therefore, terminals must ensure that the interaction latency is less than 20 ms, which does not include network and cloud latency. This technology requires larger bandwidths and accepts longer latency, which trades latency for bandwidth.

For content preparation, full-view VR contents must be coded and VR bit streams with different quality levels must be provided. The clients can select different streams to play based on bandwidth, but some data transmitted to clients is wasted due to the impact of FOV.

In FOV transmission, terminals receive frame data produced based on users' viewpoints instead of all information about spherical visions that users view. Frame data contains some visual information obtained by the equal or greater angle of FOV. If users turn their heads to change viewpoints, terminals need to determine the motions, send interaction signals to the cloud, and apply for frame data corresponding to the new positions. Therefore, good interaction experience requires a latency of less than 20 ms, which includes terminal processing latency as well as network and cloud latency. This technology requires lower bandwidths and shorter latency, which trades bandwidth for latency.

#### **2.4.5 VR-oriented Network O&M and Evaluation Are Important for Improving User Experience**

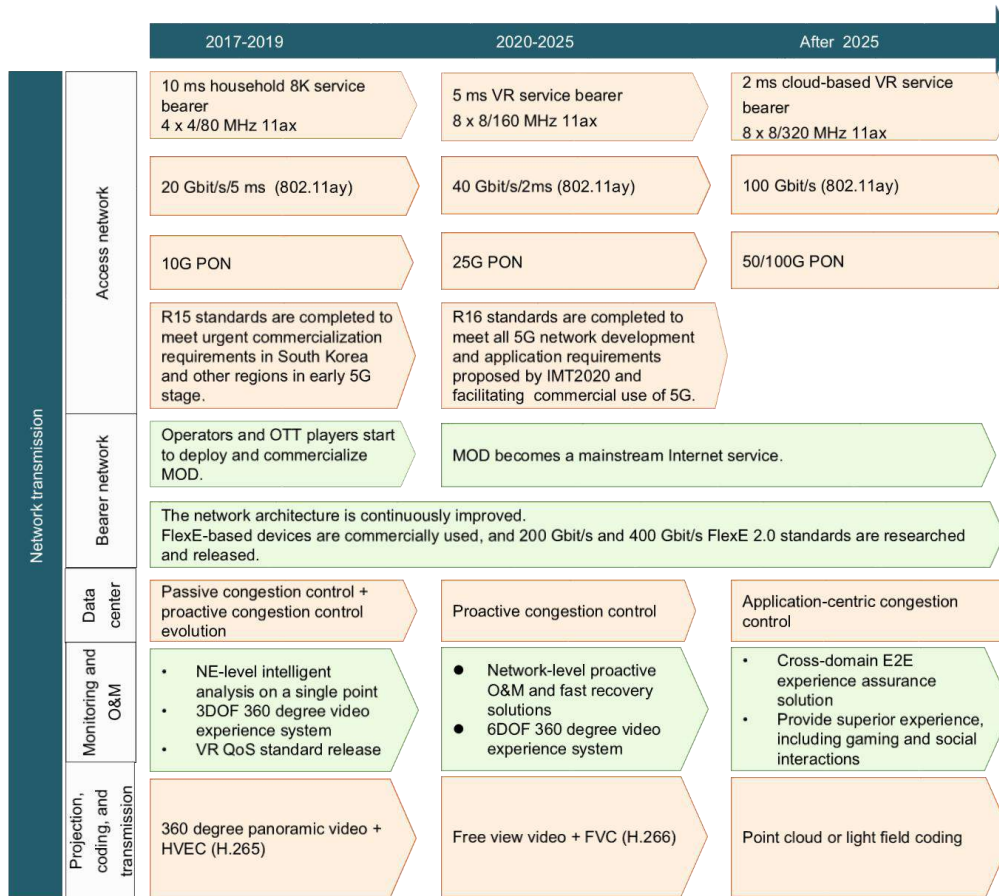
VR-oriented E2E network O&M and Quality of Experience (QoE) evaluation system become focuses. Processes from VR video content creation to playback on clients are more complex than those of common 4K videos. In this case, there are various types of faults, and video experience deterioration is hard to be demarcated. Effective methods are required for fast fault demarcation and diagnosis. Therefore, detectors

must be deployed at different points of the cloud-pipe-device architecture to implement real-time monitoring. This helps quickly detect and demarcate faults, facilitating E2E network O&M. Additionally, a well-developed user experience evaluation system helps manufacturers improve their products and service providers enhance service experience, promoting industry chain development. A hierarchical multi-sensing VR QoE solution is proposed to model user experience and measure the perceivable media quality during network transmission.

Vertical industrial services impose higher requirements on latency, bandwidth, reliability, and mobility, promoting VR network transmission development. Additionally, the evolution of bearer network technologies facilitates an agile, open, and increasingly flexible operation mode that implements isolation of different services. Technological trends are as follows:

- High bandwidth access.
  - For fixed access, indoor Wi-Fi provides 100 Mbit/s or 1000 Mbit/s coverage and the next-generation Wi-Fi offers up to 10,000 Mbit/s coverage.
  - For wired access, 10G PON is mature and can be deployed in appropriate scenarios. To meet higher bandwidth requirements, 25G PON and 100G PON standards are being defined.
  - For wireless access, realizing 10GE throughput is one of the three key requirements for 5G communication.
- High-capacity bearer network. Bearer networks mainly use fiber-optic rings as access rings, which require 50-100 GHz bandwidth. Tail microwave requires 10-40 GHz bandwidth, and the aggregation core requires 200-400 GHz bandwidth.
- Low-latency networks. Each network element (NE) in this architecture must maintain an ultra-low-latency forwarding capability, reaching an approximately 10  $\mu$ s-level latency of each hop. A manageable architecture featuring measurable latency is desirable for building low-latency bearer networks.
- Agile and open networks. SDN enables IT-based networks and automation transformation to realize agile networks and network capability exposure. This effectively enhances deployment efficiency for cloud-based network and complex traffic models, and implements E2E resource management and service provisioning. Additionally, it shields complicated lower-layer network interactions, and facilitates real-time network resource scheduling to enhance resource utilization.
- Service isolation. As various services have distinctly different requirements, Different services must be isolation on the data plane, and the management and control planes must be separately maintained. Bearer networks use network slicing to isolate services and allow multiple tenants to control network slices.



**Figure 2-12** Roadmap of key VR network transmission technologies

## 2.5 Combination of Enhanced Rendering Algorithms and Capabilities Are Trend of Rendering Processing Technologies

Rendering algorithms are optimized to reduce invalid computing and rendering loads based on visual characteristics, head-motion-based interaction, and deep learning. VR rendering processing aims to achieve high quality images, low overhead, and low latency. This enables users to enjoy smooth, clear, and real-time VR visual experience. Insufficient rendering capabilities hinder the improvement of user experience. Therefore, reducing computational overheads and rendering latency has become a development trend. The following measures can be taken to solve such problems:

- Reduce GPU workload, for example, using foveated rendering. Early VR rendering technologies implement high-resolution rendering for all details in an FOV. As cone cells for handling visual clarity are centrally distributed in the fovea instead of in an FOV, areas away from the fovea have blurry visual perception. High-resolution rendering in most areas wastes larger number of resources. Therefore, foveated rendering is proposed, using high-resolution rendering for the fovea area while gradually reducing the resolution for surrounding areas. This technology reduces the rendering overhead of GPUs by at least 30%.
- Reduce the CPU overhead, for example, using multi-view rendering. Early VR rendering must simultaneously render binocular image for both eyes. A rendering request must be submitted for each image of each eye in each frame, doubling the CPU/GPU resource usage over common rendering. Therefore, multi-view rendering technology is proposed, reusing similar information of images for left and right eyes. After a rendering request and binocular disparity processing request are submitted to

the CPU, the GPU can complete binocular rendering. This saves a large number of CPU resources, and improves the GPU frame rate.

- Reduce the rendering latency, for example, using asynchronous time warping (ATW), asynchronous space twist (ASW), and buffer before rendering (FBR). Complex contents cannot be rendered in a frame refresh cycle. As a result, no new contents are generated, causing frame freezing. Therefore, ATW and ASW rendering are proposed. These technologies predict the next head motion and generate an intermediate image based on the motion differences. Frame freezing occurs due to lack of current frames, and cannot be tackled currently. These key VR rendering technologies can ensure smooth visual experience in most cases.

Light effects require large computing resources. The deep-learning-based global light rendering technology can generate images by simulating physical effects of interactions between lights and objects. The contents that the GPU does not render are displayed using deep learning. This improves the image generation speed by several times and smoothness of the interactive rendering process while reducing image noises.

VR rendering capabilities are improved by using multiple new technologies, such as cloud-base rendering, new-generation graphic interfaces, heterogeneous computing, and light field rendering. Cloud-based rendering expects to bring more vivid immersive experience, requiring 16K or 32K resolution, 120 or higher frame rate, and real-time lighting. The same standard is hard to achieve with existing GPUs. This technology performs content rendering on the cloud. Terminals transmit various types of sensor data and interactive control information such as 6DoF motions and positions in real time. Cloud computing clusters render and send images to terminals.

Cloud-based rendering implements a large amount of computing on the cloud to ensure that users can enjoy high-quality 3D rendering effects using lightweight VR devices. This lowers hardware requirements for VR devices and consumers can choose different levels of immersive experience based on financial costs. Currently, cloud-based rendering is mainly used in film post-production and industrial design. Although cloud-based rendering is not mature and commercialized in VR services or games, they are mainstream future-oriented technologies.

The next-generation graphic interfaces represented by Vulkan achieve stable, low-latency, and low-overhead rendering in VR applications. These graphic interfaces have the following characteristics:

- Adaptability to multiple platforms. Unlike Microsoft's DirectX and Apple's Metal, Vulkan supports operating systems of Linux, Windows, and Android, desktops of AMD, NVIDIA, Adreno, PowerVR, and Mali, and mobile GPUs.
- More flexible and accurate GPU control. Complex drive interfaces of OpenGL and OpenGL ES consume excessive CPU resources and cause unpredictable equipment operations. VR applications cannot predict the overhead of invoking an interface so that rendering cannot be optimized. Vulkan provides a simpler drive interface and transfers memory and multi-thread management to applications, featuring smaller computation overhead and better device consistency.
- More efficient multi-core CPU usage. Each independent thread can deliver commands to the command buffer through Vulkan. This prevents insufficient CPU resources from hindering the graphics rendering process.

Heterogeneous computing handles distortion correction, dispersion correction, and ATW or ASW rendering, and content rendering shared by the same CPU or GPU resources may cause resource competition. If application contents are complex, post-processing procedures may not be completed within the specified time, causing an unstable output frame rate. Therefore, ASIC, FPGA, and other heterogeneous modes are proposed, stabilizing the output frame rate and cutting MTP latency.

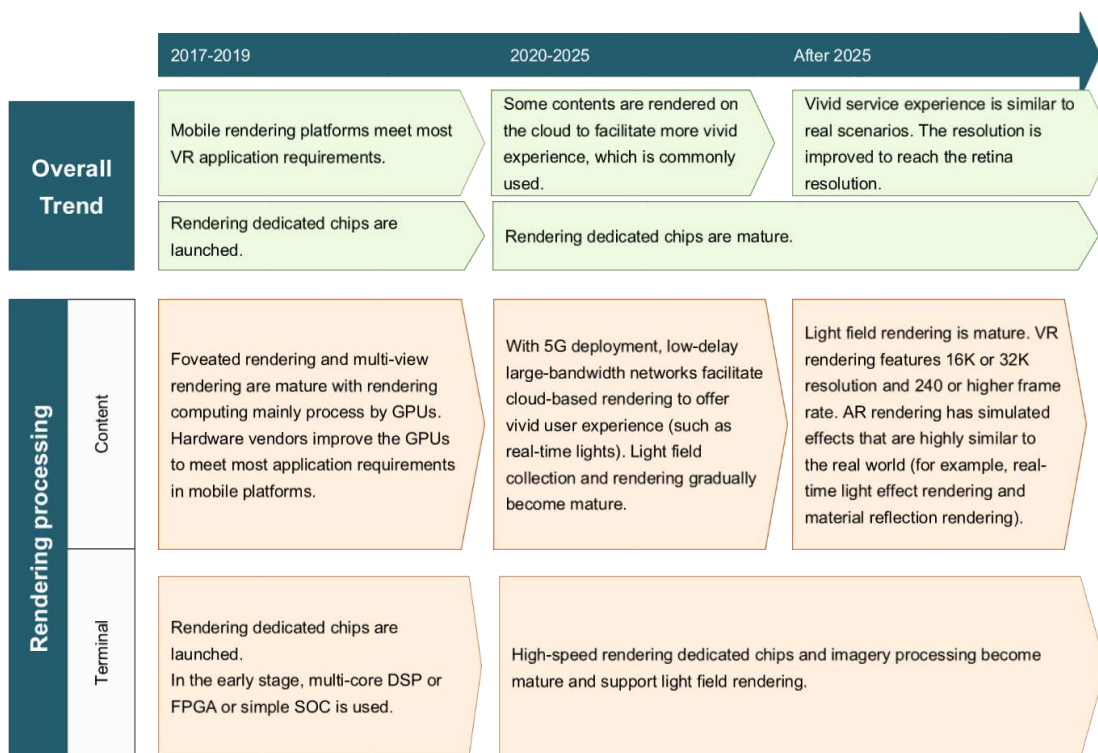
For light field technologies, current VR services use two-dimensional imaging which generates binocular parallax, causing VACs and dizziness. Eyes collect lights reflected by object surfaces at different distance, positions, and directions by changing the focal length. These lights form a light field. Vendors have produced light field cameras. Light field rendering can restore the collected light field information to offer better immersive experience. Currently, light field information collection, storage, and transmission face

multiple basic problems such as huge volumes of data. Light field rendering is still in its early stage of development. Light field rendering may become a key rendering technology in the future to meet higher VR experience requirements.

In conclusion, rendering processing involves two parts: content rendering and terminal rendering. Content rendering projects the 3D virtual space on a plane and forms planar images. Terminal rendering corrects optical distortion and dispersion of planar images generated by content rendering, and inserts frames based on user postures. All rendering technologies aim to improve rendering performance, maximize the resolution with minimum costs, and generate more perceivable details.

The biggest challenge for VR rendering is complicated content computing, for example, two-fold GPU calculation volume of common 3D applications and real-time light effects. Rendering technologies of AR services are similar to those of VR services, but AR application scenarios focus on integration with the real world, for example, virtual and physical blocking, light effect rendering, and material reflection rendering. In the future, VR rendering technologies will offer more diversified and vivid immersive experience. Foveated rendering, cloud-based rendering, rendering-dedicated chips, and light field rendering will become mainstream technologies as hardware capability, costs, and power consumption are limited and 5G will be commercialized around 2020.

**Figure 2-13** Roadmap of key VR rendering technologies



### 3 Industry Ecosystem Trend

#### 3.1 Components/Devices, Tools/Platforms, and Content Applications Dominate in the VR Industry Ecosystem

The VR industry greatly differs from the mature electronic information industry (including mobile phones, TVs, and PCs). Although these two industries have similar industry chain participants, "excessive performance" of devices such as mobile phones becomes a "performance threshold" of VR.

Figure 3-1 VR industry in China



Source: CAICT

In addition, to ensure user experience such as immersion, the tools or platforms for preparing VR contents change significantly. Focusing on the core feature of human-machine interaction, the industry deeply optimizes fields such as image and audio capture, development engine, network transmission, and SDK/API and even redevelops or redesigns them. In general, VR and AR industry systems are similar, but differ mainly in the following aspects:

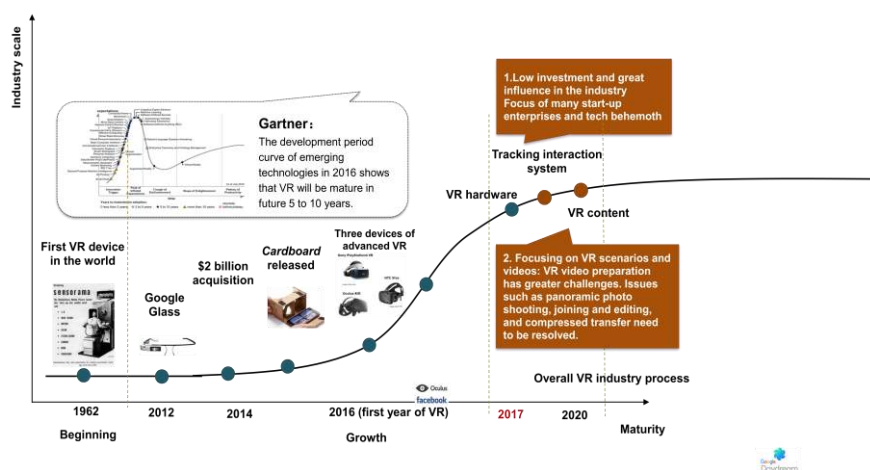
- Number of enterprises: The number of enterprises in the AR field is significantly less than that in the VR field.
- Focuses of segmented product markets in some fields: For example, in respect of content creation and video capture, VR focuses on 360-degree panoramic photo shooting, but AR focuses on three-dimensional scenario measurement.



### 3.2 Perception and Interaction and Content Creation Become Development Focuses of VR in the Next Phase

VR industry development complies with the "hardware-content" development pace. Hardware is the foundation and contents/algorithms are used to improve experience. HTC Vive, Oculus Rift, and Sony PSVR products were launched in sequence before and after 2016. The launch of these products indicates that the hardware threshold of VR is significantly reduced and that the industry has completed the first hardware development phase and started to develop towards the second development phase dominated by perception, interaction, and content creation.

**Figure 3-2** Perception, interaction, and content creation becoming the development direction in the next phase



**Perception and interaction:** The features of low investments and large output bring active investment, cooperation, and merging among VR enterprises. Various start-up companies are emerging, and tech behemoths such as Apple, Microsoft, Intel, and Google actively reserve related technologies by means of merging or self-development. IPR analysis shows that 3D modeling becomes the focus in the industry. In addition, multi-channel interaction, including eye tracking and force feedback, continuously and quickly develops.

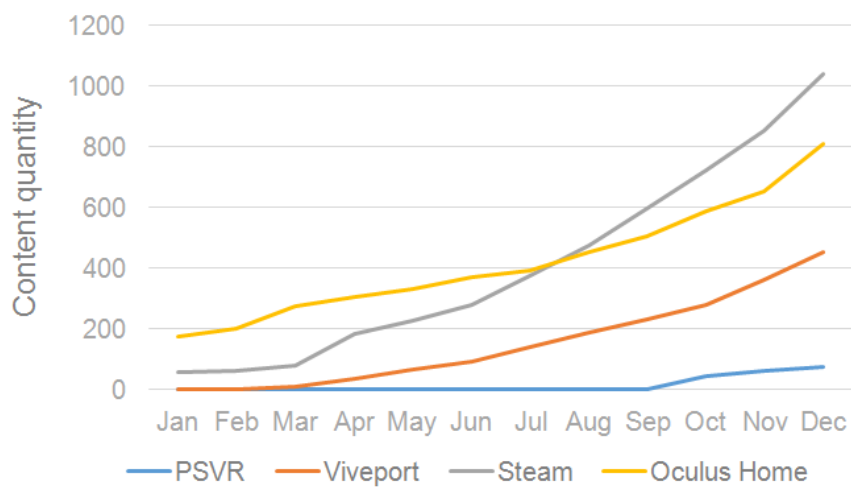
**Content creation:** focuses on two fields in the consumer market, namely, gaming/social interaction and video/live broadcast. The production mode of video/live broadcast greatly differs from the traditional mode. Specifically, photography and narration modes change, production procedures including splicing and seaming, engine development, and distortion processing are added, and compression and transmission technologies become technical bottlenecks of VR video live broadcast. YiVian statistics show that consumer contents on the four content delivery platforms Oculus Home, Steam, Viveport, and PlayStation Store were only about 200 in 2015 but rapid increased 12-fold to about 2400 in 2016. On the Steam platform, the average number of users of each application is 75,000, and the average number of users of the contents developed by Chinese teams is 4000.

**Figure 3-3** Numbers (hot degrees) of patents in the perception and interaction field

Year	Eye Tracking	Force Feedback and Tactile Feedback	3D Modeling	SLAM
2016	178	215	1091	48
2015	115	134	888	54
2014	109	99	602	32
2013	76	74	538	21
2012	74	75	418	31
2011	22	66	348	10
2010	16	56	261	11
2009	8	34	195	7
2008	8	39	186	5
2007	7	24	119	6
2006	11	24	114	4
2005	6	12	96	5
2004	4	16	91	2
2003	2	23	79	2
2002	4	19	86	0
2001	9	17	75	0
2000	6	21	86	0
1999	5	13	58	0
1998	3	10	57	0

Source: CAICT

**Figure 3-4** VR content growth rates on different platforms in 2016



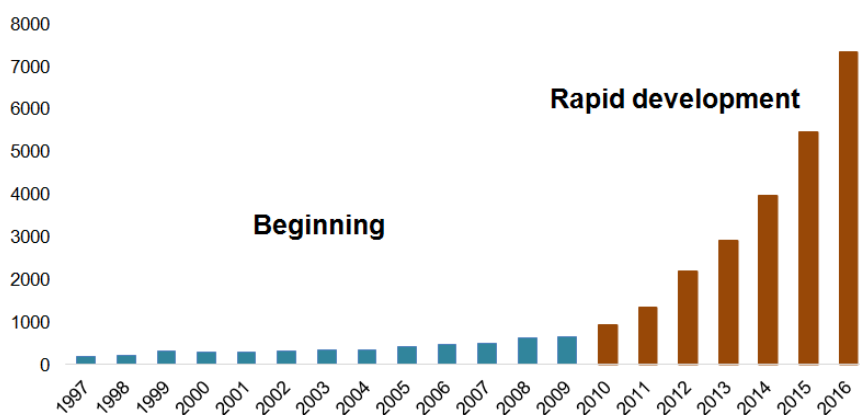
Source: YiVian

### 3.3 IPR Competition Represents the Industry Development Trend

#### 3.3.1 Global IPR Development Entered a Rapid Growth Stage, and the Growth Rate in China Is Notable

With the popularization and industrialization of VR, global ICT giants actively participate in VR development. The number of patent applications quickly increased after 2010 and will exceed 10,000 before 2020. By the end of May 2017, the number of global valid VR patents had reached 59,000, and there are about 29,000 patent families. Because there is a time period from patent application to patent release, the number of patents in 2016 shown in the following figure is less than that of patents actually applied for in 2016. The numbers in the following figures are those of patent families.

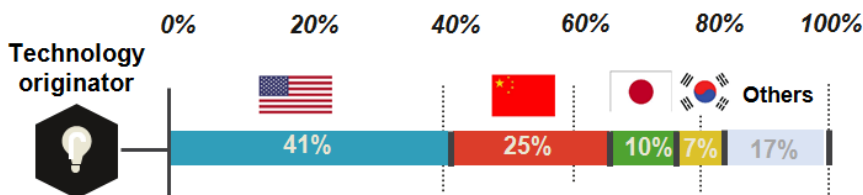
**Figure 3-5** Number of global VR patent applications in each year



Source: CAICT

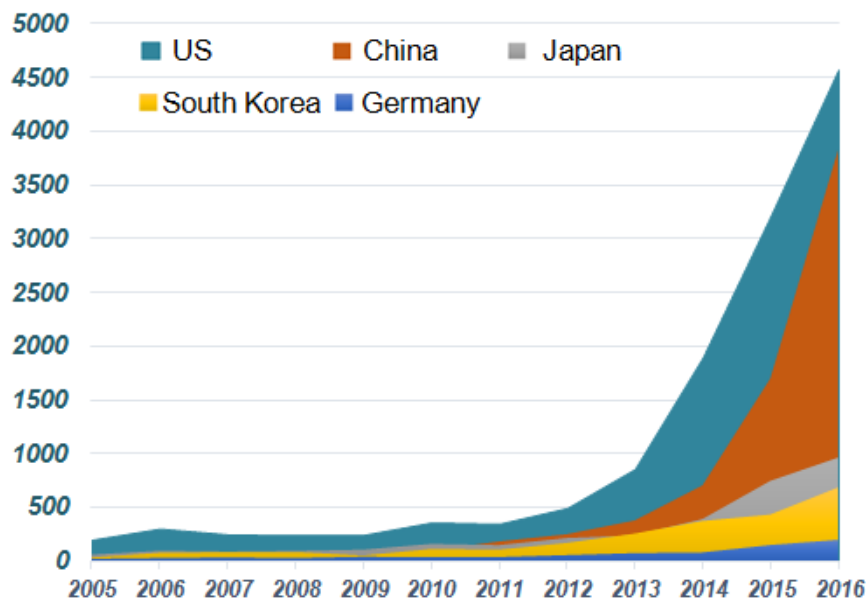
Among countries originating VR patent technologies, the US, China, Japan, and South Korea apply for a larger number of patents and they are prospective. The US started patent application earlier than other countries, and patent application in China developed quickly after 2013 with a notable annual increase.

**Figure 3-6** Percentage of patent applications of VR technologies originated by different countries



Source: CAICT

**Figure 3-7** VR patent application numbers in different countries

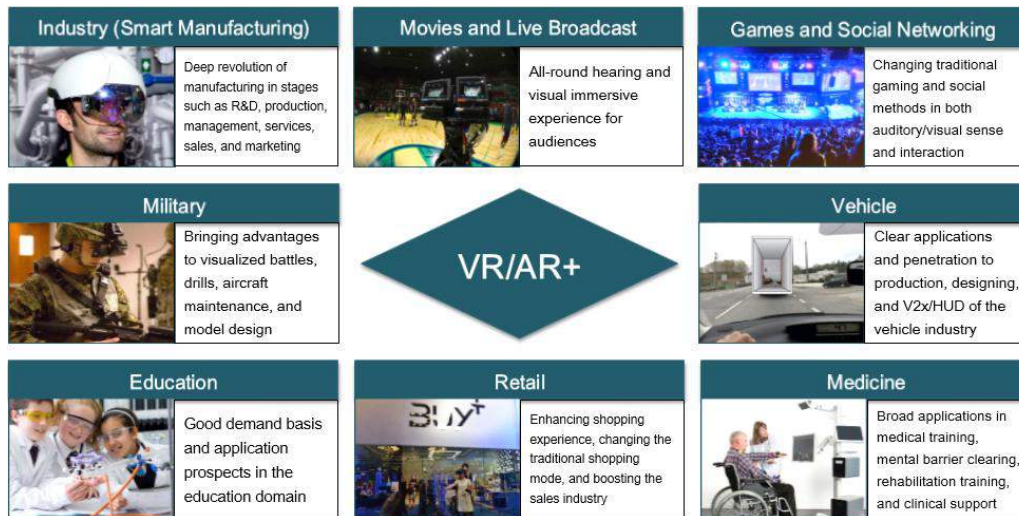


Source: CAICT

### 3.4 Emergence of the VR/AR+ Era

VR services come in various forms and have great industrial potential. These services will bring dramatic social benefits. The new technological and industrial revolution, represented by VR, is ready to take place. The combination of virtual economy and real economy will bring revolutionary changes to people's production modes and lifestyle. Currently, VR applications can be classified into industry applications and public applications. The former includes areas such as industry, medicine, education, military, and e-commerce; the latter includes games, social networking, movies, and live broadcast. The penetration of VR applications to manufacturing and life is accelerating. The VR/AR+ era has emerged. According to Goldman Sachs, the market share of global VR software applications will reach USD 45 billion in 2025, with areas such as games, social networking, videos, and live broadcast promoted by the mass and other areas promoted by enterprises and public departments.



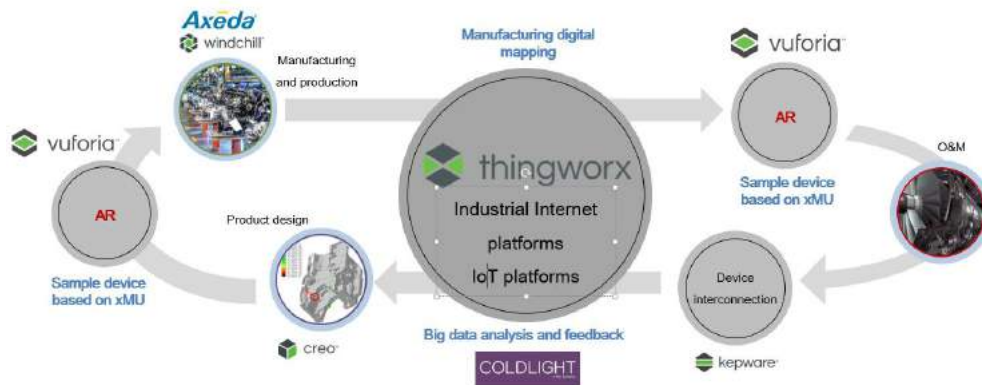
**Figure 3-8** Typical VR application domains

### 3.4.1 In VR+ Industries, VR is the key in smart manufacturing

In the technology roadmap for major areas, the policy document "Made in China 2025" lists VR as one of the key technologies of core information equipment in smart manufacturing. Fundamentally, VR is used for information collection and real-time communication in all stages during smart manufacturing, implementing dynamic interaction, and decision-making analysis and control. In the auto industry, VR implements the integration of design, manufacturing, and testing in stages such as requirement analysis, overall design, process design, manufacturing, testing, and maintenance. Vehicle vendors can achieve visibility, interactivity, and other technological features with VR. The R&D period of a new vehicle can be greatly shortened and the cost can be reduced by virtual design, simulated manufacturing, process analysis, and virtual testing. In a virtual environment sized the same as a real vehicle, dynamic adjustments of design details and the overall model can be achieved and a series of tests such as road tests, collisions, and wind tunnel tests can be performed. Currently, the US Internet of Vehicles (IoV) and automatic driving test base MCity intend to send virtual testing information to vehicle decision-making control systems in real scenarios to perform tests on different traffic scenarios. In addition, Audi, Ford, BMW, Chrysler, Toyota, Volvo, and other mainstream vehicle vendors are proactive in introducing the VR technology to vehicle R&D. Audi launched the VR-based virtual assembly line verification to enable pipeline workers to implement the assembly estimation and calibration of real products in 3D virtual space, significantly increasing the production and assembly efficiency. Ford uses the VR technology to check the appearance and interior design of a vehicle, and check specific details to optimize the ergonomic design. BMW plans to introduce the VR technology to the early stage of its vehicle R&D process. In the vehicle design stage, development and design teams based in different locations remotely collaborate through VR and help engineers rapidly modify the design draft based on simulated driving scenarios.

Based on industrial Internet/IoT platforms, VR is one of the keys to implementing Digital Twins. For example, the industrial software giant PTC integrates its core advantages in product design and PLM to the Thingworx platform and releases a digital mapping-based framework and a packaged solution. Relying on platform software such as Creo, Windchill, and Axeda, a completely symmetrical digital mirror of the physical world is constructed in the virtual environment. Such an environment forms the basis for integrating product R&D, manufacturing, and commercial promotion data. The VR solution Vuforia achieves interaction between data information and physical environments, offering the basis of phase data verification and service procedure referencing.

**Figure 3-9 Industrial Internet/IoT platforms**



### 3.4.2 In VR+ Medicine, VR is applied in surgery training/guidance, psychiatry, and rehabilitation.

The combination of VR and the medical industry is gradually expanding and will become one vital in the future medical industry. For example, Shanghai Ruijin Hospital successfully used the VR technology to broadcast a 3D laparoscopic operation live in 2016, marking the emergence of China's live VR operation. Doctors who were unable to attend the operation were able to learn techniques of the challenging operation remotely through VR headsets. Google tested its Google Glasses with multiple hospitals. Doctors used Google Glasses to project the CT scan and MRI results and scanned barcodes to obtain medicine information, improving medical efficiency. MindMaze used the vivid immersive experience of VR to help patients recover, such as helping patients with "phantom limb pain" to overcome their psychological barriers and offering stroke patients clinical treatment. According to Goldman Sachs, revenue of VR+ Medicine will reach USD 1.2 billion in 2020 and USD 5.1 billion in 2025 with up to 3.4 million users. Currently, doctors are reluctant to use VR, mainly due to hardware costs and software applicability.

### 3.4.3 In VR+ Games/Social Networking, VR gaming will become the biggest drive of the current VR industry.

VR and video games together will provide users with a more real and stronger sensory stimulation. The large user base and open attitude of key players towards new technologies make video games likely to be the first mass market to develop. Taking VR+ e-Sports as an example, the number of attendees of US e-Sports events reached 36 million in 2015, twice the number that attends NBA games. The global market share of e-Sports reached USD 86 billion in 2016. According to the Super Data report, the market share of global VR games market was expected to be USD 5.1 billion in 2016 and is on an upward trend. Currently, VR+ Games still have some issues to be resolved. The first issue is that VR differs a lot in interactive operations and content design compared to traditional games. Another issue is that the dizziness caused by prolonged wearing of a VR device affects the game experience. The third issue is that hardware requirements result in greater game costs.

VR marks the emergence of the social networking 2.0 era. VR social networking breaks the confines of traditional social networking, improving online social experience through virtual avatars, expression recognition, and more refined and rich communication modes. Currently, typical VR social networking products include High Fidelity and Facebook. High Fidelity inherited the concept of *Second Life* and aims to build a small-scale diversified virtual society, where people can have life-like and beyond social experience. Facebook also established a social VR department and launched the VR-related application SocialTrivia in consideration of merging VR and social networking.

### 3.4.4 In VR+ Movie/Live Broadcast, VR becomes a new form of display.

Movies are considered the "seventh art" after literature, drama, painting, music, dance, and sculpture. VR, as a new form of display, is capable of expressing richer details. VR movies provide audiences with immersive experience. VR poses continuous demands for higher image quality, that is, higher resolution, higher refresh rate, higher color depth, higher FOV, better 3D, and lower latency. The core characteristic of human-machine interaction is alienated due to the more game-like movie watching, where audiences can spontaneously choose the angle of view and even affect the direction of the scene, implementing "a thousand Hamlets in a thousand people's eyes." VR movies are far from mass popularization even if they are produced by well-known VR studios such as Oculus Story Studio and Baobab Studios. The main challenges hindering mass popularization include: The VR movie production and shooting technology are not yet mature, with complicated shooting process and high costs; VR terminals are expensive, resulting in a low user penetration rate; VR movie contents are relatively lacking; high-quality physical VR experience stores are not available on a large scale.

Live VR has become a new norm. According to statistics published by Goldman Sachs, the estimated market share of live VR will be USD 750 million in 2020 and USD 4.1 billion in 2025 with the number of users reaching 95 million. VR broadcast is widely used in sport events, news, concerts, and launch events. Many Internet sensations are also proactive in trying it. Currently, there are over 20 live VR platforms in China. Traditional OTT broadcast platforms (more than 200) have also gradually started to support the live VR function. WhaleyVR broadcast this season of Chinese Super League (CSL) live through VR together with China Sports Media and Flycat. Sports lenses, such as the 3D cableway shooting system and track shooting system, are first introduced to live VR. A 180-degree image covering the full view is displayed after onsite synthesizing. A concert by Faye Wong (a famous Chinese female singer/song writer), titled "Faye's Moments Live 2016," which was held in Shanghai, had 90 thousand people paying for the live VR. The broadcast platform of Tencent had over 2 thousand users watching the concert and over 2 million users reserved for the VR broadcast. In addition, NextVR, the world-renowned live VR company, has a number of excellent IP resources, for example, collaborated live broadcast of concerts with Live Nation, collaborated live broadcast of Democrats debates with CNN, and contracted live broadcast of International Champions Cup 2017 and other major sports events with Fox Sports.

**Figure 3-10** Some live VR companies and platforms

Live VR company	NextVR (Sports events and concerts)	Tencent (Concerts)	WhaleyVR (Variety shows and concerts)	LeTV (News conferences and fashion shows)	NYT VR (News events)
Live VR platform	Facebook, Twitch (Vreal), YouTube, Twitter, Huajiao, and 3D-BoBo				

## 4 Recommendations on the Development of the VR Industry for Operators

With continuous breakthroughs being made in key technologies such as near-eye display, perception and interaction, and rendering processing, the problem of VR sickness has gradually been overcome, and immersive user experience has continued to improve. The VR industry chain is becoming increasingly mature. The demand for further improvements to immersive experiences through upgrading network transmission technologies is on the rise. Now is the best time for operators to enter the VR industry. Operators need to keep pace with business development, leverage their distinct advantages to invest in the technological iteration of the VR industry, and cooperate with industry partners to promote the prosperity of the industry.

## **4.1 Operators Are One of the Core Links of the VR Industry Chain, Providing Assurance for VR Service Experience and Enabling Service Popularization**

The online distribution of high-quality VR content and real-time perception and interaction have much higher requirements on network transmission than traditional internet services. Based on more efficient network architectures and transmission technologies, stable, high-bandwidth, and low-latency network services will provide assurance for continuous improvements to immersive user experience. This poses new requirements on operators' network architecture and services, and only operators have the necessary capabilities to provide such assurance. The VR industry can achieve continuous upgrade, evolution, and popularization only with support from operators, who are one of the core links of the industry chain.

First, operators can provide assurance for VR service experience. New types of household Wi-Fi, large-capacity fiber-to-the-home (FTTH) services, and emerging 5G networks ensure high-quality broadband access anytime and anywhere, meeting VR service experience requirements of households and individuals. In addition to requirements for high bandwidth and low latency, VR applications in the industry also require E2E service isolation, SLA/QoS assurance, and fast service go-to-market. The enterprise private line and 5G network slicing services provided by operators will provide assurance for VR applications. For example, some operators have provided innovative services concerning the interactive live VR mode based on dedicated viewing locations. Users at those locations can use VR terminals to enjoy the live broadcast of performances at the main venue in a centralized manner. Operators use 5G E2E slicing or dedicated fixed VR service connections between the main venue and the dedicated viewing locations to provide an immersive and interactive experience with guaranteed QoS, and try new business models for ticket revenue and terminal cooperation.

Second, the large-scale user base, channels, and video service platforms enable operators to speed up the popularization of VR applications. Operators can provide VR services as upgraded video (IPTV) services, such as on-demand playbacks of VR content, VR live videos, and VR games, improving user experience and achieving revenue growth. Currently, some operators have upgraded their IPTV platforms and set-top boxes (STBs) to implement multi-screen live VR platforms of TV screens and VR terminals, and have piloted live VR services.

## **4.2 Operators Drive the Evolution of VR Service Forms: From Virtual Cinemas to Live VR, and Then to Cloud-based VR Services**

VR services take various forms, and operators can deploy services at a proper pace based on their advantages and the evolution of service experience.

With virtual cinemas, operators are starting their VR services. Video service innovation and experience improvement are both pointing towards a VR-dominant future due to the immersive features of VR and the availability of high-bandwidth and low-latency networks that permit VR services to be delivered. The continuous pursuit of high image quality, in aspects such as resolution, refresh rate, color depth, FOV, and depth rendering, matches operators' persistent focus on improving user experience of video services. In addition, VR video is currently the most mature form of VR services. Taking the existing IPTV service as an example, virtual cinemas can be the starting point for promoting IPTV VR applications. Operators can use the existing IPTV platform and the substantial number of HD/UHD videos to preferentially run services that can be rolled out, thereby asserting a well-established position in the VR market with minimized investment. Subsequently, 360-degree VR video content that delivers a more immersive experience is being developed to further enhance the appeal of VR services. In this way, VR videos can be a breakthrough point for operators to attract users, enhance user loyalty, and provide competitive service content.

By providing live VR services, operators can quickly cash in on platform capabilities. During major events such as sports games, artistic performances, breaking news, and product launch events, 360-degree live VR enables users to not only know what is happening on-site, but also participate in the events without going anywhere. 360-degree live VR will become an important way for people to watch and discuss

entertainment events and sports games. The booming fan economy will become a major driving force for the development of 360-degree live VR. Operators can cash in on VR services by providing value-added services (VASSs). Leveraging abundant channel resources and a large number of users, operators can provide live VR services as a type of VAS and charge fees, either by themselves or jointly with content providers, so as to obtain commercial benefits.

With Cloud VR gaming, operators can gain greater commercial benefits in the games industry, which has huge market potential. VR gaming brings more authentic and intense sensory stimulation to users. Thanks to the large user base and core players' open attitude to new technologies, VR video gaming is expected to become the VR service scenario that boasts the greatest monetization capability. Through cloud-based game content and rendering, operators can lower the user cost while ensuring low-latency interactions between the players of VR games, so as to attract a large number of mainstream user groups to Cloud VR gaming. In addition, operators can cooperate with content providers to build a games industry ecosystem and share more benefits in this huge market.

### 4.3 VR Service Development Recommendations for Operators

The VR industry is quickly transitioning from the entry-level immersion phase to the partial immersion phase. In the entry-level immersion phase, most VR users are early adopters. The VR industry has completed concept popularization, and provided consumer-oriented VR applications and industry applications that have preliminarily verified business scenarios. Meanwhile, bottlenecks in terminals, content, and networks have been exposed. Operators need to push forward network technology innovation (upgrade to Gigabit access and deployment of 5G networks) to drive evolution to the partial immersion phase and make data networks ready to promise an optimal VR experience, so that future networks will be more suitable for VR transmission and a wide variety of service scenarios. In this way, VR services can be popularized among mainstream users.

With the improving experience and popularization of VR services, VR terminals will become the next important entry to user network access and service experience. Operators should proactively embrace VR terminals to attract users, learn from the experiences of mature cooperation models like mobile phones and TV STBs to foster partnerships with leading VR terminal vendors, provide bundled packages to enhance commercial competitiveness, and deliver competitive service experiences through pipe-device collaboration. In doing so, operators can increase user loyalty and build competitive advantages.

By proactively conducting business model innovation and pilots, operators can promote ecosystem cooperation. The VR industry requires a wide variety and a great number of market participants. Collaborating with best-in-class partners for content and technology in innovative trials will be the only mechanism to determine the most successful business cases and ensure that ultimate immersive experiences are provided to meet user expectations – a great network without content is just a network, an excellent VR video film is lost if shown on a poor quality terminal. Operators should make full use of their unique advantages in the industry chain to actively innovate and pilot VR service cooperation models, so as to differentiate from competitors in terms of VR experience quality and create opportunities for the ecosystem to cooperate with third-party content providers and service providers. These pilots can also help identify bottlenecks in the industry and specify which parties should be responsible for addressing said bottlenecks. This ensures that operators can take the lead in business model creation as VR enters a new stage of commercial use.

CAICT

TEL: 86-10-62301618

FAX: 86-10-62304346

ADD: No.52, Hua Yuan Bei Road, Haidian District, Beijing, China

Post Code: 100191



WEB: <http://www.caict.ac.cn/english/>

Author: Chen Xi

Email: [chenxi1@caict.ac.cn](mailto:chenxi1@caict.ac.cn)