Technology Reports | XR | AR/MR Cloud | Cyber-physical Fusion

AR/MR Cloud Technology to Provide Shared AR/MR Experiences across Multiple Devices

Communication Device Development Department Kouki Hayashi Osamu Goto
Shinji Kimura Taishi Yamamoto

One of the new services expected to emerge in the 5G era is the "AR/MR Cloud". This article describes NTT DOCOMO's R&D efforts in the three major functions - the self-localization function, the spatial 3D model generation/management function, and the content space management function - necessary to realize the AR/MR Cloud that provides shared AR/MR experiences across multiple devices.

1. Introduction

NTT DOCOMO is working toward the realization of "cyber-physical fusion*1", which aims to improve and optimize the value of services and lifestyles in real space by converting information on people, things, and events in real space into data and predicting the future by utilizing AI technology on the collected data. Cyber-physical fusion refers to a world in which information about people, things, and events in real space is acquired through various sensors, and the information is aggregated in digital space in the cloud using communications such as 5th generation mobile communication systems (5G) to create digital twins*2. XR*3 are promising technologies for providing new services in the 5G era and more realistic experiences of the world of cyber-physical fusion. XR is a generic term for technologies such as Virtual Reality (VR)*4, Augmented Reality (AR)*5, and Mixed Reality (MR)*6 that provide new experiences by fusing virtual space with real space. XR technologies enable new experiences peering into the world of digital twins using Head-Mounted Displays (HMD)

*1 Cyber-physical fusion: Services and systems for realizing a better and more advanced society by collecting information in real space (physical space) from various sensors, etc. and linking it to virtual space (cyber space).

*2 Digital twin: A real-time reproduction in digital virtual space of the position, shape and various sensor information of an object in the real world.

such as glasses-type AR/MR devices, VR goggles or smartphones. Notable among these and attracting great attention in recent years are AR/MR Cloud technologies, which enable interactive and shared AR/MR experiences across multiple devices by superimposing cyberspace created with digital twins by aggregating various sensor information into real space.

"AR/MR Cloud" refers to technological infrastructure for superimposing AR/MR content on real space and enabling interactive sharing of AR/MR content across multiple devices. Services achievable with AR/MR Cloud technologies include:

- Highly immersive and interactive games such as virtually painting buildings in real space
- E-commerce that enables users to virtually draw furniture in a real room to check the fit before buying actual furniture or other items
- Advertising services that display advertisements and coupons, etc. that match the user's tastes and preferences or situation on storefronts, building walls, etc. as the user walks around a town
- Bulletin board services that enable users to freely write comments on the walls of buildings

At "DOCOMO Open House 2020 - Dawn of the 5G era and the Future Beyond -" held at Tokyo International Exhibition Center Aomi Exhibition Hall in January 2020, we held a demonstration that enabled visitors to experience the world of the AR/MR Cloud using AR/MR Cloud technologies developed by NTT DOCOMO R&D **(Figure 1)** [1].

(a) Visitors enjoying content on Magic Leap 1, Mirage Solo and iPad.

(b) Actual content displayed on iPad

Figure 1 The DOCOMO Open House 2020 Exhibition

*3 XR: A general term for technologies such as VR, AR, and MR that provide new experiences through the fusion of virtual space and real space.

*4 VR: Technology that enables users to immerse themselves in a virtual space separate from real space using a head-mounted display or other device.

*5 AR: Technology that uses digital technology to add information to real space using a glasses-type terminal or other device.

*6 MR: Technology that enables a virtual world to feel more real by more closely integrating real space and virtual space through the use of a glasses-type terminal or other device.

This demonstration showed three types of devices, Magic Leap 1*7, Mirage Solo (video pass-through type VR goggles*8), and iPad, sharing AR/MR content in an exhibition booth that mimics a town. Many visitors wanted to try Magic Leap 1, and many expressions of wonderment were heard as they experienced the world of the AR/MR Cloud with this cutting-edge spatial computing*9 device.

This article describes the technologies required to realize the world of the AR/MR Cloud developed by NTT DOCOMO R&D and introduces future prospects.

2. The Set of Functions Necessary to Realize the AR/MR Cloud

Figure 2 shows the set of functions necessary to realize the AR/MR Cloud. This set of functions consists of three major functions - the self-localization function, the spatial 3D model generation/management function, and the content space management function, which are described below.

2.1 Self-localization Function

Self-localization is technology that recognizes the exact position and orientation of each AR/MR device in real space. Recognizing position makes it possible to superimpose AR/MR content as if it fits into the real space while sharing the position of each device makes it possible to provide interactive AR/MR experiences.

One existing technology that could be used is Global Positioning System (GPS), but the positioning error of GPS is large, especially in indoor environments, and even in an open sky environment with no occlusions (obstructions), GPS has an error of several meters. Therefore, the positioning recognition accuracy of GPS is insufficient to provide

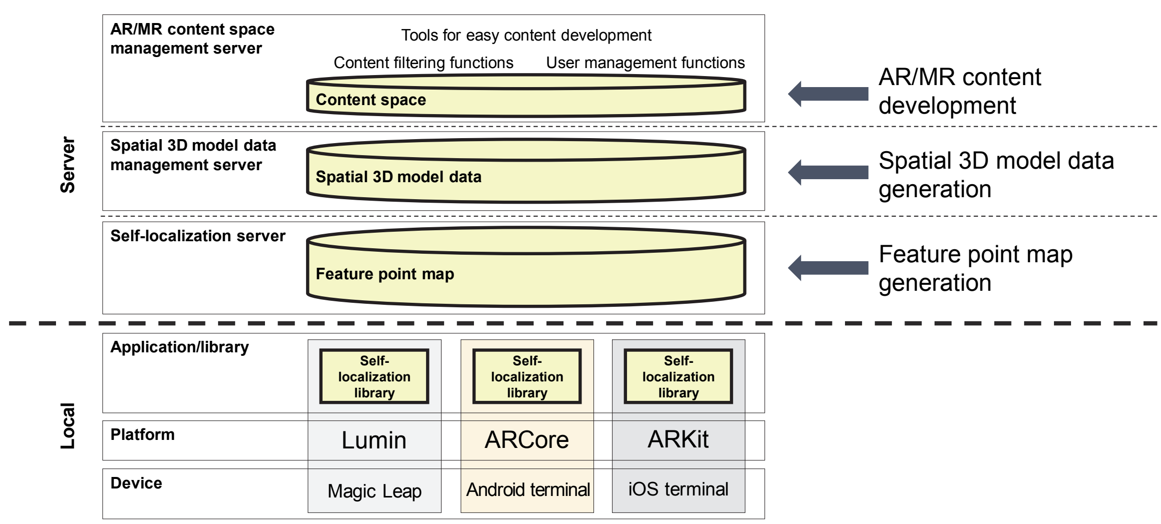

Figure 2 Set of Functions Necessary to Realize the AR/MR Cloud

*7 **Magic Leap 1:** "MAGIC LEAP 1", the Magic Leap logo, and all other trademarks are trademarks of Magic Leap, Inc.

*8 **Video pass-through type VR goggles:** VR goggles that project images recognized by a front-mounted camera onto a display inside the goggles, enabling the wearer to see the outside environment even while wearing the goggles.

*9 **Spatial computing:** Technologies that recognize objects and spaces in the real world and fuses them with digital information. Using these technologies makes it possible to transcend the limitations of two-dimensional displays, integrate real space and the digital world into one, and interact with the digital world in the same way as the real space.

AR/MR experiences in which AR/MR content is accurately superimposed in real space. Another method would be to use AR markers*10 to recognize absolute coordinates in real space, but the issue is AR markers need to be placed so that self-localization can be performed.

To address these issues, NTT DOCOMO has developed a system that enables self-localization by applying Simultaneous Localization And Mapping (SLAM) technology*11, which is a feature point-based positioning technology that uses feature points in images obtained from cameras. The basic process flow of this system is shown in **Figure 3**. The system uses a stereo camera and takes pictures of the area in which to place the AR/MR content. Next, it extracts feature points and creates the feature point map for aligning the AR/MR content with the real world. Then, the system matches the extracted feature points to the coordinates of the real space, makes global coordinates*12, and uploads the feature points matched to the coordinates of the real space (a feature point map) to the server. Meanwhile, each device has a built-in self-localization library, and periodically sends camera images (one every few seconds) to the self-localization server. The server extracts the feature points of the images sent from each device, checks them against the feature point map, and calculates the global coordinates where the images were taken. Using the built-in self-localization library, each device performs real-time tracking of each device's local coordinates and corrects the coordinates in real time using the global coordinates sent by the server. Maintaining a common global coordinate system and

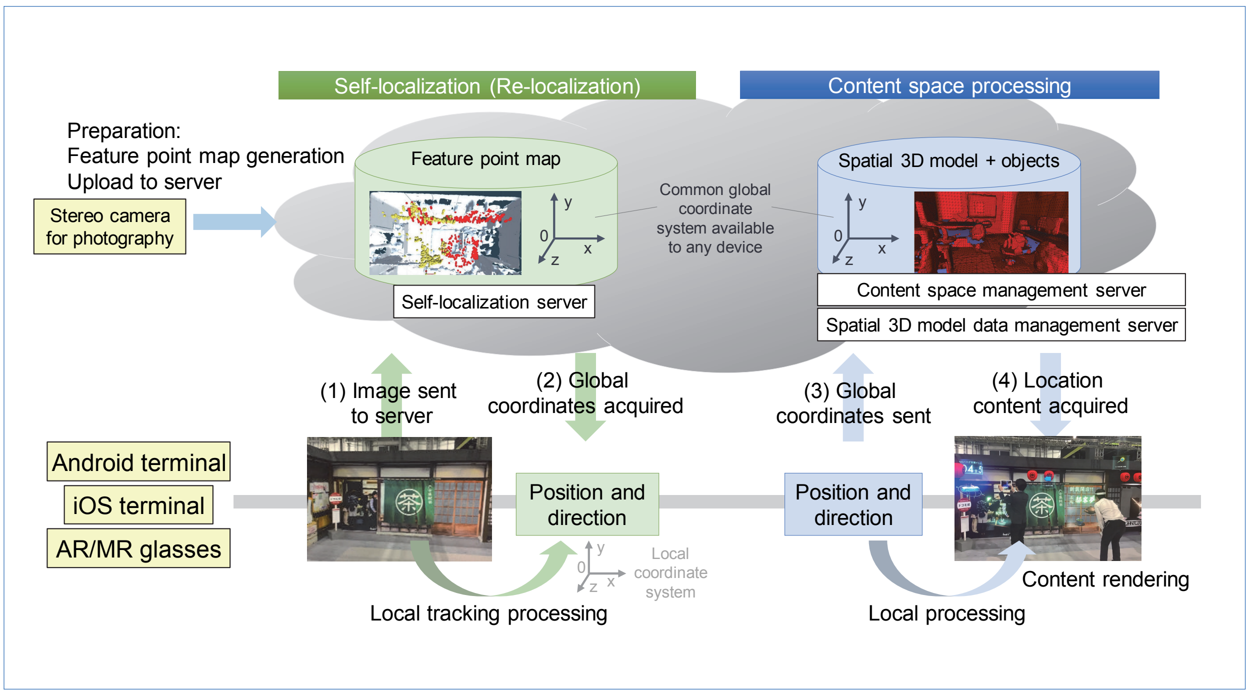

Figure 3 Basic process flow of the system

---

*10 **AR marker:** A mark or image used to display digital content on a device screen. Reading AR markers with image recognition technology makes it possible to recognize the location of the AR markers.

*11 **SLAM technology:** Technology that uses device camera image and sensor information to create a map of a surrounding environment while recognizing the device's own position.

*12 **Global coordinates:** A coordinate system that represents the entire space where AR/MR content is placed. Transforming the local coordinate system representing the position of each device into a global coordinate system through self-localization makes it possible to view shared AR/MR content across multiple devices.

linking data between the self-localization server, spatial 3D model data management server, and content space management server makes it possible to draw AR/MR content precisely aligned to the location.

In this way, self-localization using SLAM technology enables more accurate position recognition than GPS, without the need to arrange AR markers, etc. Due to the characteristics of SLAM technology, there are some environments where self-localization is easy and others where it is difficult. Thus, performance depends on how the feature point map is created. Currently, the system is capable of self-localization indoors and in some outdoor environments.

In addition, there are issues such as the need to capture images for feature point maps in advance, the photographic know-how required to create highly accurate feature point maps, and the fact that self-localization may not be possible due to changes in the surrounding environment, etc.

2.2 Spatial 3D Model Generation/ Management Functions

Spatial 3D model data is used for two main purposes (**Figure 4**). One is to express physical phenomena such as occlusions and bounces of AR/MR content, and the other is to increase visibility for AR/MR content developers for consideration of how and where to place AR/MR content when developing it.

The spatial 3D model data used to represent physical phenomena is not point cloud data, which is a collection of points, but mesh data, which is a collection of surfaces. In the example of throwing a ball in AR/MR content, aligning and arranging transparent surfaces represented by mesh data to match real space makes it possible to express the

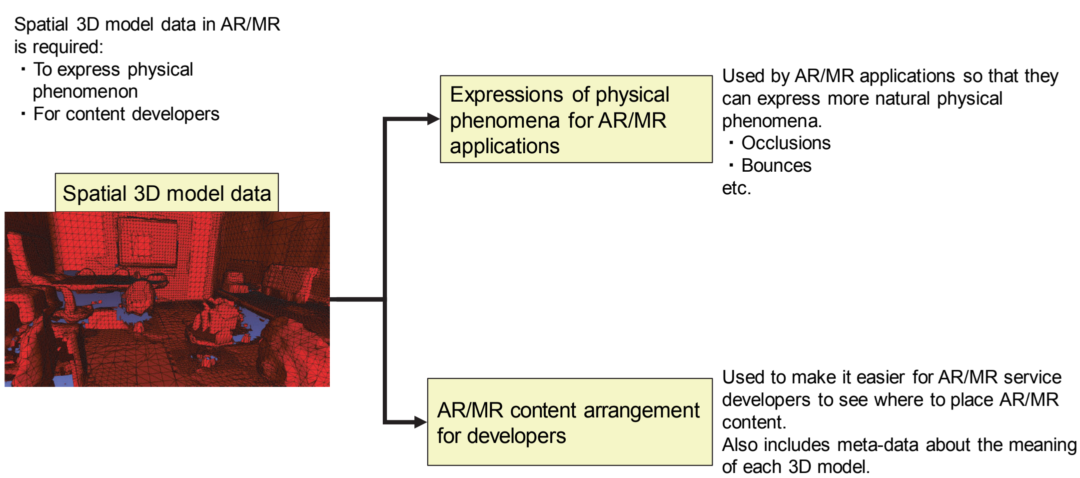

Figure 4　Uses of spatial 3D model data

ball bouncing off walls (bounce) or express rendering (occlusion) such that the ball cannot be seen on the other side of an obstacle. Such expression of physical phenomena is a very important technology to make AR/MR content feel more realistic.

Generating spatial 3D model data by photographing real space in advance and managing it on a spatial 3D model data server makes it possible to share spatial 3D models among users and express physical phenomena through spatial 3D models. As spatial 3D model data generally involves large file sizes, it is important for real-time interaction to ensure the minimum accuracy required to represent physical phenomena and ensure that the file size does not become too large.

Also, developers need to know where cyberspace is in real space so they can place AR/MR content. Since the spatial 3D model data used here only requires a certain level of understanding of locations, point cloud data or mesh data can be used, although high accuracy locations and shapes may be required depending on the content to be placed.

2.3 Content Space Management Function

The content space management function manages the tastes and preferences of device users, as well as attributes data and current status (user management function), and selects and outputs content according to the various attributes and states of the users of each device (content filtering function) (Fig. 2). This function also provides tools to make it easy to develop this content (a user-friendly development environment). The content space in which AR/MR content is placed is managed with a common coordinate system with the feature point map used in self-localization technology and the spatial 3D model data of real space, and has an interface for smooth linking with the self-localization function and the spatial 3D model management function.

3. Conclusion

This article has described the self-localization function, spatial 3D model generation/management function, and the content space management function for realizing the AR/MR Cloud.

NTT DOCOMO R&D is making efforts to develop technologies to realize an AR/MR Cloud world where all users of various devices, including glasses-type AR/MR devices, smartphones, and video pass-through VR goggles, can experience shared AR/MR content. We believe that the world of the AR/MR Cloud - spatial computing where AR/MR content in cyberspace is superimposed onto real space - can provide a completely new and never-before-seen experience. We will continue to develop toward the realization of a world in which all users of AR/MR devices can experience the AR/MR Cloud.

Please go to the official NTT DOCOMO website to find out more about NTT DOCOMO's XR efforts [2].

REFERENCES

[1] M. Tamaoki: "DOCOMO Open House 2020 - Dawn of the 5G era and the Future Beyond -," NTT DOCOMO Technical Journal, Vol.22, No.1, pp.52-60, Jul. 2020.

[2] NTT DOCOMO: "docomo XR|NTT DOCOMO." http://xr.docomo.ne.jp/