INTERNATIONAL STANDARD

ISO/IEC 18040

First edition
2019-05

Information technology — Computer graphics, image processing and environmental data representation — Live actor and entity representation in mixed and augmented reality (MAR)

Technologies de l'information — Infographie, traitement de l'image et représentation des données environnementales — Représentation d'acteurs et d'entités réels en réalité mixte et augmentée (MAR)

Reference number
ISO/IEC 18040:2019(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see http://patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 24, *Computer graphics, image processing and environmental data representation*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

This document defines the scope and key concepts of a representation model for a live actor and entity (LAE) to be included in a mixed and augmented reality (MAR) world. The relevant terms and their definitions, and a generalized system architecture, together serve as a reference model for MAR applications, components, systems, services and specifications. It defines representing and rendering an LAE in an MAR scene, and interaction interfaces between an LAE and objects in an MAR scene. It defines a set of principles, concepts and functionalities for an LAE applicable to the complete range of current and future MAR standards. This reference model establishes the set of required modules and their minimum functions, the associated information content, and the information models that shall be provided and/or be supported by a compliant MAR system. It includes (but is not limited to) the following content:

— an introduction to the mixed and augmented reality standards domain and concepts;

— a representation model for including an LAE in an MAR scene;

— 3D modelling, rendering and simulation of an LAE in an MAR scene;

— attributes of an LAE in an MAR scene;

— sensing representation of an LAE in an MAR scene;

— representation of the interfaces for controlling an LAE in an MAR scene;

— functionalities and base components for controlling an LAE in an MAR scene;

— interactive interfaces between an LAE and an MAR scene;

— interface with other MAR components;

— relationship to other standards;

— use cases.

The objectives of this document are as follows:

— provide a reference model for LAE representation-based MAR applications;

— manage and control an LAE with its properties in an MAR environment;

— integrate an LAE into a 2D and/or 3D virtual scene in an MAR scene;

— achieve interaction of an LAE with a 2D and/or 3D virtual scene in an MAR scene;

— provide an exchange format necessary for transferring and storing data between LAE-based MAR applications.

This document has the following document structure:

— Clause 4 describes the concepts of LAE-based systems represented in MAR.

— Clause 5 illustrates how a sensor captures an LAE in a physical world and a virtual world.

— Clause 6 describes mechanisms to track the position of an LAE and specifies the role of a spatial mapper between physical space and the MAR space.

— Clause 7 describes mechanisms to recognize the behaviour of an LAE and specifies an association or event between an MAR event of an LAE and the condition specified by the MAR content creator.

— Clause 8 describes a scene, which consists of a virtual scene, sensing data, a spatial scene, events, targets and so on, for an LAE.

— Clause 9 describes how the MAR scene system renders the scene, LAE mapping, event and so on for presentation output on a given display device.

— Clause 10 describes types of displays, including monitors, head mounted displays, projectors, haptic devices and sound output devices for displaying an LAE in an MAR scene.

— Clause 11 identifies and describes virtual LAE, such as virtual 3D model (avatar) and virtual LAE, such as real human model in an MAR system.

— Clause 12 makes statements regarding any system performance related issues of an LAE in MAR.

— Clause 13 makes statements regarding any operational safety related issues of an LAE in MAR.

— Clause 14 makes statements regarding any conformance related issues of an LAE in MAR.

— Annex A gives examples of representative LAE representation systems in MAR.

Information technology — Computer graphics, image processing and environmental data representation — Live actor and entity representation in mixed and augmented reality (MAR)

1 Scope

This document defines a reference model and base components for representing and controlling a single LAE or multiple LAEs in an MAR scene. It defines concepts, a reference model, system framework, functions and how to integrate a 2D/3D virtual world and LAEs, and their interfaces, in order to provide MAR applications with interfaces of LAEs. It also defines an exchange format necessary for transferring and storing LAE-related data between LAE-based MAR applications.

This document specifies the following functionalities:

a) definitions for an LAE in MAR;

b) representation of an LAE;

c) representation of properties of an LAE;

d) sensing of an LAE in a physical world;

e) integration of an LAE into a 2D/3D virtual scene;

f) interaction between an LAE and objects in a 2D/3D virtual scene;

g) transmission of information related to an LAE in an MAR scene.

This document defines a reference model for LAE representation-based MAR applications to represent and to exchange data related to LAEs in a 2D/3D virtual scene in an MAR scene. It does not define specific physical interfaces necessary for manipulating LAEs, that is, it does not define how specific applications need to implement a specific LAE in an MAR scene, but rather defines common functional interfaces for representing LAEs that can be used interchangeably between MAR applications.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 18039, *Information technology — Computer graphics, image processing and environmental data representation — Mixed and augmented reality (MAR) reference model*

3 Terms, definitions and abbreviated terms

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 18039 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1.1
augmented object
object with augmentation

3.1.2
geographic coordinate system
coordinate system which is provided by sensor devices for defining a location of *LAE* (3.1.4)

3.1.3
head mounted display
HMD
device which displays stereo views of virtual reality

Note 1 to entry: It has two small displays with lenses and semi-transparent mirrors which can adapt to the left and right eyes.

3.1.4
live actor and entity
LAE
representation of a living physical or real object, such as a human being, animal or bird, in the MAR content or system

Note 1 to entry: A live actor can be animated, moved and interacted with virtual objects in an MAR scene by capturing gesture from a camera. Entity refers to 3D objects and entities that exist in MAR content.

3.1.5
LAE recognizer
MAR component that recognizes the output from an *LAE capturer* (3.1.6) and an LAE sensor, then generates MAR events based on conditions indicated by the content creator

3.1.6
LAE capturer
MAR component that captures an *LAE* (3.1.4) in a virtual world and a physical world, which includes depth cameras, general cameras, 360° cameras and so on

Note 1 to entry: LAE's information can be processed by an LAE recognizer and an LAE tracker to extract background or skeleton.

3.1.7
LAE tracker
MAR component (hardware and software) that analyses signals from *LAE capturers* (3.1.6) and sensors and provides some characteristics of a tracked *LAE* (3.1.4) (for example position, orientation, amplitude, profile)

3.1.8
physical camera coordinate system
coordinate system which is provided by a camera for capturing *LAE(s)* (3.1.4) in physical world

3.1.9
physical coordinate system
coordinate system that enables locating an *LAE* (3.1.4) and is controlled by a geospatial coordinate system sensing device

3.1.10
virtual actor and entity
VAE
virtual reality representation of an *LAE* (3.1.4)

Note 1 to entry: The virtual actor and entity is obtained by a 3D capturing technique and can be reconstructed, transmitted or compressed in the MAR scene. A virtual actor and entity can be captured in one place or transmitted to another place in real time using holography technology.

3.1.11
world coordinate system
universal system in computer graphics that allows model coordinate systems to interact with each other

3.2 Abbreviated terms

For the purposes of this document, the abbreviated terms given in ISO/IEC 18039 and the following apply.

DDR	dance dance revolution
EID	event identifier
FOV	field of view
GNSS	global navigation satellite system
LAE-MAR	live actor and entity representation in mixed and augmented reality
RGB	red, green, blue
SDK	software development kit
SID	sensor identifier
UI	user interface
UTM	universal transverse mercator
VR	virtual reality

4 Concepts of LAE representation in MAR

4.1 Overview

As illustrated in ISO/IEC 18039, MAR represents a continuum which encompasses all domains or systems that use a combination of reality (for example live video) and virtuality representations (for example computer graphic objects or scene) as its main presentation medium[1][2]. Figure 1 illustrates the MAR that is defined according to a mixture of reality and virtuality representations excluding pure real environment and pure virtual environment with viewpoints of an LAE representation. The real environment refers to the physical world environment where an LAE and physical objects are located. Virtual environment commonly refers to virtual reality, that is the computer-generated realistic images and hypothetical world that replicate a real environment. Augmented reality refers to the view of the real-world environment whose elements include LAE and objects that can be augmented by computer-generated sensory, and augmented virtuality is the virtual environment within which physical world elements including LAE can be mapped and interacted. In Figure 1, an LAE wears an HMD device to see the virtual world and interacts directly with virtual objects.

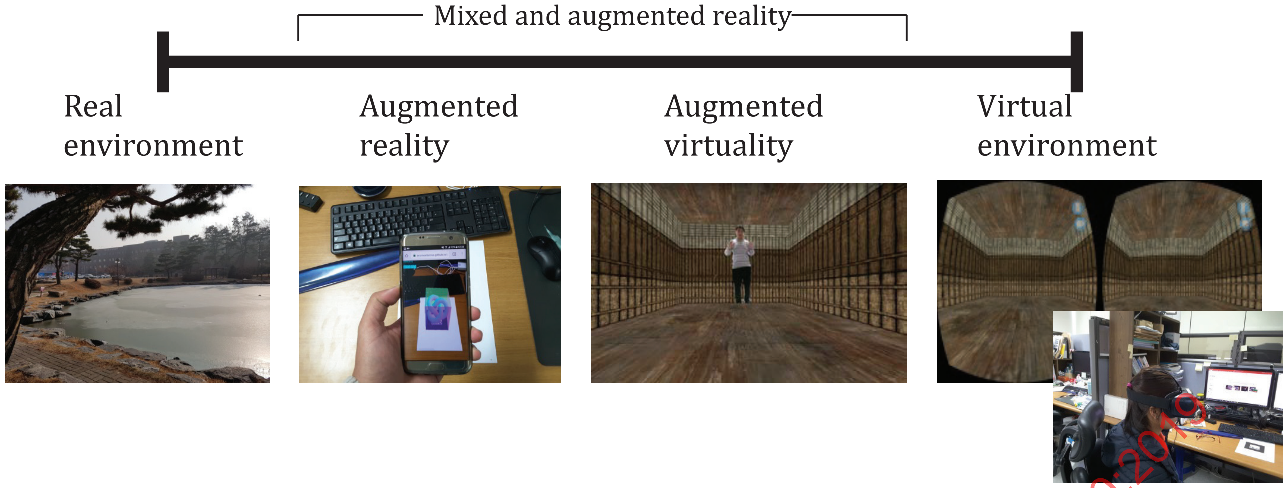


Figure 1 — Mixed and augmented reality (MAR)

This clause describes the concepts of LAE representation in an MAR scene based on the MAR reference model (MAR-RM) of ISO/IEC 18039, which includes objectives, embedding, interaction and functions of the system for representing an LAE in an MAR scene. In general, an actor is an individual who portrays a character in a performance. In this case, an actor represents a human captured by a depth camera or a general camera, which can then perform actions that are embedded into an MAR scene. A 3D object that exists in an MAR scene and can interact with a live actor is called an entity. The entity can be moved or interact with an actor's motion via an event mapper. An LAE in this document is defined as a representation of a physical living actor and an object in an MAR content or system. For example, human beings, birds and animals are all represented as LAEs in an MAR scene.

Figure 2 shows the examples of LAE representation in an MAR scene which consists of 2D virtual world and 3D virtual world that can be described as the following.

Figure 2 a) shows an LAE integrated into a 2D virtual world that is a real or virtual image. The LAE can be captured from general camera and/or depth camera sensors. This subfigure shows a real-like action where a man is captured by cameras in a green screen studio and is integrated as an actor into a 2D virtual world image of the White House.

Figure 2 b) shows multiple LAEs integrated into a 3D virtual world. This scenario can be applicable in various situations, such as news studios, education services, virtual surgical operations or games[3][4]. It supports an integrative combination application of 3D videoconferencing, reality-like communication features, presentation/application sharing and 3D model display within a mixed environment.

Figure 2 c) shows an MAR scene constructed by integrating a 3D virtual world and a live actor[5]. The live actor interacts with objects in the 3D virtual world by using a joystick or by motion captured by a depth camera. An HMD device is used to display 360° 3D views, including real time and real-like action, in the virtual world[6]. The figure shows a man in the studio wearing an HMD through which he sees the bow sling training field and handling a joystick. By handling the joystick, it appears that he is handling the arrow and bow sling. As a result, he can shoot at objects in the virtual world.

Figure 2 d) shows a virtual actor and entity with representation in the physical world[7]. The virtual LAE and LAE can communicate and interact with each other, for example to have a natural face-to-face conversation. When combined with MAR, this technology allows an LAE to see, hear and interact with a virtual LAE in a 3D virtual space, just like a real presentation in physical space.

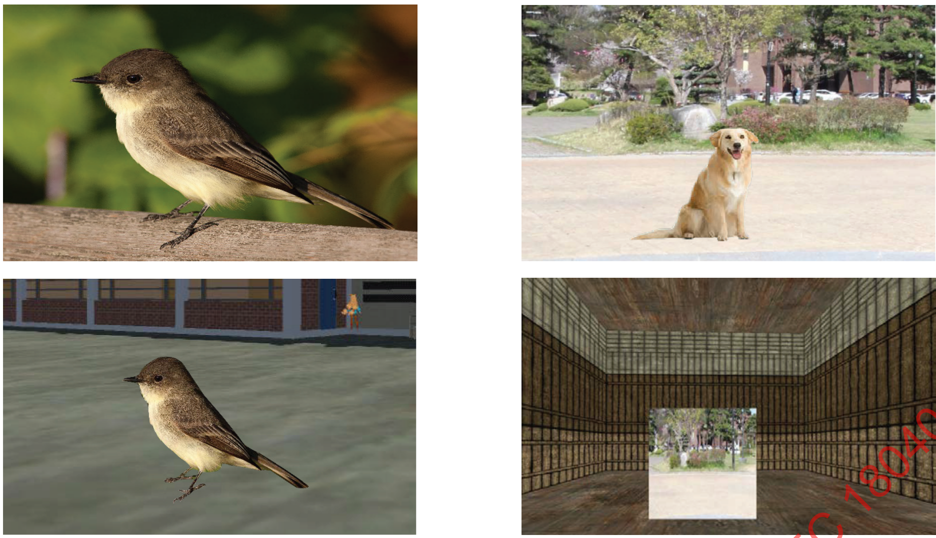
Figure 2 e) describes the LAE representation as a bird and a dog from which it can be inferred that LAE can be animals, birds or humans to be represented in MAR scene.

a) An LAE integrated into a 2D video virtual world after Chroma-keying[8]

b) Two LAEs integrated into a 3D virtual world

c) An LAE interacting with a virtual object in a 3D virtual world[5]

d) Virtual representation of an LAE in an MAR scene[7]



e) LAE representation as a bird and a dog

Figure 2 — Examples of LAE representation in an MAR scene

Once an LAE in the physical world is integrated into a 3D virtual world, its location, movements and interactions should be represented precisely in the 3D virtual world. In an MAR application, an LAE that needs to be embedded in a 3D virtual world shall be defined and then information, such as the LAE's location, actions and sensing data from a handled device, shall be able to be transferred between the physical world and the 3D virtual world, and between MAR applications. This document aims at defining how an LAE can be managed and controlled with its properties in a 3D virtual world; how an LAE can be embedded in a 3D virtual world; how an LAE can interact with virtual cameras, virtual objects and AR content in a 3D virtual world; and how MAR application data can be exchanged in heterogeneous computing environments.

4.2 Components

4.2.1 General

An LAE in an MAR scene can be captured from the physical world, then represented in a 3D virtual world, and can interact with cameras, objects and AR content in the 3D virtual world according to an input of sensing information.

In order to provide a 3D virtual world with the capability of representing an LAE based on the MAR-RM, the MAR system requires the following functions:

— sensing of an LAE in a physical world from input devices such as a (depth) camera;

— sensing of information for interaction from input sensors;

— recognizing and tracking an LAE in a physical world;

— recognizing and tracking events made by LAEs in a physical world;

— recognizing and tracking events captured by sensors in a physical world;

— representation of the physical properties of an LAE in a 3D virtual world;

— spatial control of an LAE in a 3D virtual world;

— an event interface between an LAE and a 3D virtual world;

— composite rendering of an LAE into a 3D virtual world.

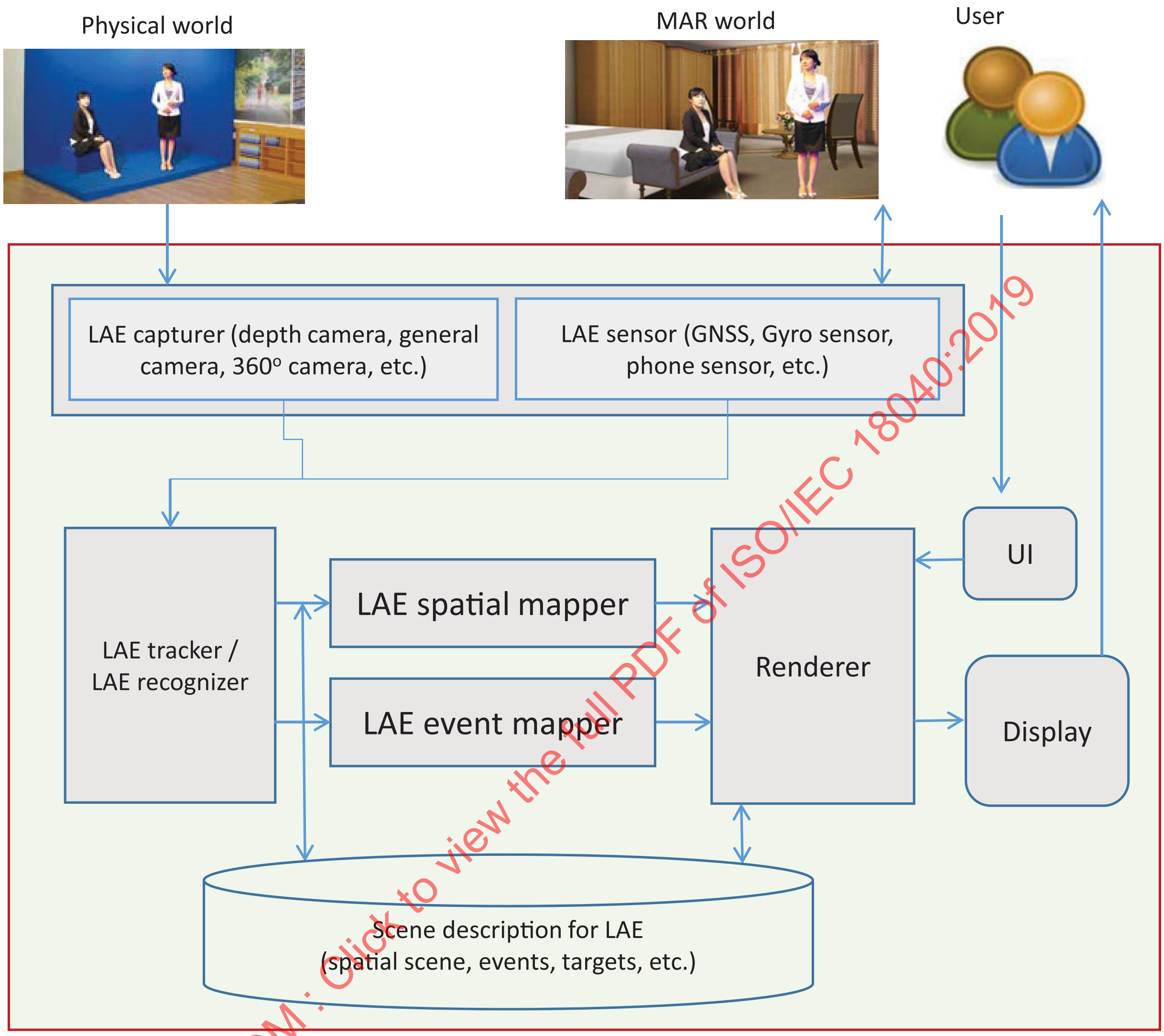


Figure 3 — System framework for an LAE in an MAR scene

Conceptually, as shown in Figure 3, the system for implementing an MAR scene with LAEs includes several components necessary for processing the representations and interactions of the LAEs to be integrated into the 3D virtual world. This clause provides a conceptual overview of these components. Each component is described in more detail in Clauses 5 to 10.

4.2.2 LAE capturer and sensor

An MAR scene with an LAE can receive a captured image, or a sequence of captured images, and sensing data triggered by the LAE in the physical world. The data for the LAE representation can be obtained from sensing devices/interfaces and is classified into two types: one for the LAE, and one for the sensing data for the LAE's actions. The data for the LAE can be used to perform both spatial mapping and event mapping of the LAE into the 3D virtual space, and the sensing data can be used to perform event mapping between the LAE and the 3D virtual environment.

The MAR system can use devices such as Web cameras and depth-like devices. For Web cameras, the sequence of images is captured and the images themselves are used as input. A depth camera is a movement sensing interface device for capturing the movement of an LAE. The device consists of an

RGB camera, 3D depth sensors, multi-array microphones and a motorized tilt. The device uses an IR (infrared radiation) camera and an IR laser projector to take 3D depth information.

4.2.3 LAE recognizer

Recognition refers to finding and identifying the actions of an LAE in the captured image, or a sequence of captured images, and the sensing data. When a sensor processes the output of an MAR component and generates MAR events, the LAE recognizer identifies each event and matches the event with an event ID from the database. The LAE recognizer analyzes signals of LAE motion or interaction from the physical world by comparing them with a local or remote signal. Then, an event function is processed.

4.2.4 LAE tracker

Modules are required to pre-process the captured image, or a sequence of captured images, and the sensing data obtained from the sensing devices and interfaces. Pre-processing examples include chroma-keying, colour conversion and background removal. Tracking an LAE refers to finding the location of the LAE in each image of the sequence, which can be implemented by an image processing library.

4.2.5 LAE spatial mapper

An LAE in a physical world is embedded into a 3D virtual world in an MAR application. The spatial mapper's role is to support more natural movement of the LAE within the 3D virtual world. The LAE spatial mapper provides spatial information, such as position, orientation and scale, between the physical world space and the MAR scene space by applying transformations for calibration. The LAE spatial mapper maps the physical space and the LAE into the MAR scene by supplying explicit mapping information. The mapping information can be modelled by characterizing the translation process of a sensor's given information.

4.2.6 LAE event mapper

An LAE performs actions in the physical world that can trigger events. These actions can be reflected in the 3D virtual world where an LAE is participating and interacting with virtual cameras, virtual objects and AR content. The LAE event mapper's role is to support the actions of the LAE. It creates an association between an MAR event and the events which are identified or recognized by the LAE recognizer.

4.2.7 Renderer

The results of spatial and event mappers for an LAE are transferred into the rendering module. This module integrates the results and a 3D virtual scene and displays the final rendering result of the integrated outcomes. The MAR scene can be specified by various capabilities of the renderer; thus, the scene can be adapted, and simulation performance can be optimized.

4.2.8 Display and user interface

The final rendering result of the integrated outcomes of LAEs and an MAR scene can be displayed on a variety of devices, such as monitors, head mounted displays, projectors, scent diffusers, haptic devices and sound speakers. A user interface provides users with a way to modify the state of the MAR scene. WebXR[9] is proposed to display a rendered scene on an HMD device[10].

4.2.9 Scene representation

The MAR scene describes all information related to LAEs in the MAR environment. This information consists of sensing data, spatial scene, events, targets and so on. The MAR scene observes the spatiality of physical and virtual objects and has at least one physical object and one virtual object.

Table 1 summarizes the inputs and outputs for the components of an LAE-MAR system based on the discussion so far.

Table 1 — Attributes of an LAE in each component

Component	Type	
	Input	Output
LAE capturer/ sensor	Physical world signal	Sensor data related to representation of an LAE
LAE recognizer	Raw or processed signals representing the LAE (provided by sensors) and target object specification data (reference target to be recognized)	At least one event acknowledging the recognition
LAE tracker	Sensing data related to the representation of an LAE	Instantaneous values of the characteristics (pose, orientation, volume and so on) of the recognized target signals
LAE spatial mapper	Sensor identifier and sensed spatial information	Calibrated spatial information for an LAE in a given MAR scene
LAE event mapper	MAR event identifier and event information	Translated event identifier for an LAE in a given MAR scene
Renderer	MAR scene data	Synchronized rendering output (for example visual frame, stereo sound signals, motor commands and so on)
Display/UI	Render scene data/user actions	Display output/response of user actions

5 LAE capturer and sensor

5.1 Overview

An LAE in a physical world can be captured from hardware and (optionally) software sensors that are able to measure any kind of physical property. As referenced in ISO/IEC 18039, two types of sensors, “capturer” and “sensor”, are used to embed the LAE into the virtual world and to perform an interaction between the LAE and objects in the virtual world. The most common “capturer” sensors are video and/or depth cameras which capture a physical world as a (depth) video containing live actors and entities. The video is used to extract the LAE and its actions to be processed by the recognizer and tracker components. The actions, especially, can affect interaction between the LAE in the physical world and objects in the virtual world. The target physical object can generate physical or nonphysical data which can be captured from a “sensor”. The (non) physical data can be used to detect, recognize and track the target physical object to be augmented, and to process the interaction. The sensing data is input into the recognizer and tracker components.

5.2 Computational view

5.2.1 General

An LAE capturer/sensor module defines the functionalities of components and their interfaces for sensing an LAE. It specifies the services and protocols that each component exposes to the environment. This module provides two types of sensors, “capturer” and “sensor”, for sensing an LAE.

5.2.2 LAE capturer

Various types of cameras, including general cameras, depth cameras and 360° cameras, can be used to capture an LAE. Figure 4 shows an LAE capturer that captures the physical world, including an LAE, as a video, depth image, and skeleton which is used in an MAR scene. An LAE can be extracted in a pre-processing step for the LAE tracker and/or recognizer by using video processing methods such as background removal, filtering and chroma-keying. An extracted LAE can not only be embedded into a virtual world but also be used to identify an MAR event, that is, be used as input to the LAE tracker and/or recognizer.

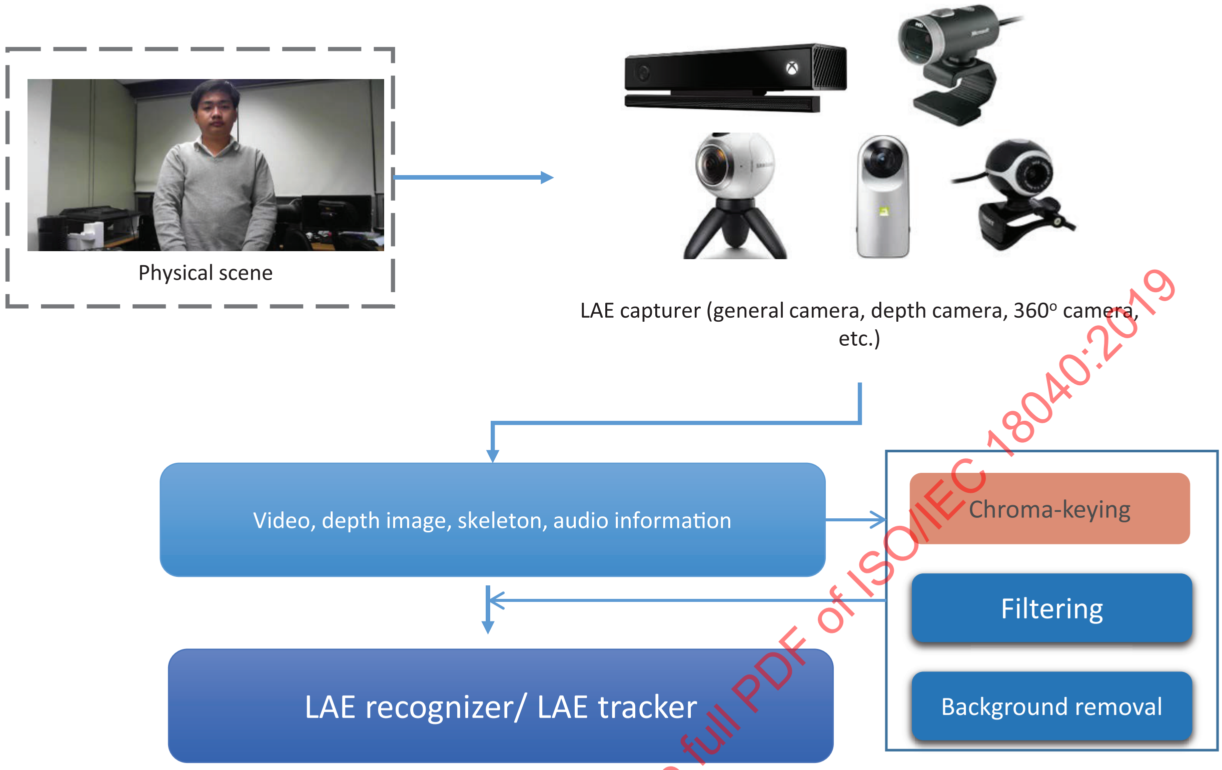


Figure 4 — LAE capturer

5.2.3 LAE sensor

An LAE sensor can measure various physical properties and interpret and convert the observations into digital signals related to the LAE. Figure 5 shows sensor capturing actions related to LAE activities. It shows that the captured data can only be used to compute the context in the LAE tracker and LAE recognizer, or it can be used to both compute the context and contribute to the composition of the scene, depending on the nature of the physical property or type of sensor device[11]. There are many types of sensors that can be used to control virtual objects, virtual cameras and augmented objects by an LAE in an MAR environment. These sensors can generate different results depending on their properties, such as position, direction, geographic coordinate system, time, motion and so on. Especially, the output of sensors can be filtered and regenerated as high-quality data.

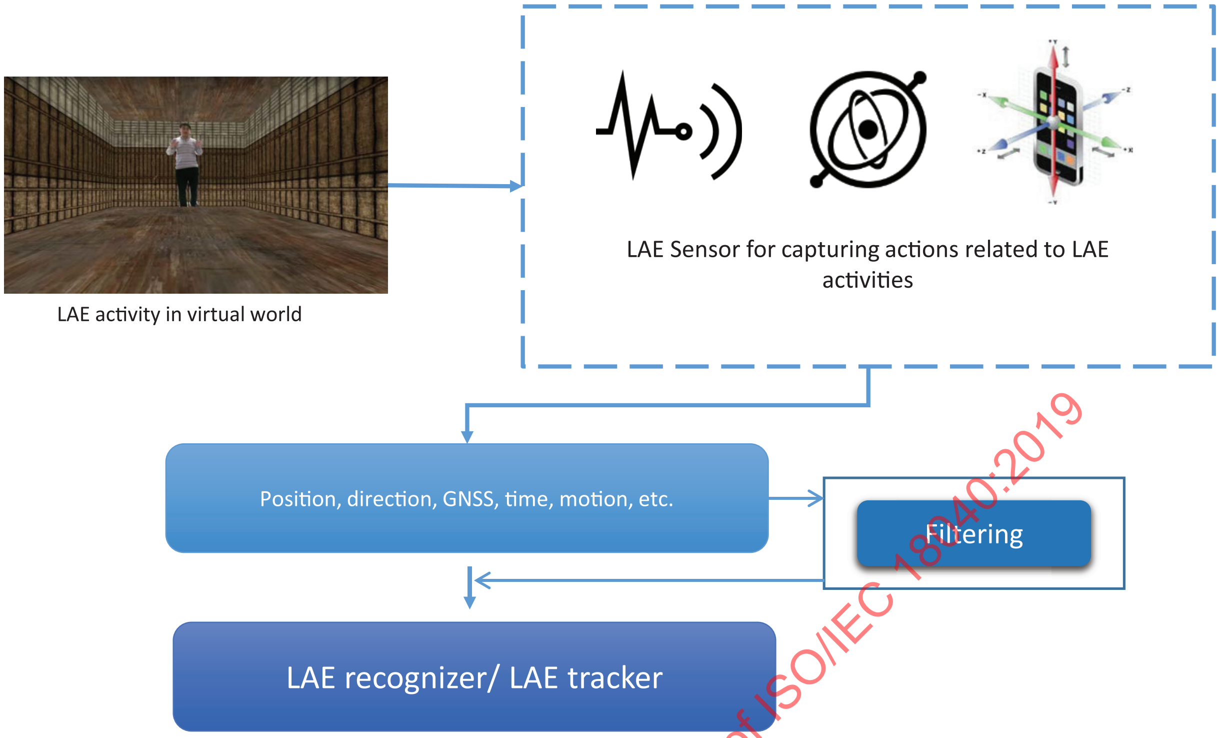


Figure 5 — Sensor capturing actions related to LAE activities

As an example, Figure 6 shows how a smartphone can be used to generate a gyroscope sensor, accelerometer, digital compass, ambient light sensor, proximity sensor, barometer, Hall Effect sensor, magnetometer, pedometer and so on. A gyroscope sensor is used to generate the angular velocity of the rotational angle per unit of time to detect the rotation of the phone. An accelerometer sensor is used to measure acceleration forces and dynamics to sense movement of the phone or vibration. A digital compass sensor is used to detect the geographic coordinate system, direction and navigation information of the phone.

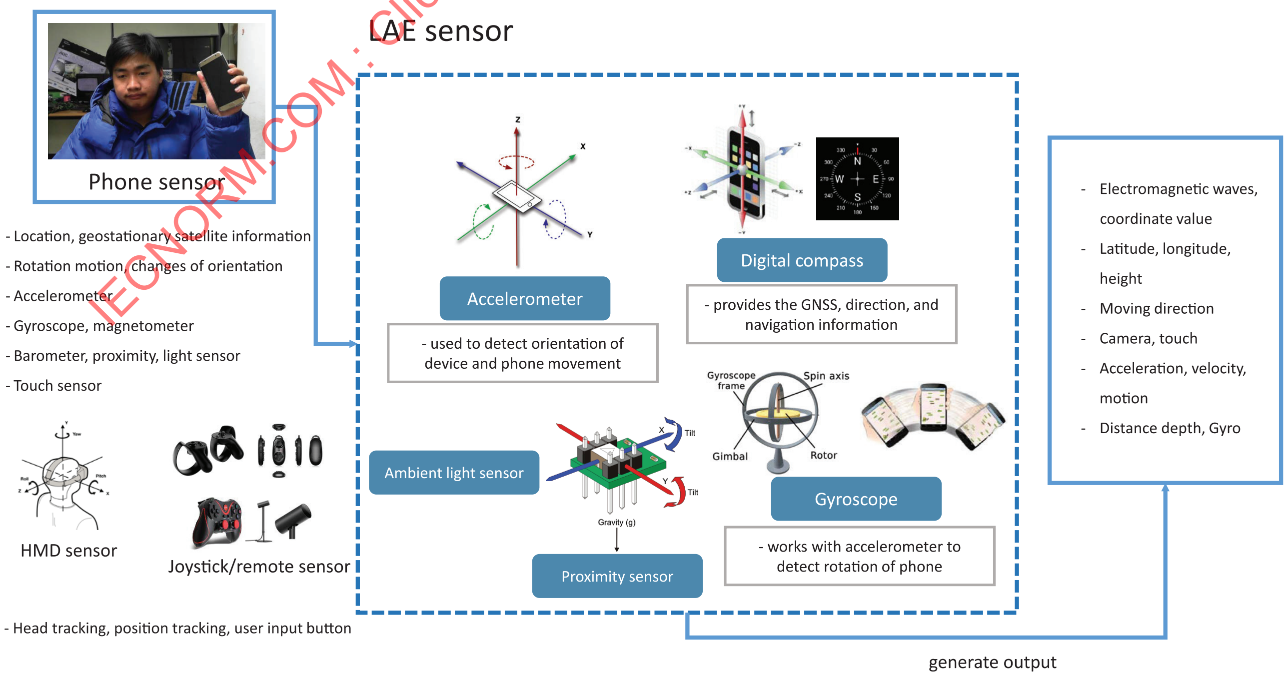


Figure 6 — Examples of sensors for LAE activities

An HMD sensor is also important for an LAE to be represented naturally in the virtual world. While an LAE is wearing an HMD device, they can see the real-like scenes of the virtual world and interact with virtual objects by using gesture front of depth or leap motion devices. There are three types of HMD devices, each of them providing different sensor information for an LAE:

— First, a PC HMD is a desktop peripheral that acts as an external monitor. It provides the deepest and most immersive VR experience. It can track position and orientation smoothly, allowing for easy calculation of the user's position, then generate real time position movement in virtual space. Most PC HMDs come with input devices, such as a camera for tracking position or a joystick for interacting with objects in the virtual world.

— Second, a mobile HMD is a device that can be connected via smartphone in order to visualize the virtual reality and can be customized using a built-in SDK. Mobile HMDs can provide orientation and head tracking sensing data. This type is supported strictly for smartphones.

— Third, there are other, lower-cost devices that can view the stereo scenes of virtual reality, such as a drop-in phone viewer. Drop-in phone viewers are supported by many smartphones using simple stereo rendering and accelerometer tracking. They can only track orientation and, therefore, do not provide an as smooth or immersive virtual reality experience as the first two.

Head tracking allows an LAE to feel present in a 360° scene due to sensors, such as gyroscope, accelerometer and magnetometer.

5.3 Informational view

Sensors receive various types of signals in the physical world as inputs and obtain sensing data related to the representation of an LAE in an MAR scene.

The input and output of sensors related to an LAE in an MAR scene are:

— input: physical world signals and/or device signals;

— output: sensor data related to representation of the LAE.

Physical sensor devices have a set of capabilities and parameters. Based on their properties, sensor devices generate wide-ranging types of sensing data, including camera intrinsic parameters (for example focal length, field of view (FOV), gain, frequency range);camera extrinsic parameters (for example position and orientation); resolution; sampling rate; skeleton; mono, stereo, or spatial audio; and 2D, 3D (colour and depth), or multi-view video. To provide the sensing data in a universally-consistent way, sensor output consists of <Identifier>, <Type>, and <Attributes>.

In order to composite a 3D virtual space and an LAE, consideration needs to be paid to the LAE sensing devices themselves. This document can use the following types of sensing devices: general cameras such as Web cameras and depth cameras such as depth-like devices. For general cameras, the sequence of RGB images is captured and the image itself is used as the sensing data. Depth cameras provide the following: an RGB color image, a 3D depth image, multi-array microphones and a motorized tilt. The captured image and/or the depth image can be used to embed the LAE into the 3D virtual space. Table 2 summarizes the inputs and outputs of the LAE capturer and LAE sensor.

Table 2 — Input/output of LAE capturer and LAE sensor

	Capture type/Sensor type	Input	Output	Devices to be used
LAE capture	Video	Physical world including LAEs	Image/video (gray, color, depth) camera information	RGB camera Depth camera
	360° video		360° image/video camera information	360° camera
	Stereo video		Stereo image/video camera information	Stereo camera
	Skeleton		Skeleton camera information	Depth camera
	Audio	Physical world	Audio signal (spatial or stereo) audio information	Microphone
LAE sensor	Geographic coordinate system sensor	Location, geostationary satellite information	Electromagnetic waves, coordinate value, latitude, longitude, height	
	Gyro sensor	Rotation motion, changes of orientation	Electromagnetic waves, coordinate value, moving directions	
	Smartphone	Accelerometer, gyroscope, magnetometer, barometer, proximity, light sensor, touch sensor	Electromagnetic waves, coordinate value, camera, touch, acceleration, accelerometer, motion	
	Wii remote™[a]/joystick	User input button function	Electromagnetic waves, coordinate value, distance depth, force	
	HMD	Head tracking, position tracking, user input button	Electromagnetic waves, coordinate value, motion Gyro, accelerometer	

[a] Wii remote is an example of a suitable product available commercially. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO/IEC of this product.

6 Tracker and spatial mapper for an LAE

6.1 Overview

Figure 7 shows that an LAE, which is represented in the physical world, can be captured, and sensing data extracted, by an LAE capturer and LAE sensor. The sensing data that relates to the LAE from the LAE capturer and/or LAE sensor needs to be pre-processed in order to provide optimized sensing data for doing spatial mapping of the LAE in the MAR scene. LAE capturer refers to physical world capturing devices, such as depth cameras, general cameras and 360° cameras. The video, image, skeleton and audio information are parsed to a pre-processing function to compose the filtering, colour conversion, background removal, chroma-keying and depth information extraction by using the technique of computer vision. The results of pre-processing are transmitted to an LAE tracker. Furthermore, the LAE sensor for capturing the event of the LAE's motion, such as position, direction, geographic coordinate system sensing, time and motion, provides that sensing data to a filtering function. The filtering function is used for processing the sensing data and recognizing the target object. The raw signal for representing an LAE and target object specification data is parsed to the LAE tracker. The LAE tracker can track a variety of information related to an LAE, such as real camera information and the LAE itself in the physical world[12].

Tracking data from the LAE tracker, such as camera information (FOV, position, orientation and so on) and LAE information (chroma-keying results, skeleton, geographic coordinate system and so on.) in the physical world, are parsed to a spatial mapper for mapping the LAE's spatial data. The role of the spatial mapper for an LAE is to provide spatial relationship information (position, orientation, scale and unit) between the physical world and the world of the MAR scene by applying the necessary transformations for the calibration of the LAE. The tracking data is mapped with a spatial scene description to embed the LAE into the spatial scene, then the calibrated spatial information is parsed to a scene compositing module.

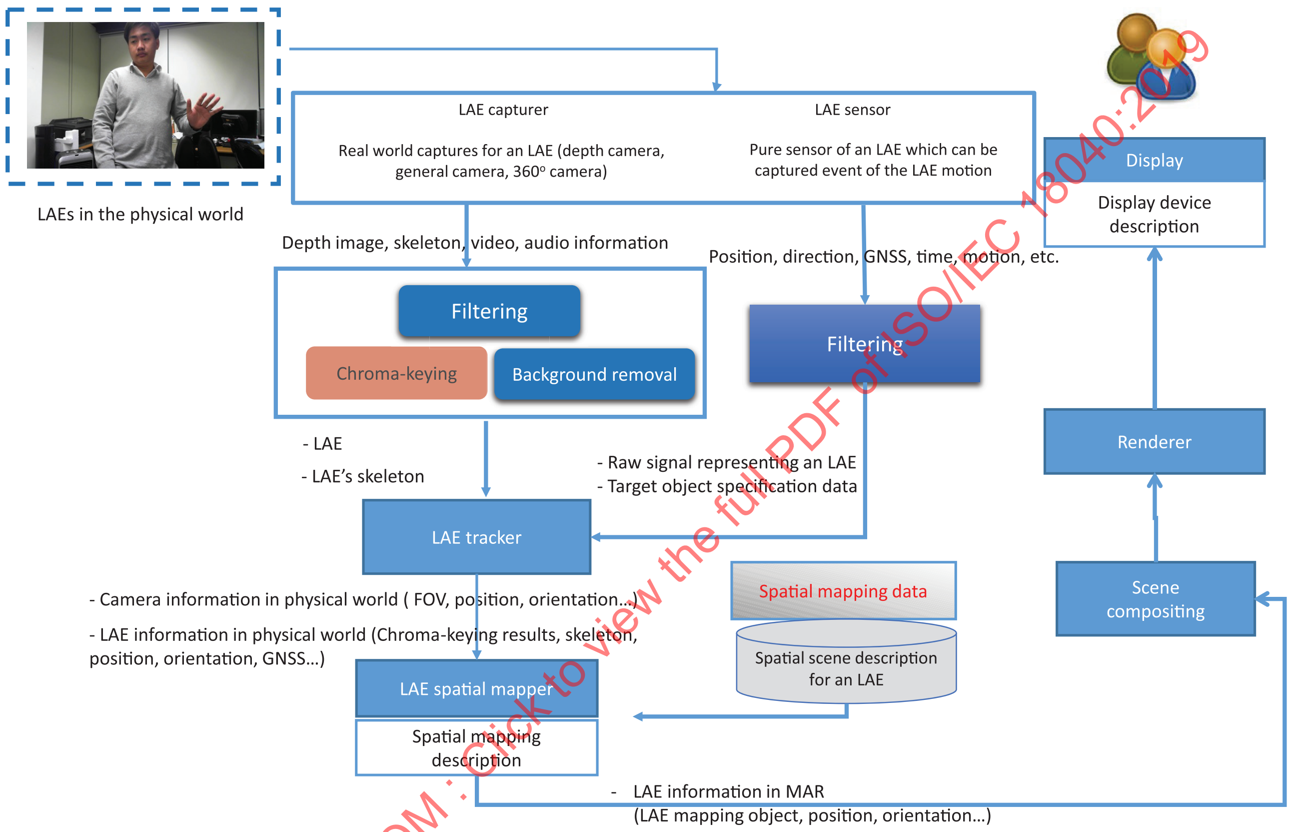


Figure 7 — Tracker and spatial mapper of an LAE in MAR

6.2 Computational view

The work of the spatial mapper components can be done using information stored in the spatial scene for an LAE. Sensing data includes a coordinate system which refers to an LAE sensor that extracts sensing data from the physical coordinate system, the physical camera coordinate system and the world coordinate system.

Figure 8 illustrates the mapping between the physical coordinate system and the world coordinate system, and details are provided in the following:

— Physical coordinate system refers to a coordinate system that enables locating an LAE and which is controlled by a geographic coordinate system sensing device. This means that the coordinate system can be extracted from the physical world through an implementation mechanism. The physical coordinate system retrieves ground coordinates, such as latitude/longitude or UTM coordinates, and can also measure ground distances and areas. The physical coordinate system can be applied to any kind of object or LAE that can be related to a geographical location.

— Physical camera coordinate system refers to an LAE capturer, which is the camera for capturing an LAE in the physical world. The physical camera also has its own coordinates, which can provide coordinate data to the spatial mapper.

— Object LAE refers to the position, orientation, FOV and spatial information of an LAE. The position is the location relative to the reference point of the object/LAE in the origin of the coordinate system. Orientation is the object's state of being oriented with a vector quantity and rotation of the LAE.

— Spatial mapping is used for combining and mapping the object LAE in physical world coordinate with virtual world coordinate. It provides tracking information from the LAE capturer and LAE sensor in order to map LAE into virtual world more naturally.

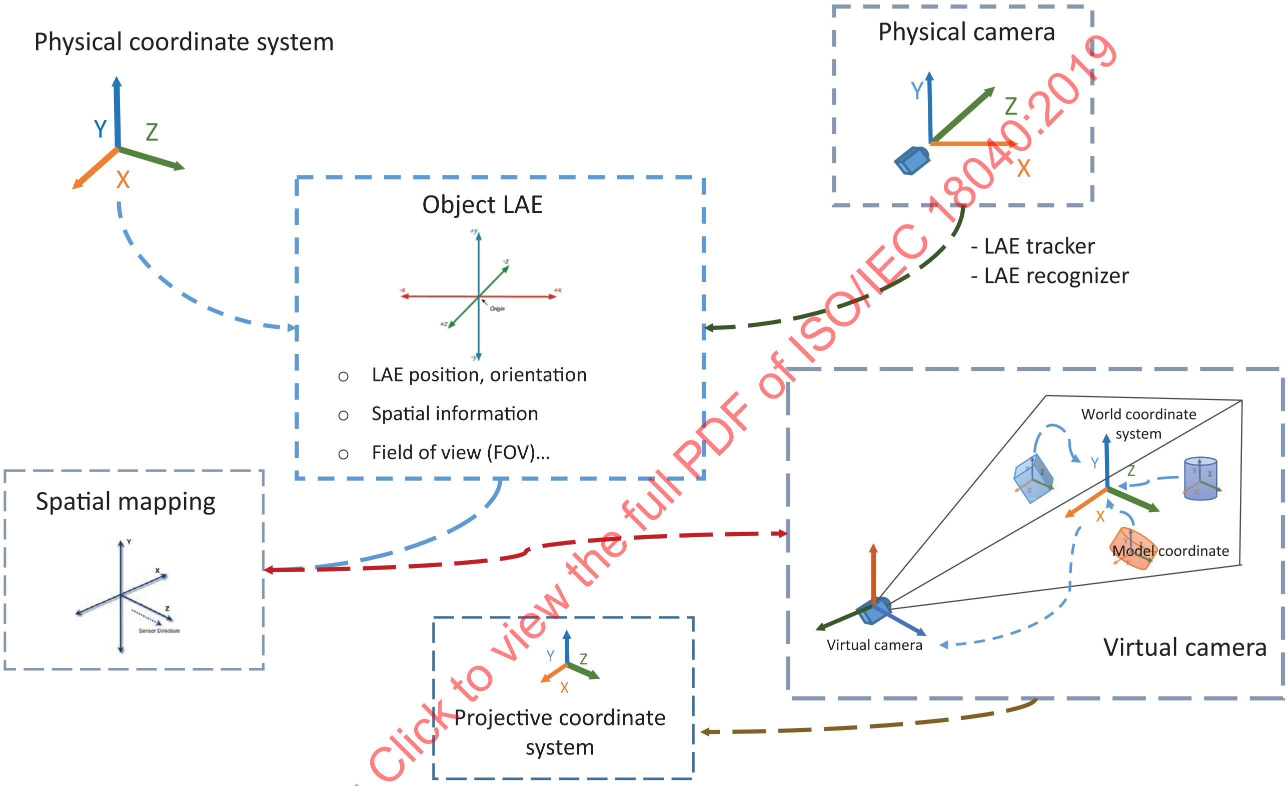


Figure 8 — Mapping between physical and world coordinate systems

The model coordinate system is the coordinate system where the entity's object and LAE model are initialized and created. It is a unique coordinate space of the model. Two distinct models, each with their own coordinate system, cannot interact with each other. Thus, there needs to be a universal coordinate system that allows any model to interact with any other. That universal system is called the world coordinate system. When interaction occurs, the coordinate system of each model and entities are transformed into a world coordinate system. The world coordinate system is then transformed into a coordinate system called the virtual camera coordinate system.

The virtual camera coordinate system is what users see on the screen and is related to a viewer which acts as a camera. A change in the camera's orientation and position changes what the viewer sees. The final transformation that converts three-dimensional scenery into a two-dimensional image is called the projective coordinate system or display coordinate system. What is perceived on screen as three dimensional is just an illusion. It is simply a two-dimensional image making use of the projective coordinate system.

When the image taken by the physical camera is rendered according to the position of the virtual camera, any image distortion can be made. The MAR scene rendering involving an LAE should be done with this in mind.

6.3 Informational view

The LAE tracker is able to detect and measure changes in the properties of sensing data related to the representation of an LAE in an MAR scene. Tracking information, such as position, orientation, location, geographic coordinate system and so on, from the sensor is tracked in the physical world.

— The input and output of the tracker, as shown in Table 3, are:

  — input: sensing data related to the representation of the LAE;

  — output: instantaneous values of the characteristics of the recognized target signals of the LAE.

Table 3 — Tracker categories

Dimension	Type	Category
Input	LAE capturer	— Camera information — 2D image/video of LAE — Object specification data — 3D primitives (points, lines, polygons, shapes) — 3D mode
	LAE sensor	— Sensing information
Output	LAE capturer	— Video for LAE — Position, orientation, volume, location — Haptic (force, direction and so on) — Aural (intensity, pitch and so on)
	LAE sensor	— Sensor information (coordinate value, latitude and so on)

The spatial reference frames and spatial metrics used in a given sensor need to be mapped into that of the MAR scene so that the sensed LAE can be correctly placed, oriented and sized. The spatial representation relationship between a particular sensor system and an MAR space is provided by the MAR experience creator and is maintained by the spatial mapper.

— The input and output of the spatial mapper, as shown in Table 4, are:

  — input: sensor identifier (SID) and sensed spatial information of the LAE;

  — output: calibrated spatial information of the LAE for the given MAR scene.

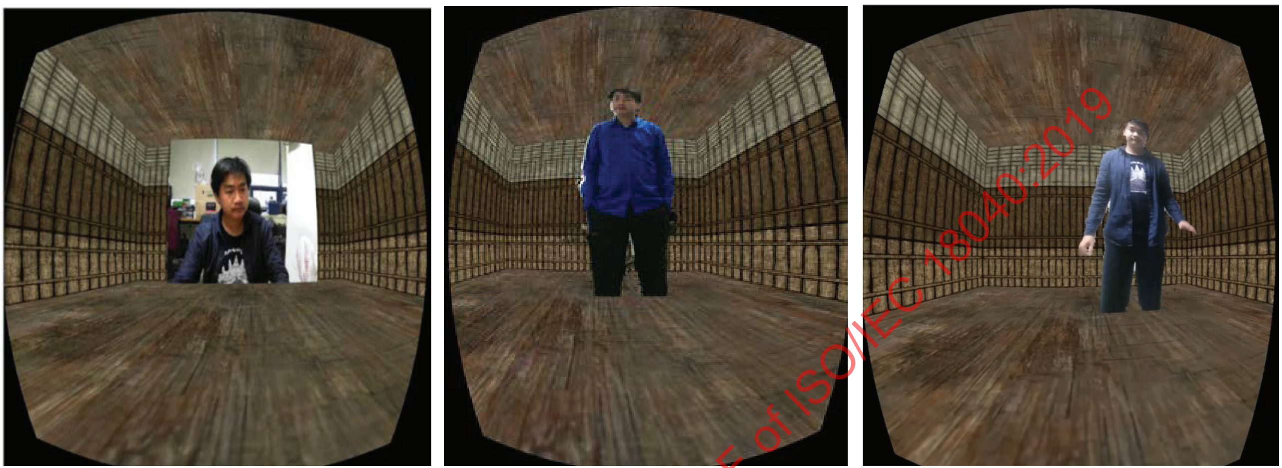
Table 4 — Spatial mapper input/output

Dimension	Type
Input	— LAE identifier spatial data — LAE consecutive spatial data — Tracking spatial data — Spatial information
Output	— Calibrated spatial information — Audio (direction, amplitude, units, scale)

The notion of a spatial mapper can be extended to mapping other domains, such as audio (for example direction, amplitude, units, scale) and haptics (for example direction, magnitudes, units and scale).

6.4 An example of LAE tracking and spatial mapping in MAR

Figure 9 a) shows an LAE embedded into an MAR scene without a chroma-keying function to remove the background of the LAE in the physical scene. Figure 9 b) shows that, after applying the chroma-keying filter on the LAE video stream, the LAE can be embedded in the MAR scene without its background. Figure 9 c) shows an LAE moment in time in the MAR scene. The LAE can be tracked from the physical world by the LAE capturer; and the position, orientation, FOV, coordinates and other information of an LAE can be mapped into the spatial representation. Spatial mapping is used to map the spatial data of an LAE and the scene description which contains the 3D virtual world.



(a) An LAE embedded into an MAR scene without a chroma-keying function to remove the background

(b) An LAE embedded in the MAR scene without its background

(c) An LAE movement in time in the MAR scene

Figure 9 — Examples of tracking and spatial mapping of an LAE in MAR

7 Recognizer and event mapper for an LAE

7.1 Overview

After obtaining sensing data from the LAE sensor, the data go through to the LAE recognizer for recognizing events of LAE gestures, object collision and so on.

The recognizer is a component that analyzes sensing data related to the representation of an LAE in an MAR scene and produces MAR events and data activated from the LAE through comparison with a local or remote target signal stored in the MAR scene. That is the target for augmentation.

7.2 Recognizer

Recognition can only be based on previously captured target signals. Both the recognizer and the tracker can be configured with a set of target signals provided by or stored in an external resource (for example a third-party database server) in a manner consistent with the scene definition. The recognizer can be independently implemented and utilized. There are two types of data used by the recognizer: the output from the LAE capturer and the output from the LAE sensor.

Figure 10 shows an event of gesture of an LAE in the physical world captured by the LAE sensor and the LAE capturer. After filtering the LAE's raw data, the filtered data (position, orientation, direction, motion and so on) are parsed to the LAE recognizer for recognizing the specific target event. The LAE recognizer is processed according to the event mapper and event description for the LAE. The event mapping data is stored in the event description for the LAE and can be used by the LAE recognizer according to the specific target and Event_ID. The LAE's event information, such as Event_ID, Event_Type and event function, is parsed to the event mapper for mapping the event. By analyzing the LAE information, the event mapper produces the MAR event in the MAR scene. The MAR event results from the detection of a relevant condition from the LAE in the physical world and augmentation. The event result is parsed to the scene compositing module and renderer in order to render the event function for display and UI.

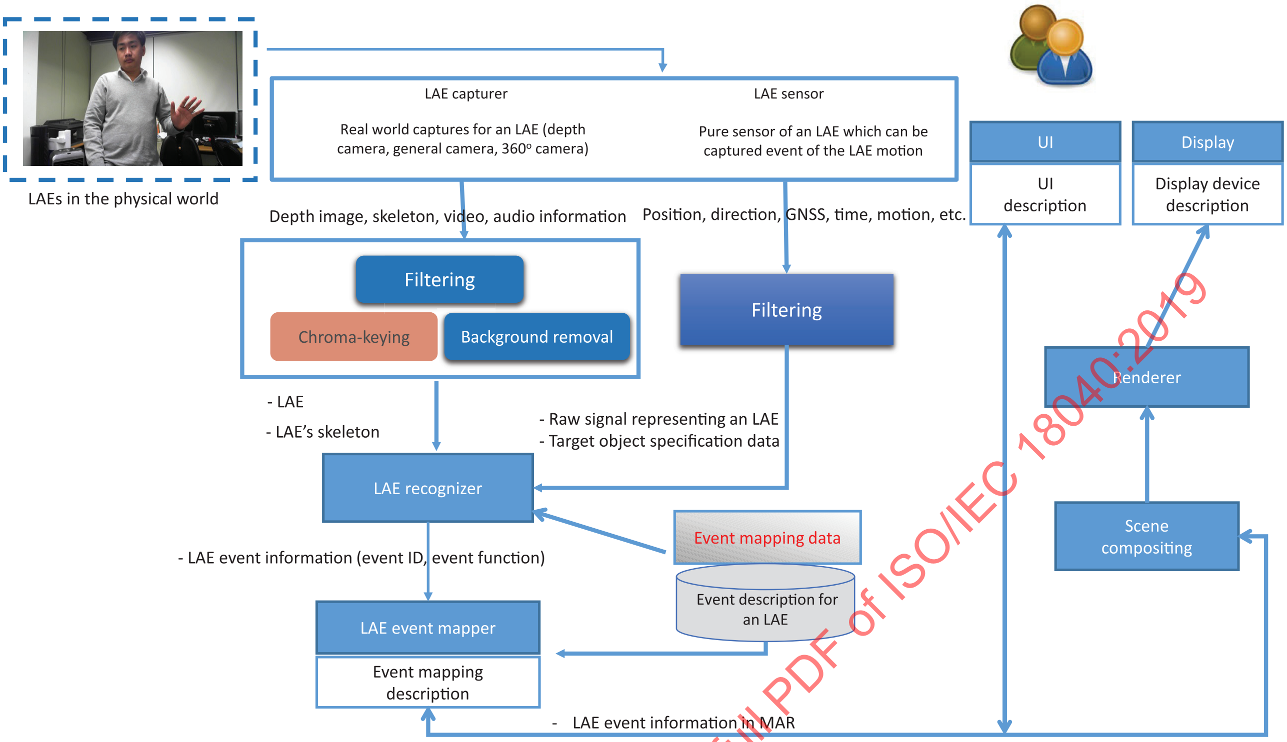


Figure 10 — Recognizer and event mapper of an LAE in MAR

— The inputs and outputs of the recognizer, as shown in Table 5, are:

— input: sensing data related to the representation of the LAE in the MAR scene. The input data model of the recognizer is the output of the sensors. The other input to the recognizer, the LAE raw data, should contain the following elements: first, it should have an identifier indicating the event when the presence of the LAE is recognized. Second, the LAE specification may include raw template files used for the recognition and matching process, such as image files, 3D model files, sound files and so on. Third, it may include a set of feature profiles. The types of features depend on the algorithms used by the recognizer. For instance, it can be a set of visual feature descriptors, 3D geometric features and so on;

— output: at least one event acknowledging the recognition, which identifies the recognized target and optionally provides additional information that should follow a standard protocol, language and naming convention.

Table 5 — Recognizer categories

Dimension	Type	Category
Input	Physical world	— Camera information — 2D Chroma-keying image — Audio — Gesture
	Sensor	— Sensor information
Output	Physical world	— Event indication only of the recognized LAE — Event by gesture and audio information
	Sensor	— Event indication only of the recognized LAE

7.3 Event mapper

The event mapper creates the relationship to the MAR event of an LAE that is obtained from the LAE recognizer and/or LAE tracker. It describes the MAR event that can occur due to the controlling event (gesture, movement, head tracking, sensor and so on) and the condition specified in the MAR scene. In order to map the MAR event within the MAR scene, as well as the events identified and recognized by the LAE recognizer, Event_ID, Event_Type and eventDB (event database) are needed. The event refers to the sequence of the controlling event of a gesture by the LAE. The gesturing data are recognized by the LAE recognizer to determine the type of event, and then the event function is retrieved from the eventDB.

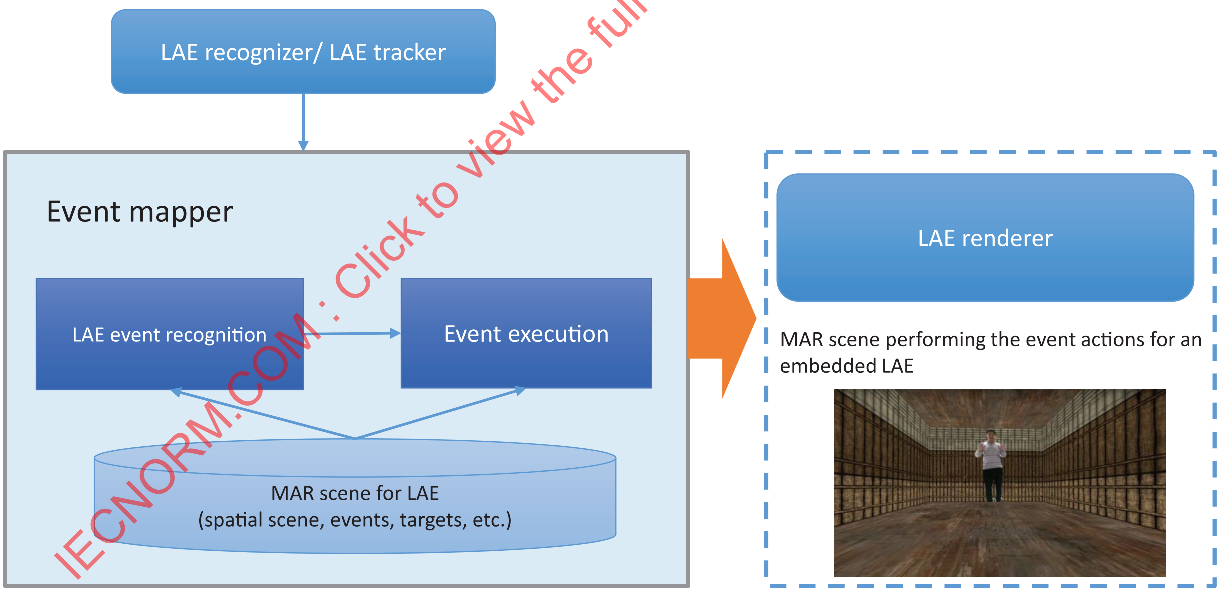


Figure 11 — Procedure for controlling an LAE event in an MAR scene

After retrieving the sensing data of the LAE, the event control model proposed in Figure 11 can be concretely developed with two modules, a recognition module and an execution module. Especially, the objective is to obtain faster and more accurate sensing data from the LAE itself and its handling devices in real time. If a hand gesture is used to control the events of an LAE, a depth camera, which allows skeleton images of one or two LAEs to be tracked, can be employed as a motion-sensing input device. Position information for the left and right hands can be obtained via depth. A position in 3D real space is represented by three Cartesian coordinates (x, y, depth) at each joint of the skeleton. An acquired position is first filtered to suppress image noise and then reprocessed with depth calibration and depth cueing to reduce image flickering. Depth supports some filtering features, but the resulting object boundaries remain very unstable and there is flickering. After all filtering processes have been carried out, hand gestures are recognized by the recognition module, and the recognized event is executed by the execution module in a virtual space, according to the gestures.

The event relationship between a particular recognition system and a target scene is provided by the MAR experience creator and is maintained by the event mapper.

— The input and output of the event mapper, as shown in Table 6, are:

  — input: event identifier and event information;

  — output: translated event identifier for the given MAR scene.

Table 6 — Event mapper input/output

Dimension	Types
Input	— Event information — Event identifier (virtual camera control, virtual object control, AR content control)
Output	— Translated event identifier for the given MAR scene — Interaction of a virtual object event

7.4 Event execution

A detected event is transferred in the event class *EventClass*, in an execution module. If the detected event is one of the events defined in the MAR system, the event corresponding to the recognized event is applied to the object in the 3D virtual space by the event executor *EventExecutor*. In other words, the object in the 3D virtual space is manipulated according to the corresponding event function.

Figure 12 shows the structure for executing an event to control an object in a 3D virtual space. The class *EventExecutor* employs registered object-control functions to control objects in the 3D virtual space by analyzing events in *EventClass*. If a virtual object can be controlled by an event, the class *EventExecutor* is set as a member of the object instance, a callback function is created to be executed on the gesture event for the object, and the generated function is linked to a corresponding gesture delegate function in *EventExecutor* based on the event type. Then, when an event occurs, the MAR system executes the function *executeCommand*, to which the event is passed as an *Event* instance and executes the generated call back function via the corresponding delegate function in *EventExecutor*. As a result, an LAE can interact with 3D virtual objects in an MAR environment just as it can with real-world objects.

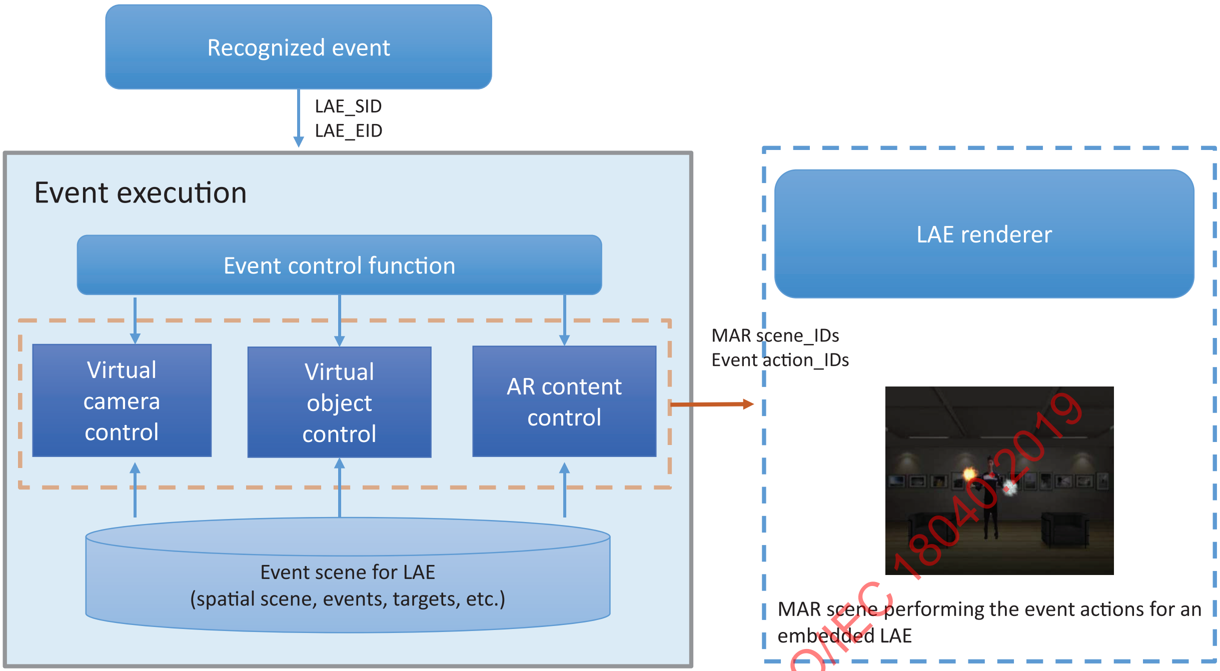
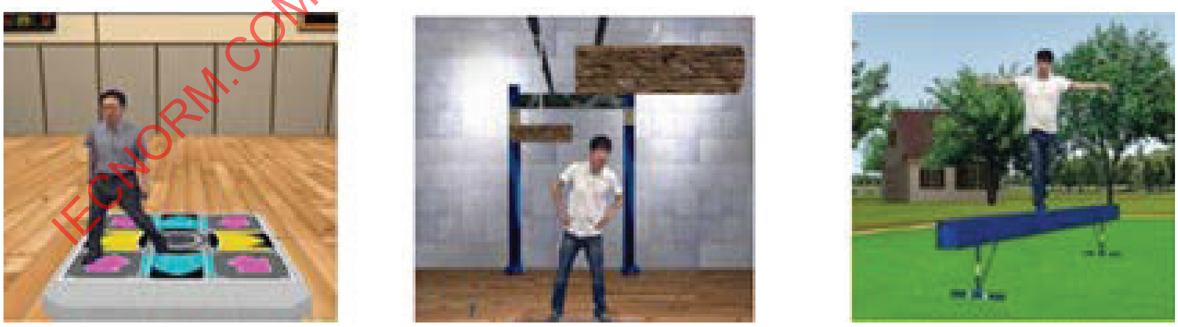


Figure 12 — Execution structure of events performed by an LAE

7.5 Examples of LAE recognizing and event mapping in MAR

Figure 13 shows the interaction of an LAE with virtual objects in an MAR scene[13]. Examples of interactions include dancing on a virtual dance dance revolution (DDR)[14] stage, dodging a piece of wood and walking on a balance beam. If the LAE follows prescribed conditions accurately, they win the game, otherwise they lose. In the first example [Figure 13 a)], the man moves (dances) in front of the camera, the motion is captured and the music plays according to the motion of his feet. In the second example [Figure 13 b)], the piece of wood is a virtual object in the 3D MAR scene and the man tries to dodge the object. If he is unsuccessful, a collision between his head and the virtual object occurs. In the third example [Figure 13 c)], the man is walking in the physical world, but it appears as if he was crossing a balance beam in the MAR scene. If he falls, an event occurs.



a) Virtual DDR dancing b) Dodging a piece of wood c) Crossing a balance beam

Figure 13 — Interaction of a virtual object in an MAR scene and Event_IDs of an LAE

Figure 14 shows events of an LAE in an MAR scene based on recognition of gestures and Event_IDs[15]. The man is embedded into the MAR scene by applying the spatial mapper and filtering with the chroma-keying function to remove the background. The man can use his hands to create and control the events in the MAR scene by doing hand gestures. The gestures are recognized by the LAE recognizer. The event IDs are retrieved from the eventDB and mapped with a module/function in the MAR scene by event mapping. An event occurs based on an Event_ID, such as lifting the right hand creates fire, or lifting two hands creates both fire and ice, or holding the hands together makes the fire disappear.

a) Lift a right hand   b) Lift two hands   c) Hold two hands

Figure 14 — Augmenting a virtual object based on event IDs of an LAE

8 Scene representation for an LAE

8.1 Overview

MAR scene refers to a scene that represents a virtual scene and placeholders. It serves as the implementation structure that combines the physical world scene/objects and the virtual scene/objects. It is the observable spatiotemporal organization of physical and virtual objects that have been tracked and recognized by the LAE tracker and LAE recognizer functions. The scene representation can be based on several proposed AR related formats so far[16][17].

An LAE event can occur based on the LAE's own actions, as well as based on output obtained from sensing interface devices handled by the LAE. The LAE can be captured from general cameras and/or depth cameras. Its actions can be obtained by recognizing LAE sensing data. Meaningful events to be used in the MAR scene can be classified into one of two types: sensing data by the LAE itself or sensing data of devices handled by the LAE. The defined events are stored in the eventDB.

— Sensing outputs captured from a "capturer" for the LAE

  — Gestures — hands, fingers, head, body

  — Facial expressions

  — Speech and voice

— Sensing outputs captured from a "sensor" for the LAE

  — AR marker

  — Global navigation satellite system (GNSS)

  — Wii remote™ motion data

  — Other sensing data by a smart device: three-axis accelerometer, magnetometer (digital compass)

Sensing data are used to define the events of the LAE. For example, a virtual camera, a virtual object or an AR object in a 3D virtual space is controlled by human body gestures of the LAE. To define the gestures more efficiently, basic primitive postures of a human body are defined, as shown in Figure 15. The postures consist of position, vector, image and skeleton. By detecting the postures of a human body and combining them, gestures can be defined. Three common primitive gestures of a human body can be defined as follows: linear, wave, clockwise/counterclockwise. A variety of events for a human body can be defined using these predefined primitive gestures.

An MAR scene describes all the information of an LAE that relates to an MAR scene and consists of a virtual scene, sensing data, spatial scene, events, targets and so on. The MAR scene representation is the middleware of the spatial mapper and event mapper, which observes the spatial of physical objects, virtual objects and events. An MAR scene has at least one physical object and one virtual object.

Figure 15 shows the scene representation for handling an LAE in an MAR scene. After receiving sensing information from the sensor, the raw signal representing the LAE and target object's specification data is parsed to the LAE tracker and LAE recognizer. Thus, the LAE and physical world objects can be captured and detected. The LAE's information can be calculated and mapped in the virtual scene according to the LAE's position, orientation and so on. Calibration and spatial mapping between real and virtual objects are needed in an MAR scene. A physical world object is mapped to a placeholder in the MAR scene by, for example, image feature, marker pattern definition, or geographic coordinate system, and provides the information data/method to recognize and track the object for semantic information as to what the object actually is. The video of the entire scene, including the background, is processed in a background removal function to extract the LAE and its information. The LAE is embedded in the virtual scene via real time streaming, and the virtual objects of the virtual scene can be interacted with by the LAE's gesture/event according to the event of the LAE recognizer. When there are changes to a virtual object, MAR behaviour invokes the MAR event and the LAE recognizer recognizes then produces the event.

Sensor/device parameters are important for correctly presenting an MAR scene, for example the relationship of a physical world capture sensor parameter to the MAR scene viewpoint and image scale. The viewpoint and image scale shall be changed according to the information captured by the physical world capture sensor.

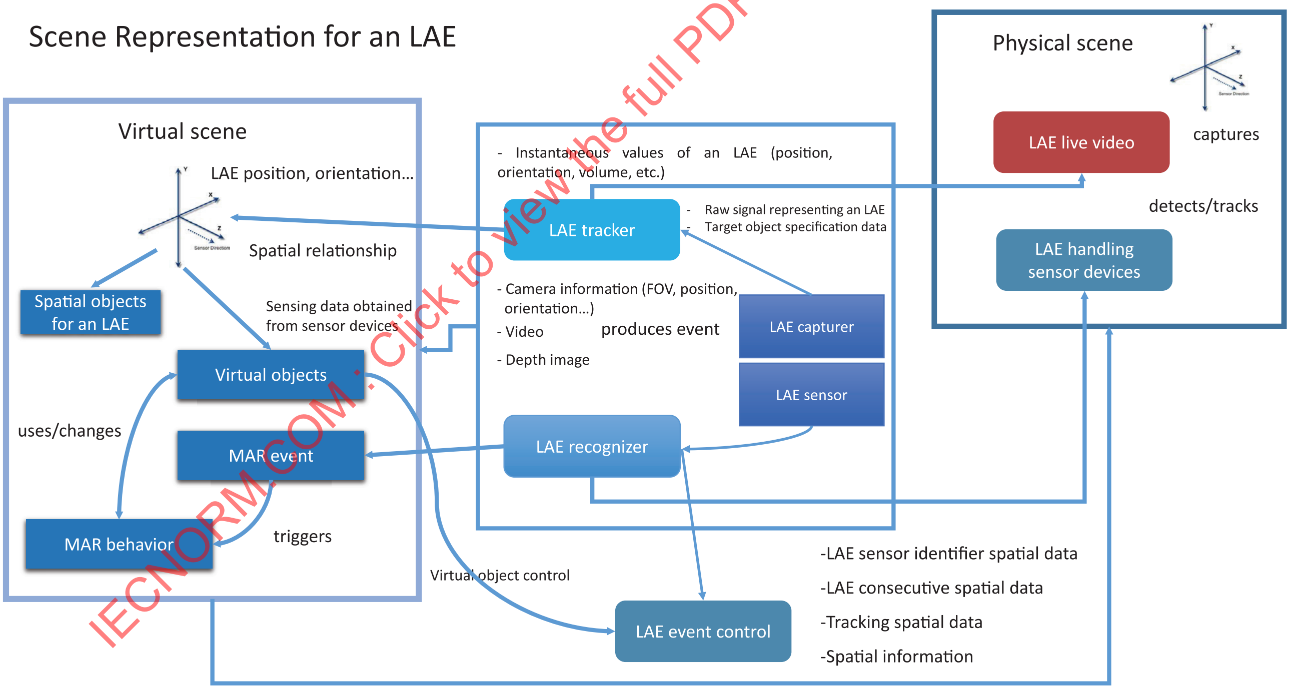


Figure 15 — Scene representation for handling an LAE in an MAR

8.2 Scene description

MAR scene description is used to perform more effective spatial mapping of and event handling of an LAE in an MAR scene. Two types of scene description are provided.

First, spatial mapping is used for mapping spatial representation relationship information between a physical world and an MAR scene. It describes how an LAE in the physical world can be mapped into an MAR system. LAE information provided by the LAE tracker and/or LAE recognizer sets an LAE_ID and initializes the spatial mapping information. The spatial reference frames and spatial metrics used in a given sensor need to be mapped into the MAR scene so that the LAE mapping object can be correctly placed, oriented and scaled. As shown in Table 7, the spatial mapping information can be modelled as a table, with each entry characterizing the transformation process from one aspect (for example lateral unit, axis direction, scale and so on) of the spatial property of the sensor to the given MAR scene[18].

Table 7 — Scene description for spatial mapping of an LAE

LAE ID	Initial spatial mapping information	LAE mapping object	Spatial mapping function
LAE_ID	Position, orientation and so on	Virtual 2D/3D object	Spatial mapping from physical world to MAR scene

Second, event mapping describes how an LAE in an MAR scene can produce an event, such as interaction, gesture, movement, voice recognition and so on. It creates an association between an MAR event obtained by the LAE recognizer and/or LAE tracker and a condition specified in the eventDB. For example, an LAE in an MAR scene can use their hands to control/interact with virtual objects by doing hand gestures (palms up/down, hands clutched, hands raised and so on). An event occurs when the LAE does the gesture or movement or handles a sensing device (for example smartphone, joystick and so on). The sensor device handled by the LAE can generate the sensing data for creating an event in the event mapper. Event mapping matches the Sensor ID and Event_ID, which are retrieved from the eventDB as shown in Table 8. Then, it can know the type of event (Event_Type) and call a function to be executed.

Table 8 — Scene description for event mapping of an LAE

LAE ID	Event ID	Event type	Event function
LAE_ID	Event_ID	Event_Type	Event call functions

9 Renderer

9.1 Overview

Rendering refers to the process of generating data from a 2D/3D model, updating a simulation and rendering the presentation output for a given display device. After the processes of the spatial mapper and event mapper have completed, the data are parsed to the renderer. Thus, the renderer has the job of rendering LAEs, virtual objects and events to a display device.

Rendering of an LAE in an MAR system needs to be done based on the type of display, such as Web rendering to a browser, or PC, mobile, and stereo rendering. Stereo rendering has to do with the stereoscopic separation of the left and right eye. Rendering needs to be smooth and in real time, based on the user's HMD.

There are two issues when it comes to rendering. The first has to do with spatial mapping rendering, which refers to an LAE being embedded into an MAR scene. Based on the embedding process, the renderer needs to generate a high-quality and smooth display. The second has to do with event mapping, which refers to the events of an LAE, such as natural movement rendering, gesture, voice recognition and so on. The renderer needs to render the event perfectly in order to produce a quality, streamlined event. In other words, when an LAE moves or does gestures, this should be reflected in the MAR scene in a streamlined fashion. Furthermore, the renderer needs to control problems that can occur during LAE movement, such as moving to an unacceptable position and orientation, or gestures and movement being too slow, which can result in low quality rendering.

9.2 Computational view

An MAR system can specify various capabilities of the renderer, so a scene can be adapted, and simulation performance can be optimized. The rendering of virtual reality, a stereoscopic HMD and a mobile device can require different rendering performance. Multimodal output rendering sometimes necessitates careful millisecond-level temporal synchronization. The output is a visual, aural and/or haptic stream of data such as a video frame, stereo sound signals and so on, to be fed into a display device.

9.3 Information view

- The input and output of the renderer, as shown in Table 9, are:
  - input: MAR scene graph data;
  - output: synchronized rendering output (for example visual frame, stereo sound signals, motor commands and so on).

The renderer can be categorized in the following way:

Table 9 — Renderer categories

Dimension	Type			
Modality	Visual	Aural	Haptics	Other

10 Display and UI

Display refers to a hardware component that produces the actual presentation of an MAR system for representing an LAE to the end user. Displays include monitors, HMDs, projectors, scent diffusers, haptic devices, sound speakers and so on. The display needs to meet specific requirements in order to provide a good quality display to the end user. An actuator is a special type of display that does not directly stimulate the end user's senses but rather produces a physical effect in order to change properties of physical objects or the environment.

A UI is a component of hardware used to capture user interactions (for example touch and click) for the purpose of modifying the state of an MAR scene. A UI requires sensors to achieve this purpose. There are many types of UI sensors for capturing user interactions, such as gesture, voice, haptic, Gyro and so on. These sensors have a usage similar to that of LAE sensors. However, the difference is that, for UI sensors, the only physical object sensed is the user.

There are two kinds of events that can be generated:

- An LAE can generate events by itself — the LAE uses gestures, voice, head movements and so on.
- An LAE can handle a device for generating events — the LAE can use a smartphone, haptic, joystick and so on.
  - The input and output of the display are:
    - input: rendered signals;
    - output: display output.
  - The input and output of the UI are:
    - input: user action, gesture, voice, interaction;
    - output: UI event.

11 Extensions to virtual actor and entity

A virtual actor and entity (VAE) refers to a virtual representation of an LAE which is captured by 3D capturing technology and can be reconstructed in an MAR scene. The VAE appears as a real person that can be communicated with, though it is not. Figure 16 shows a VAE restricted to and communicating in an MAR scene. The virtual person is reconstructed and displayed in the MAR scene as a virtual object. Recently, a new type of 3D capture technology has appeared, which allows a high-quality 3D model of a person to be reconstructed, compressed and transmitted anywhere in the world in real time[3]. When combined with a mixed reality display, this technology allows a user to see, hear and interact with a remote participant in 3D as if they were actually present in the same physical space. Communicating and interacting with a remote user becomes as natural as face-to-face communication.

Figure 16 — A virtual actor and entity restricted to and communicating in an MAR scene[7]

An HMD device is required to view the hologram in real time, and a room surrounded with 3D cameras is necessary to create the hologram. The VAE is not viewable without wearing an HMD device, and it is not possible to make eye contact with someone who is wearing the device.

A second type of VAE refers to a virtual character (avatar) that is represented in an MAR scene, as shown in Figure 17. This kind of VAE can be used instead of an LAE in a physical scene. However, the interaction and motion of the virtual character shall be accomplished by LAE gestures in the physical scene. The avatar can be represented along with other avatars, and the avatars can be made to communicate with each other. An avatar can be constructed with a 3D model, and points, like an LAE skeleton, need to be defined in order to produce animation and gestures. The virtual character and its movements shall be transmitted in real time using an LAE and movements in the physical scene, such as gestures of the hand, foot, head and mouth.

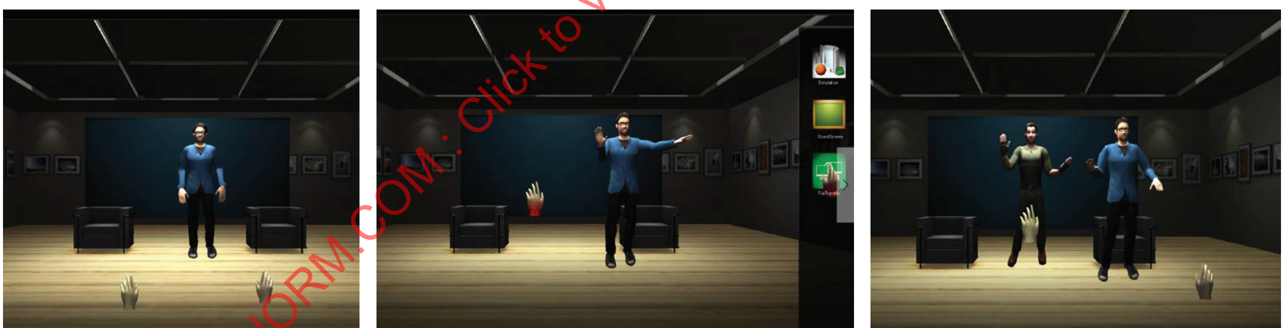


Figure 17 — Virtual character representation in an MAR scene[15]

12 System performance

System performance for an LAE-MAR application is an important part of the user's experience. The performance of an LAE-MAR system can be evaluated in measurable and technical terms. Performance is based on sensor quality, quality of LAE expression, processing speed, response time, latency, dirty chroma-keying results, frame rate and so on. Sensor refers to an LAE sensor, which is a sensor for capturing an LAE. The accurateness of the sensor affects the accuracy and quality of the LAE's information. The expression of the LAE can also be affected by system performance related to applying a chroma-keying function, background removal, natural movement and degree of freedom.

Latency can be measured as the total time needed to process the target LAE object and produce the augmentation by an LAE event. That is the time delay between the cause and the effect of a physical change in the system being observed according to an LAE event. It is a result of the limited velocity with which any physical interaction of an LAE can take place. For example, an LAE in physical world moves to the left, but the system update is slow and the LAE still shows on the right. In the LAE-MAR systems, types of latency can be defined as several viewpoints. The latency of LAE capturer is the delay of capturing time from physical world for transmitting into system. In this case, the latency of LAE tracker can be occurred by the time measure of data that is transmitted from one function to another. LAE's gesture, movement and activities in physical world can be tracked and recognized by LAE tracker and LAE recognizer. Thus, the measure time for tracking and recognizing an LAE should be considered. Latency testing and reducing latency are proposed to test the system by measuring latency with a stop watch. In some cases, hardware can cause latency. High-speed cameras and computer hardware specification can be used for reducing latency as well.

Augmentation precision in an LAE-MAR system can be measured as the error between event actions estimated by the LAE recognizer and the correct ones, or as the distance (in pixels) and angular distance (in degrees) between where the virtual object is displayed and where it should be displayed.

Operating conditions affecting performance in an LAE-MAR system may include the mobility of the target object, sensing distance, orientation and so on.

Response time in an LAE-MAR system, which is the total amount of time it takes to respond to a request for any functionality, such as an LAE detection, an LAE movement control, 3D display rendering and so on, is another consideration. There are three types of response time:

— The first, service time, refers to how long the LAE-MAR system takes to respond to a request.

— The second, waiting time, refers to how long an LAE-MAR movement and interaction request has to wait for another request queued ahead of it before it can run.

— The third, transmission time of LAE information, refers to how long it takes to process the request and respond back to the requestor.

Examples of LAE information requests include gestures for interacting with objects, an LAE's movement (updating position and movement in virtual space) and so on.

13 Safety

An LAE in an MAR scene can be presented with various safety issues, as it is a human being acting in the physical world. Thus, an LAE's attention needs to be focused. LAE safety guidelines are necessary to ensure that an LAE in an MAR scene includes components for safeguarding the LAE during system operation. These safety guidelines can be used to reduce the risk for an LAE by considering the following:

— dangerous obstructions that can lead to injury during the LAE's performance in the MAR scene;

— any information that needs to be encrypted for LAE security reasons;

— personal privacy of an LAE and potential exposure of personal information to unauthorized systems or third parties via a sensor/camera being out of scope, including authentication identity, system access to an LAE's personal data and so on;

— physical world situation during LAE movement in the MAR scene;

— virtual reality sickness from wearing an HMD;

— wearing an HMD device and being blind to potentially dangerous objects in the vicinity;

— avoiding quick acceleration or deceleration of camera movements, and using a constant velocity instead;

— keeping the frame rate up (less than 30 fps is uncomfortable);

- intermittent disconnection of the network service, leading to false confidence in the currently presented information;
- avoiding the use of depth of field or motion blur post processing because of not knowing where the eyes focus;
- avoiding sharp and/or unexpected camera rotations;
- avoiding brightness changes (use low frequency textures or fog effects to create smooth lighting transitions).

An LAE-MAR system should satisfy the safety guidelines for the user by applying the solutions of system configuration and safety guideline for the user in more detail. Here the user can be represented as an LAE model. The LAE-MAR system should consider the encryption and protection of LAE's data while users are using the system and transferring LAE's data to server. Furthermore, the developers should consider performance and sickness that can be caused by using the LAE-MAR system. For example, while the user is wearing an HMD device, the quick acceleration or deceleration of the camera can make the user feel uncomfortable and vomit. In some cases, while wearing an HMD, the user cannot see the environment outside. The safety guideline and system functionalities are useful for avoiding unexpected problems.

14 Conformance

Conformance for LAE representation in an LAE-MAR system is expressed around the aspects of how an LAE can be embedded into an MAR scene and the implementation process related to the LAE. Table 10 describes the conformance for the LAE representation and the components with their specification. Moreover, the conformance of LAE representation to this document shall satisfy at least the following requirements:

- The key architectural components that can be present for LAE representation in an LAE-MAR system are the following:
  - LAE capturer;
  - LAE sensor;
  - LAE tracker;
  - LAE spatial mapper;
  - LAE recognizer;
  - LAE event mapper;
  - scene representation;
  - renderer;
  - display and UI.
- The three types that can be controlled for LAE representation in an LAE-MAR system are the following:
  - virtual camera control in an MAR scene;
  - virtual object control in an MAR scene;
  - augmented object control.
- The implementation of LAE representation in an MAR scene shall conform to the concepts and architectural components for an LAE-MAR system as shown in Figure 3, Clause 4.

— The processing implementation of an LAE-MAR system can contain the chroma-keying, filtering, tracking, and recognizing functions of capturing and sensing LAE as specified in Clauses 5, 6, and 7.

— The movement of an LAE in an LAE-MAR system shall be mapped and moved naturally within an MAR scene if an LAE spatial mapper is provided in an LAE-MAR system as specified in Clause 6.

— The interaction between an LAE and an MAR scene shall be performed naturally within an MAR scene if an LAE event mapper is provided in an LAE-MAR system as specified in Clause 7.

— The interfaces between the architectural components of an LAE-MAR system shall contain and carry the information specified in this document. However, the specific content, format, data types, handshake, flow and other implementation details are at the discretion of the given LAE-MAR system to meet its specific needs.

— The API for an LAE-MAR system shall conform to the concepts specified in this document in order to ensure compatibility and software interface interoperability between LAE-MAR systems can be accomplished at least at the abstract API level.

Table 10 — Conformance for an LAE representation

Functions	Check	Items	Remarks
LAECapturer	[ ]	— provide information of a capturer	*There are different types of capturers such as general camera, depth camera, 3D camera and so on. *A suitable capturer in an LAE-MAR system can be used to capture an LAE according to their purpose.
	[ ]	— capture an LAE in the physical world	
	[ ]	— provide LAE information	
LAESensor	[ ]	— provide information of a sensor	*There are different types of sensors such as any camera, GNSS sensor, gesture sensor, voice sensor and so on. *A suitable sensor in an LAE-MAR system can be used to sense an action of an LAE according to their purpose.
	[ ]	— provide sensing information of an LAE	
LAETracker	[ ]	— track information of an LAE	*A suitable tracking module in an LAE-MAR system can be developed to track a movement of an LAE according to their purpose.
	[ ]	— Filtering function	
	[ ]	— Chroma-keying function	
	[ ]	— transfer the tracked result of an LAE to LAESpatialMapper	
LAERecognizer	[ ]	— provide a recognition function of an LAE event	*A suitable recognizing module in an LAE-MAR system can be developed to recognize an event of an LAE according to their purpose.
	[ ]	— provide event data base containing event ID and the corresponding event function	
	[ ]	— transfer the recognized result of an LAE event to LAEEventMapper	

Table 10 *(continued)*

Functions	Check	Items	Remarks
LAESpatialMapper	[ ]	— provide spatial information in a physical world and in MAR scene	*Since there are several types of MAR scene such as 2D virtual scene, 3D virtual scene, 360° VR scene and so on, an embedding and a movement of an LAE shall be represented in the different type of MAR scene.
	[ ]	— embed the tracked LAE into an MAR scene	
	[ ]	— transform the physical coordinate system of an LAE in a physical world into a virtual camera coordinate in an MAR scene	
	[ ]	— process a movement of an LAE in an MAR scene naturally	
LAEEventMapper	[ ]	— control event ID and the corresponding event function	*Since there are several types of MAR scene such as 2D virtual scene, 3D virtual scene, 360° VR scene and so on, an event of an LAE shall be controlled in the different type of MAR scene.
	[ ]	— provide a virtual camera control by an LAE action	
	[ ]	— provide a virtual object control by an LAE action	
	[ ]	— augment a virtual control by an LAE action	
LAESceneRepresentation	[ ]	— provide a spatial mapping table of an LAE into an MAR scene	*An LAE-MAR system shall provide at least one of both in an MARSceneRepresentation.
	[ ]	— provide an event mapping table of an LAE into an MAR scene	
Renderer	[ ]	— render integration result of an LAE and an MAR scene	*An LAE-MAR system shall render an MAR scene which is containing an LAE spatial mapping and/or event mapping results at least 30 fps.
	[ ]	— render integration result of an LAE virtual camera control and an MAR scene	
	[ ]	— render integration result of an LAE virtual object control and an MAR scene	
	[ ]	— render integration result of an LAE augmented object control and an MAR scene	
Display and UI	[ ]	— provide the rendered result on any kind of display such as 2D monitor and/or an HMD device	*Since there are different types of displays and user interfaces in an LAE-MAR system, the system shall provide them in more detail.
	[ ]	— provide user interfaces between an LAE scene and an MAR scene	