INTERNATIONAL STANDARD

ISO/IEC 3721

First edition
2023-09

Information technology — Computer graphics, image processing and environmental data representation —Information model for mixed and augmented reality content — Core objects and attributes

Reference number
ISO/IEC 3721:2023(E)

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 or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

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. In the IEC, see www.iec.ch/understanding-standards.

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 and www.iec.ch/national-committees.

Introduction

Mixed and Augmented Reality (MAR) refers to a spatially coordinated combination of media/ information components that represent, on the one hand the physical real world and its objects and on the other, those that are virtual, synthetic and computer generated. MAR, as an information medium, strives to provide rich experience based on realism, presence and augmentation.

In this document, a comprehensive set of information constructs for representing mixed and augmented reality (MAR) contents is described. This set of components extends the conventional ones used for representing virtual reality (VR) contents, as MAR environments are technically realized as virtual environments. The principles and requirements for the extension are laid out and the details of the component model including (but not limited to) those for representing *physical* real world objects, extending the virtual scene graph/structure to that for MAR (with the physical objects), how to spatially the physical objects into the MAR scene graph, associating these content components to the MAR system, and other miscellaneous constructs (e.g. event mapping, MAR events/ behaviours, video backdrops, etc.). This document is designed for the ease, generality and extendibility, and this is demonstrated with various examples and implementation results. The model will serve as a sound basis for establishing standard and interoperable file formats MAR contents in the future.

The document also provides definitions for terms as related to these MAR content informational components and their attributes.

The target audience of this document are mainly MAR system developers and contents designers interested in specifying MAR contents to be played by an MAR system or browser. The standard will provide a basis for further application standards or file formats for any virtual and mixed reality applications and content representation.

The extension will be self-contained in the sense that it is independent from the existing virtual reality information constructs, focusing only on the mixed and augmented reality aspects.

However, this document only specifies the information model, and neither promotes nor mandate to use a specific language, file format, algorithm, device, implementation method, and standard. The standard model is to be considered as the minimal basic model that can be extended for other purposed in actual implementation,

This document is based on the MAR Reference model (ISO/IEC 18039) that specifies for the contents-browser/player type reference architecture. The MAR content (in ISO/IEC 18039) is specified as the input that describes the scene and objects' behaviours, given to the browser/player which in turn parses, simulates and renders it to the display. The standard is the information model for the content.

As an extension to the virtual reality based contents or scene structure, this standard is very much related to the existing standard for VR scene representation such as ISO/IEC 19775-1 (X3D) and other related on-going standards such as the image-based object/environment representation for VR/MAR (ISO/IEC 23488) as well. There are also specific object models relevant to this standard such as those for the live actors and entities (ISO/IEC 18040 and ISO/IEC 23490) and MAR system sensor components (ISO/IEC 18038).

Information technology — Computer graphics, image processing and environmental data representation — Information model for mixed and augmented reality content — Core objects and attributes

1 Scope

This document specifies the information model for representing the mixed and augmented reality (MAR) scene/contents description, namely, information constructs for:

a) representing the virtual reality scene graph and structure such that a comprehensive range of mixed and augmented reality contents can also be represented;

b) representing physical objects in the mixed and augmented reality scene targeted for augmentation;

c) representing physical objects as augmentation to other (virtual or physical) objects in the mixed and augmented reality scene;

d) providing ways to spatially associate aforementioned physical objects with the corresponding target objects (virtual or physical) in the mixed and augmented reality scene;

e) providing other necessary functionalities and abstractions that will support the dynamic MAR scene description such as event/data mapping, and dynamic augmentation behaviours;

f) describing the association between these constructs and the MAR system which is responsible for taking and interpreting this information model and rendering/presenting it out through the MAR display device.

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*

ISO/IEC 18040, *Information technology — Computer graphics, image processing and environmental data representation — Live actor and entity representation in mixed and augmented reality (MAR)*

3 Terms, definitions and abbreviated terms

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in from ISO/IEC 18039 and ISO/IEC 18040 and the following apply.

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

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

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

3.1.1
aggregation
relation among objects/components that specifies a whole-part relationship as defined and used in the Unified Modelling Language literature

3.1.2
association
relation among objects/components (and their instances) that they can be linked to each other or combined logically or physically into an aggregation

Note 1 to entry: An aggregation is a specific form of an association as defined and used in the Unified Modelling Language literature.

3.1.3
behaviour
object or *component* that describes how certain objects and their attribute values change in time and/or in response to *events*

3.1.4
content
MAR content
MAR scene
description of mixed and augmented reality based experience to end-users by publishers or media producers

3.1.5
component
MAR component
object
MAR object
self-contained computational entity or model that has one or more input channels and/or one or more output channels

Note 1 to entry: In the context of MAR contents or system, a component has a relevant MAR functionality.

3.1.6
component model
object model
model of a collection of computational *objects* and their relationships serving a particular purpose or collective function

3.1.7
event mapping
object or *component* that maps events produced by the *MAR system* and maps/relates them to the corresponding ones in the *MAR scene/content* space

3.1.8
glass-based
type of an optical see-through device that has the form factor of optical glasses

3.1.9
inheritance
association representing a parent and child relationship

Note 1 to entry: In general, the child object, if not specified otherwise, will inherit the attribute model of the parent as defined and used in the Unified Modelling Language literature.

3.1.10
node
term referred to an actual implementation or specific format of an *object* or *component*

3.1.11
optical see-through device
apparatus or device that allows the *real/physical world* to be seen directly or through optical elements. in addition, the device allows an overlaid display of graphical (or *virtual*) elements

3.1.12
real transform group
RTG
transform group representing a group of objects situated at a *physical/real* spatial location

3.1.13
transform
abstract entity representing a relative spatial relationship (translation, rotation and scaling) with another coordinate system (or equally *transform*)

3.1.14
transform group
TG
component or object that represents a group of *objects* or *components* sharing a common physical or virtual coordinate system

3.1.15
video see-through device
apparatus or device that displays the real/physical world live through a video camera feed

3.1.16
virtual transform group
VTG
transform group representing a group of *objects* situated at a virtual spatial location

3.2 Abbreviated terms

AR	augmented reality
AV	augmented virtuality
CRT	capturer, recognizer, tracker
GNSS	global navigation satellite system
LAE	live actor and entity
MAR	mixed and augmented reality
MAR-RM	mixed and augmented reality reference model
MR	mixed reality
UI	user interface
VR	virtual reality
RTG	real transform group
TG	transform group
VTG	virtual transform group

4 Overview

MAR refers to the interactive medium that uses and merges the real (or physical) and virtual objects. [6] The two representative genres of MAR are augmented reality (AR) where virtual objects are added on to representation of the real physical world, while in augmented virtuality (AV) real world physical object representations are added to the virtual environment. The continued innovations and advances in computer vision, mobile/cloud computing and portable display devices have brought about a renewed interest in MAR, as a prominent information visualization and interaction medium. MAR not only has many application areas but also can empower users in daily activities by spatially augmenting useful information to key interaction objects. Traditionally, MAR services/contents have been developed as a single application using popular application programming interfaces (APIs). While there are moves to standardize the API[7,8],there is also a strong and natural move toward the separation of such singleton applications into contents (in some standard formats) and dedicated players/browsers, similarly to those for web documents.

X3D already offers a standard and declarative method of specifying dynamic virtual environments (ISO/IEC 19775-1) and is in fact being extended for MAR functionalities[9]. Such a service structure (and standards) can promote the proliferation of the given media and associated content form. Thus, in this document, a minimal (yet able to cover a comprehensive and reasonable, typical MAR contents) set of information constructs for representing MAR contents is established. The standard would extend the conventional constructs used for representing virtual reality contents (such as X3D), as virtual environments are used for the implementation platform for mixed reality contents as well.

First, the principles and requirements for the extension are laid out and the details of the component model including (but not limited to) those for representing real world physical objects, extending the virtual scene graph/structure to that for MAR (with the real objects), how to spatially the real objects into the MAR scene graph, associating these content components to the MAR system, and other miscellaneous constructs (e.g. event mapping, MAR events/ behaviours, video backdrops, etc.). The standard is designed for the ease, generality and extendibility, and this is demonstrated with various examples and implementation results. The model will serve as a sound basis for establishing standard and interoperable file formats MAR contents in the future.

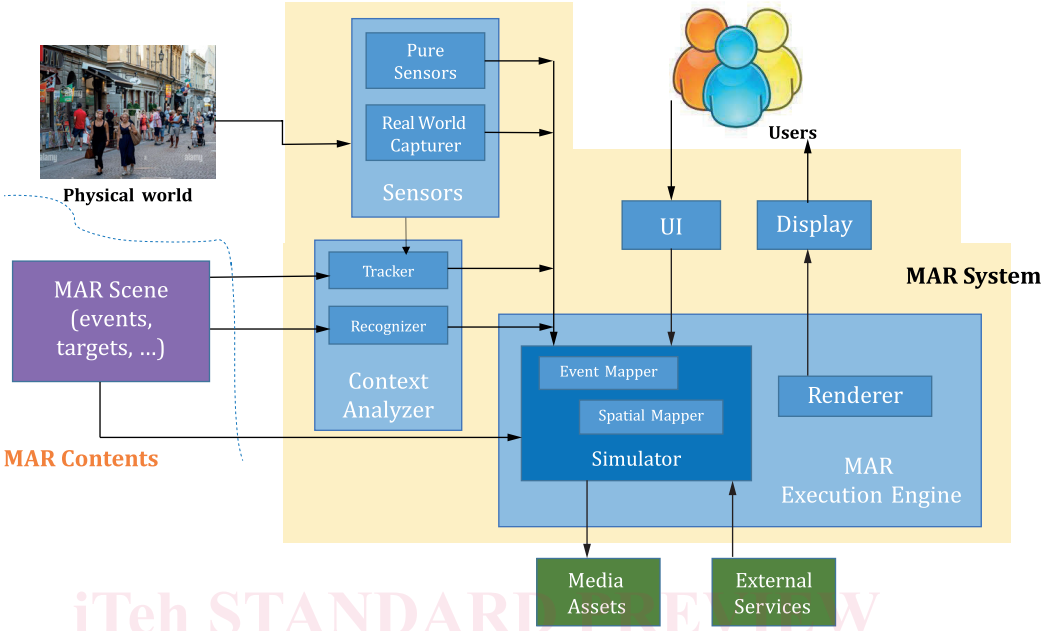
Most MAR systems have been implemented as a single application with all the contents and assets embedded in it, using programming libraries or APIs. The GPS equipped mobile and smart phones have made ways for location based augmented reality services (e.g. providing guides for commercial points of interest and tourism). Such a service necessitated the separation of contents (and its format specification) and the underlying player (to support the notion of one place-many contents). Future interoperability, reusability and proliferation of MAR contents will hinge on the comprehensible and extendable content model (and file format). For example, the file format for location based AR service called the ARML (Augmented Reality Mark-up Language) has been adopted as a standard for the Open Geospatial Consortium. ARML allows defining geographical points or landmarks of interest and associate GPS coordinates and simple augmentation contents (e.g. text, logos, and image).

The Web3D consortium has also been considering adding new components for MAR functionality to X3D, the ISO standard and declarative mark-up file format for representing virtual environments. The group has specified new and extended nodes to support e.g. video see-through based AR, such as the live video background, extended camera sensor nodes[9]. Some of its work is reflected in this standard as well. As the video augmented content is a popular form of MAR, the MPEG has proposed the ARAF (AR Application Format) that uses other MPEG standards (like MPEG-4 and MPEG-V) to specify video based and spatially augmented contents (ISO/IEC 23000-13).

Another class of approach of realizing, extending and specifying for MAR functionality and behaviour is through scripting. InstantReality[10] has developed its own extension to X3D and modules (e.g. for marker and image patch tracking allowing to express MAR contents with scripts through which the virtual objects (as expressed by the usual X3D formalism) can be associated with the position/pose tracked real world physical objects. There are several similar web-based MAR systems that can present AR contents with Javascript[1)] programming[11-13]. Despite these approaches, the current state with

1) JavaScript is a registered trademark of Oracle Corporation. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO or IEC.

respect to MAR content representation is still limited in its comprehensiveness (e.g. only covers certain genre of MAR such as location based or video based), often implementation dependent and proprietary, and lacks sufficient abstraction and clean modularization requiring lengthy and complicated script programming.



The MAR Scene or equivalently content description (purple box in the left part of the figure) is an input to the larger MAR system to be interpreted by it and be rendered to a display for the user consumption.

Figure 1 — The generic MAR system architecture as specified in the ISO/IEC 18039 MAR reference model

5 Principles and Requirements

The MAR reference model (ISO/IEC 18039) suggests the scope for which a general MAR system and content shall encompass as shown in Figure 1. Accordingly, the MAR content model should provide reasonable generality, and be able to express both AR and AV, the two notable genres along the MAR continuum under a consistent and unified representational framework. Other requirements are as follows:

— independent of particular sensor/device model, algorithm or implementation platform (such issues would be absorbed into the browser/player implementation),

— provide virtually any digital information and media, both static and dynamic, as augmentation such as text, images, videos, animation, HTML document elements, etc.,

— provide useful abstract and declarative constructs for often used content functionalities and minimize manual scripting or programming,

— make use of existing standard constructs, where possible,

— flexible/extendible to accommodate new future requirements.

To fulfil the final requirement (and others indirectly), this standard will use the component based approach, similarly to that of the X3D. Components in X3D are collections of objects that perform or represent similar operations, displays, or functions. Each component may define multiple levels. The levels indicate the complexity of the object or object features. The component framework is ideal for introducing and modelling new capabilities that satisfy MAR requirements (regardless of using X3D or

not) – new objects can be added or expanded as needed. The components form a structured environment that ensures consistent behaviour, usability and generality. The MAR reference model (ISO/IEC 18039) outlines the model architecture of a prototypical MAR system and indirectly points to the informational needs of an MAR content as an input to the system. In the next clause is described, the kind of new or expanded MAR components (equivalently functionalities) to be abstracted and represented.

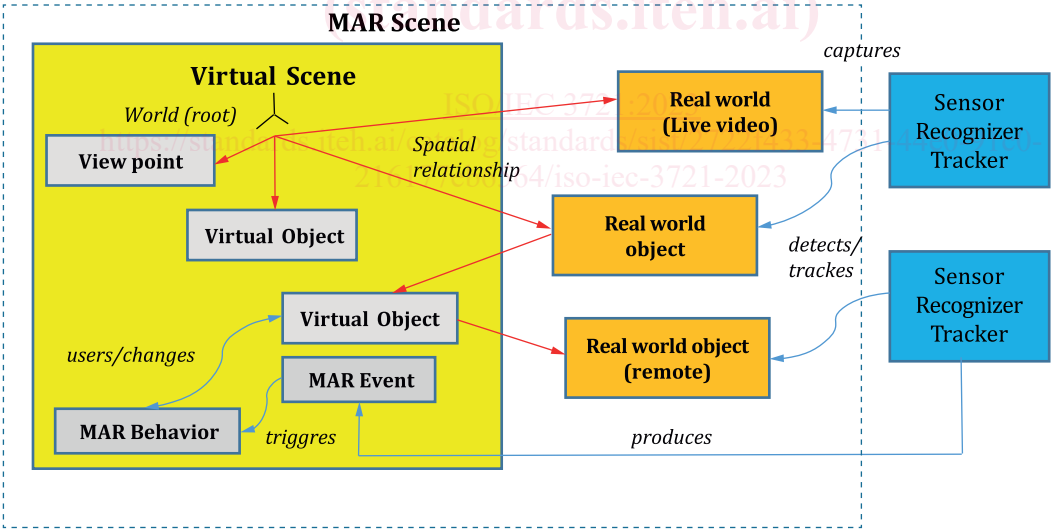
Again it is emphasized that the MAR content is the input to the system (purple box in the left part of Figure 1) which comprises the entities such as the sensor, capturer, tracker, recognizer, simulator, renderer, etc. Thus, while the content model may be associated with such system components, the system components themselves would not be part of the content description or model.

Note that the extension will be self-contained in the sense that it is independent from the existing virtual reality information constructs, focusing only on the mixed and augmented reality aspects.

6 MAR content model

6.1 Concept

MAR contents are technically realized as virtual reality contents with its scene containing special type objects that represent the real physical world/objects. In other words, one can think of a MAR scene that is technically a virtual scene with placeholders for the real world physical objects. The placeholders may be defined logically or spatially with respect to the virtual scene. These so called special type placeholder objects may need additional information of how it might be spatially registered into the implementation virtual scene. It is important to note that such a representation is (and should be) amenable to representing both AR and AV contents in a unified fashion.



Real world objects (in yellow) are associated with the virtual environment scene graph (in orange), and in turn the real objects are sensed, tracked, recognized and captured by the MAR system (in blue).

Figure 2 — The concept of extending virtual reality scene structure for MAR contents

Figure 2 illustrates the basic concept in which virtual objects (or the root of the virtual scene) provide a place for which certain real world physical object (remote or local) would map to in the MAR scene. This way, the real world physical objects are seamlessly represented within the virtual scene structure. For example, a virtual object can be spatially defined with respect to a real world physical object (or its placeholder). Note that the virtual scene may not necessarily be 3D (i.e. could be augmentation onto the 2D content), and real world physical objects can be individually recognized and tracked, or captured as a whole (e.g. live video) and implanted into the virtual scene. This is accomplished and specified by association with the MAR System components (e.g. blue boxes in Figure 2). The whole content