INTERNATIONAL STANDARD

ISO/IEC 23005-1

First edition
2011-04-01

Information technology — Media context and control —

Part 1: Architecture

Technologies de l'information — Contrôle et contexte de supports —

Partie 1: Architecture

Reference number
ISO/IEC 23005-1:2011(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword iv

Introduction v

1 Scope 1

2 MPEG-V System Architecture 1

3 Instantiations 4
3.1 Instantiation 1: Exchanges within the real world 4
3.2 Instantiation 2: Exchanges between real world and virtual world 6
3.3 Instantiation 3: Exchanges between virtual worlds 10
3.4 Instantiation 4: Control of avatars and other virtual objects by real world signals 15
3.5 Instantiation 5: Control of objects by signals from the virtual world 19

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. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

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.

ISO/IEC 23005-1 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

ISO/IEC 23005 consists of the following parts, under the general title *Information technology — Media context and control*:

— *Part 1: Architecture*

— *Part 2: Control information*

— *Part 3: Sensory information*

— *Part 4: Virtual world object characteristics*

— *Part 5: Data formats for interaction devices*

— *Part 6: Common types and tools*

— *Part 7: Conformance and reference software*

Introduction

The usage of multimedia content is becoming omnipresent in everyday life, in terms of both consumption and production. On the one hand, professional content is provided to the end user in high-definition quality, streamed over heterogeneous networks, and consumed on a variety of different devices. On the other hand, user-generated content overwhelms the Internet with multimedia assets being uploaded to a wide range of available Web sites. That is, the transparent access to multimedia content, also referred to as Universal Multimedia Access (UMA), seems to be technically feasible. However, UMA mainly focuses on the end-user devices and network connectivity issues, but it is the user who ultimately consumes the content. Hence, the concept of UMA has been extended to take the user into account, which is generally referred to as Universal Multimedia Experience (UME).

However, the consumption of multimedia assets can also stimulate senses other than vision or audition, e.g., olfaction, mechanoreception, equilibrioception, or thermoception. That is, in addition to the audio-visual content of, for example, a movie, other senses shall also be stimulated giving the user the sensation of being part of the particular media which shall result in a worthwhile, informative user experience.

This motivates the annotation of the media resources with metadata as defined in this part of ISO/IEC 23005 that steers appropriate devices capable of stimulating these other senses.

ISO/IEC 23005 (MPEG-V) provides an architecture and specifies associated information representations to enable the interoperability between virtual worlds, for example, digital content provider of a virtual world, (serious) gaming, simulation, DVD, and with the real world, for example, sensors, actuators, vision and rendering, robotics (e.g. for revalidation), (support for) independent living, social and welfare systems, banking, insurance, travel, real estate, rights management and many others.

Virtual worlds[1)] (often referred to as 3D3C for 3D visualization & navigation and the 3C's of community, creation and commerce) integrate existing and emerging (media) technologies (e.g. instant messaging, video, 3D, VR, AI, chat, voice, etc.) that allow for the support of existing and the development of new kinds of social networks. The emergence of virtual worlds as platforms for social networking is recognized by businesses as an important issue for at least two reasons:

a) it offers the power to reshape the way companies interact with their environments (markets, customers, suppliers, creators, stakeholders, etc.) in a fashion comparable to the Internet;

b) it allows for the development of new (breakthrough) business models, services, applications and devices.

Each virtual world however has a different culture and audience making use of these specific worlds for a variety of reasons. These differences in existing metaverses permit users to have unique experiences. Resistance to real-world commercial encroachment still exists in many virtual worlds where users primarily seek an escape from real life. Hence, marketers should get to know a virtual world beforehand and the rules that govern each individual universe.

Although realistic experiences have been achieved via devices such as 3-D audio/visual devices, it is hard to realize sensory effects only with presentation of audiovisual contents. The addition of sensory effects leads to even more realistic experiences in the consumption of audiovisual contents. This will lead to the application of new media for enhanced experiences of users in a more realistic sense.

Such new media will benefit from the standardization of a control and sensory information which can include sensory effect metadata, sensory device capabilities/commands, user sensory preferences, and various delivery formats. The MPEG-V architecture can be applicable for various business models for which

1) Some examples of virtual worlds are: *Second Life* (http://secondlife.com/), IMVU (http://www.imvu.com/) and Entropia Universe (http://www.entropiauniverse.com/).

audiovisual contents can be associated with sensory effects that need to be rendered on appropriate sensory devices.

Multi-user online virtual worlds, sometimes called Networked Virtual Environments (NVEs) or massively-multiplayer online games (MMOGs), have reached mainstream popularity. Although most publications tend to focus on well-known virtual worlds like *World of Warcraft*, *Second Life*, and *Lineage*, there are hundreds of popular virtual worlds in active use worldwide, most of which are not known to the general public. These can be quite different from the above-mentioned titles. To understand current trends and developments, it is useful to keep in mind that there is large variety in virtual worlds and that they are not all variations on *Second Life*.

The concept of online virtual worlds started in the late 70s with the creation of the text-based Dungeons & Dragons world MUD. In the eighties, larger-scale graphical virtual worlds followed, and in the late nineties the first 3D virtual worlds appeared. Many virtual worlds are not considered games (MMOGs) since there is no clear objective and/or there are no points to score or levels to achieve. In this report we will use “virtual worlds” as an umbrella term that includes all possible varieties. See the literature for further discussion of the distinction between gaming/non-gaming worlds. Often, a virtual world which is not considered to be an MMOG does contain a wide selection of mini-games or quests, in some way embedded into the world. In this manner a virtual world acts like a combined graphical portal offering games, commerce, social interactions and other forms of entertainment. Another way to see the difference: games contain mostly pre-authored stories; in virtual worlds the users more or less create the stories themselves. The current trend in virtual worlds is to provide a mix of pre-authored and user-generated stories and content, leading to user-modified content.

Current virtual worlds are graphical and rendered in 2D, 2.5 D (isometric view) or 3D, depending on the intended effect and technical capabilities of the platform: web-browser, gaming PC, average PC, game console, mobile phone, and so on.

“Would it not be great if the real world economy could be boosted by the exponential growing economy of the virtual worlds by connecting the virtual - and real world”; in 2007 the Virtual Economy in *Second Life* alone was around 400 MEuro, a factor nine growth from 2006. The connected devices and services in the real world can represent an economy of a multiple of this virtual world economy.

Virtual worlds have entered our lives, our communication patterns, our culture, and our entertainment never to leave again. It's not only the teenager active in *Second Life* and *World of Warcraft*, the average age of a gamer is 35 years by now, and it increases every year. This does not even include role-play in the professional context, also known as serious gaming, inevitable when learning practical skills. Virtual worlds are in use for entertainment, education, training, obtaining information, social interaction, work, virtual tourism, reliving the past and forms of art. They augment and interact with our real world and form an important part of people's lives. Many virtual worlds already exist as games, training systems, social networks and virtual cities and world models. Virtual worlds will change every aspect of our lives: the way we work, interact, play, travel and learn. Games will be everywhere and their societal need is very big and will lead to many new products and require many companies.

Technology improvement, both in hardware and software, forms the basis of this. It is envisaged that the most important developments will occur in the areas of display technology, graphics, animation, (physical) simulation, behavior and artificial intelligence, loosely distributed systems and network technology.

The figures in this part of ISO/IEC 23005 have been reproduced here with the permission of Samsung, Sharp Electronics, ETRI, University of Klagenfurt, Institute of Science and Technology, Myongji University, Institut national des télécommunications and the partners of the ITEA2 project Metaverse1: Philips, Forthnet S.A., Alcatel-Lucent Bell N.V., Innovalia, Alcatel-Lucent France, Technicolor, Orange Labs, DevLab, CBT, Nextel, Carsa, Avantalia, Ceesa, Virtualware, I&IMS, VicomTECH, E-PYME, CIC Tour Gune, Artefacto, Metaverse Labs, Technical University Eindhoven, Utrecht University, University of Twente, Stg. EPN, VU Economics & BA, VU CAMeRA, Ellinogermaniki Agogi, IBBT-SMIT, UPF-MTG, CEA List and Loria/Inria Lorraine.

Information technology — Media context and control —

Part 1: Architecture

1 Scope

This part of ISO/IEC 23005 specifies the architecture of MPEG-V (media context and control).

2 MPEG-V System Architecture

A strong connection (defined by an architecture that provides interoperability trough standardization) between the virtual and the real world is needed to reach simultaneous reactions in both worlds to changes in the environment and human behavior. Efficient, effective, intuitive and entertaining interfaces between users and virtual worlds are of crucial importance for their wide acceptance and use. To improve the process of creating virtual worlds a better design methodology and better tools are indispensible. For fast adoption of virtual worlds we need a better understanding of their internal economics, rules and regulations.

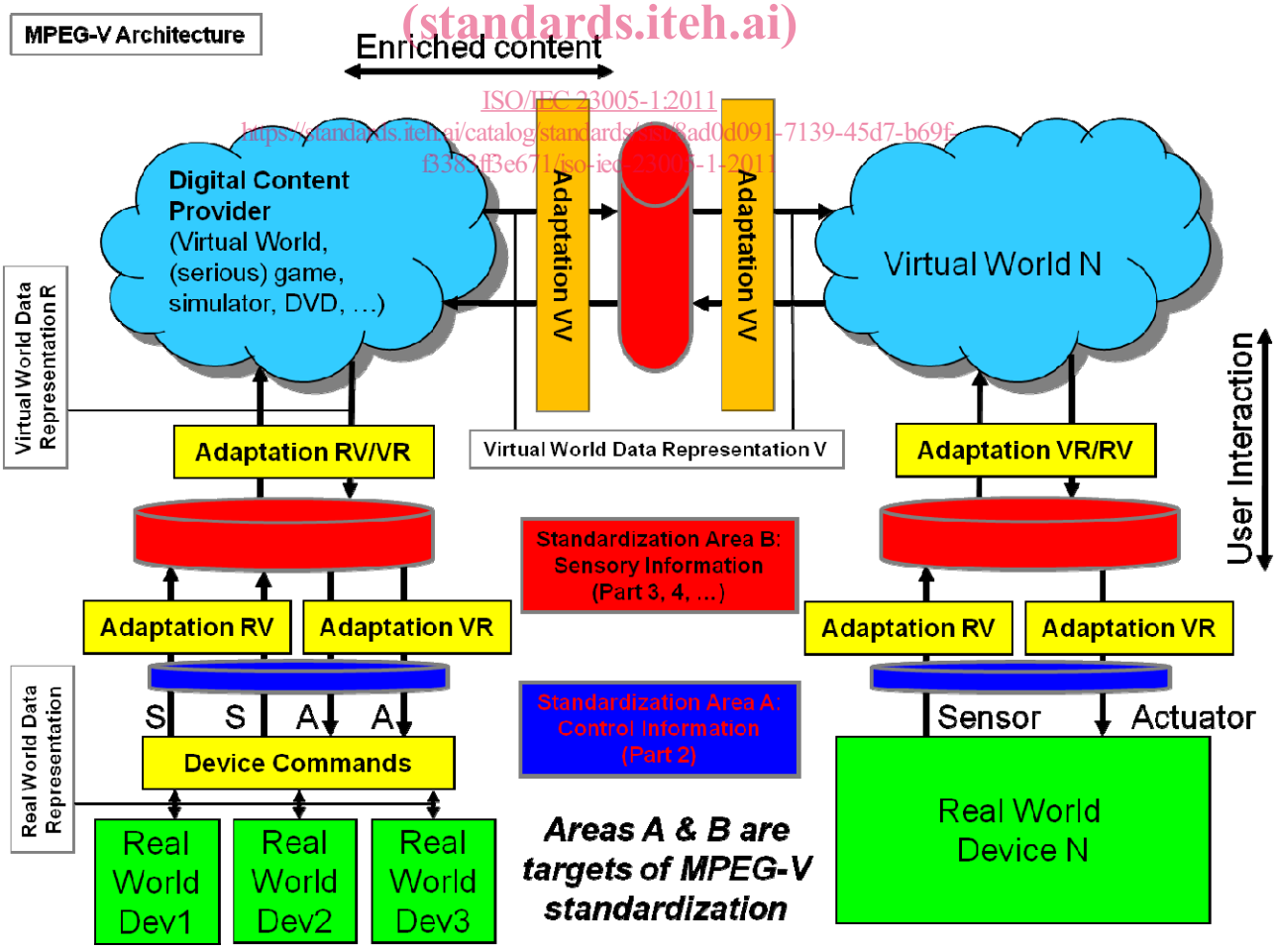

Figure 1 — System Architecture of the MPEG-V Framework

The overall system architecture for the MPEG-V framework is depicted in Figure 1 comprising the standardization areas a: control information and b: sensory information. Please note that standardization area b may be composed of multiple parts of the MPEG-V standard.

The individual elements of the architecture have the following functions:

— **Digital Content Provider**

A provider of digital content, real time or non real time, of various nature ranging from an on-line virtual world, simulation environment, multi user game, a broadcasted multimedia production, a peer-to-peer multimedia production or packaged content like a DVD or game.

— **Virtual World Data Representation R**

The native representation of virtual world related information that is intended to be exchanged with the real world (either exported or imported).

— **Virtual World Data Representation V**

The native representation of virtual world related information that is intended to be exchanged with another virtual world (either exported or imported).

— **Adaptation RV/VR**

The adaptation of the native representation of virtual world related information (that is intended to be exchanged with the real world) to the standardized representation format of MPEG-V in the standardization area B (e.g. sensory information, haptic information, emotion information …) in both directions: that is from the standardized representation into the native representation and vice versa.

— **Adaptation VV**

The adaptation of the native representation of virtual world related information (that is intended to be exchanged with another virtual world) to the standardized representation format of MPEG-V in the standardization area B (e.g. avatar information …) in both directions: that is from the standardized representation into the native representation and vice versa.

— **Sensory Information**

The standardized representation format of MPEG-V in the standardization area B (Sensory Information) (e.g. sensory information, haptic / tactile information, emotion information, avatar information …).

— **Adaptation RV**

The adaptation of the standardized representation of real world related information in the standardized representation format of MPEG-V in the standardization area A to the standardized representation of virtual world related information in the standardized representation format of MPEG-V in the standardization area B.

— **Adaptation VR**

The adaptation of the standardized representation of virtual world related information in the standardized representation format of MPEG-V in the standardization area B to the standardized representation of real world related information in the standardized representation format of MPEG-V in the standardization area A.

— **Control Information**[2)]

The standardized representation format of MPEG-V in the standardization area A (Control Information) (e.g. bi-directional control information, preference information, capability information …) related to the following elements of the architecture:

— Virtual World Data Representation R

— Virtual World Data Representation V

— Real World Data Representation

— **Real World Data Representation**

The native representation of real world related information that is intended to be exchanged with the virtual world (either exported or imported).

— **Device Commands**

Device commands is responsible for the adaptation of the native representation of real world related information (that is intended to be exchanged with the virtual world) to the standardized representation format of MPEG-V in the standardization area A (control information) (e.g. bi-directional control information, preference information, capability information …) in both directions: that is from the native representation into the standardized representation and vice versa.

— **Real World Device S**

A real world device containing a sensor (e.g. a temperature, light intensity, blood pressure, heartbeat …)

— **Real World Device A**

A real world device containing an actuator (e.g. a display, speaker, light speaker, fan, robot, implant …).

NOTE Real world devices can contain any combination of sensors and actuators in one device.

In the MPEG-V standard the following areas are addressed:

— **Standardization Area A: Control Information**

This area covers the information representation of the control information to and from devices in the physical world and into and from the virtual world. Examples of these representations are the representation of sensory input devices like smart vision systems, environmental and body sensors and the like and sensory output rendering devices like lights, heaters, fans, displays, speakers and the like.

— **Standardization Area B: Sensory Information**

This area covers the (bidirectional) information representation of information exchanged between the physical world and the virtual world as well as the information exchange between virtual worlds. Examples of these representations are the representation of haptic, emotion and avatar information.

2) In general, control information is strongly related to de-facto industry solutions for e.g. sensors, actuators and virtual worlds.

3 Instantiations

3.1 Instantiation 1: Exchanges within the real world

3.1.1 Instantiation 1.1: Representation of Sensory Effects (RoSE)

3.1.1.1 Introduction and Motivation

Traditional multimedia with audio/visual contents have been presented to users via display devices and audio speakers as depicted in Figure 2. In practice, however, users are becoming excited about more advanced experiences of consuming multimedia contents with high fidelity. For example, stereoscopic video, virtual reality, 3-dimensional television, multi-channel audio, etc. are typical types of media increasing the user experience but are still limited to audio/visual contents.

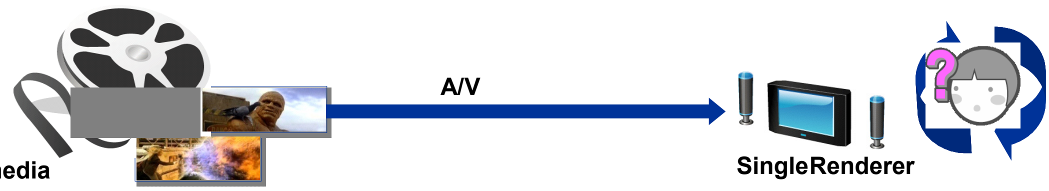

Figure 2 — Traditional Multimedia Consumption

From a rich multimedia perspective, an advanced user experience would also include special effects such as opening/closing window curtains for a sensation of fear effect, turning on a flashbulb for lightning flash effects as well as fragrance, flame, fog, and scare effects can be made by scent devices, flame-throwers, fog generators, and shaking chairs respectively. Such scenarios would require enriching multimedia contents with information enabling consumer devices to render them appropriately in order to create the advanced user experience such as described above. Figure 3 shows an example configuration adopting a multimedia multiple device (MMMD) approach for an advance user experience compared to the multimedia single device (MMMD) approach as illustrated in Figure 2. In this configuration, the multimedia contents are not rendered by a single device but with multiple devices in a synchronized manner.

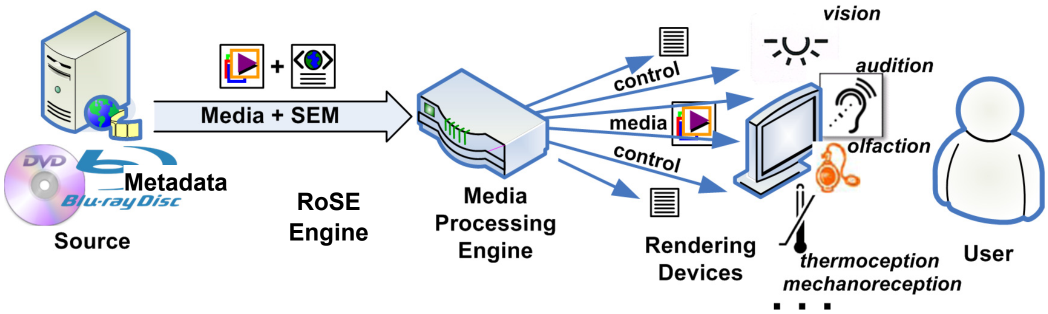

Figure 3 — RoSE-enabled Multimedia Consumption for Advanced User Experience

From a technical perspective, this requires a framework for the *Representation of Sensory Effects (RoSE)* information which may define metadata about special or sensory effects, characteristics of target devices, synchronizations, etc. The actual presentation of the RoSE information and associated audio/visual contents allows for an advanced, worthwhile user experience.

The next Subclause provides a brief overview of the RoSE system architecture.

3.1.1.2 RoSE System Architecture

The overall system architecture for the RoSE framework is depicted in Figure 4 comprising Sensory Effect Metadata, Sensory Device Capabilities, Sensory Device Commands, User Sensory Preferences, and a so-called RoSE Engine which generates output data based on its input data.

It is important to note that the *Sensory Effect Metadata*, *Sensory Device Capabilities*, *Sensory Device Commands*, and *User Sensory Preferences* are within the scope of standardization and, thus shall be normatively specified. On the other side, the *RoSE Engine* as well as *Provider* entities and *Consumer Devices* are informative and are left open for industry competition.

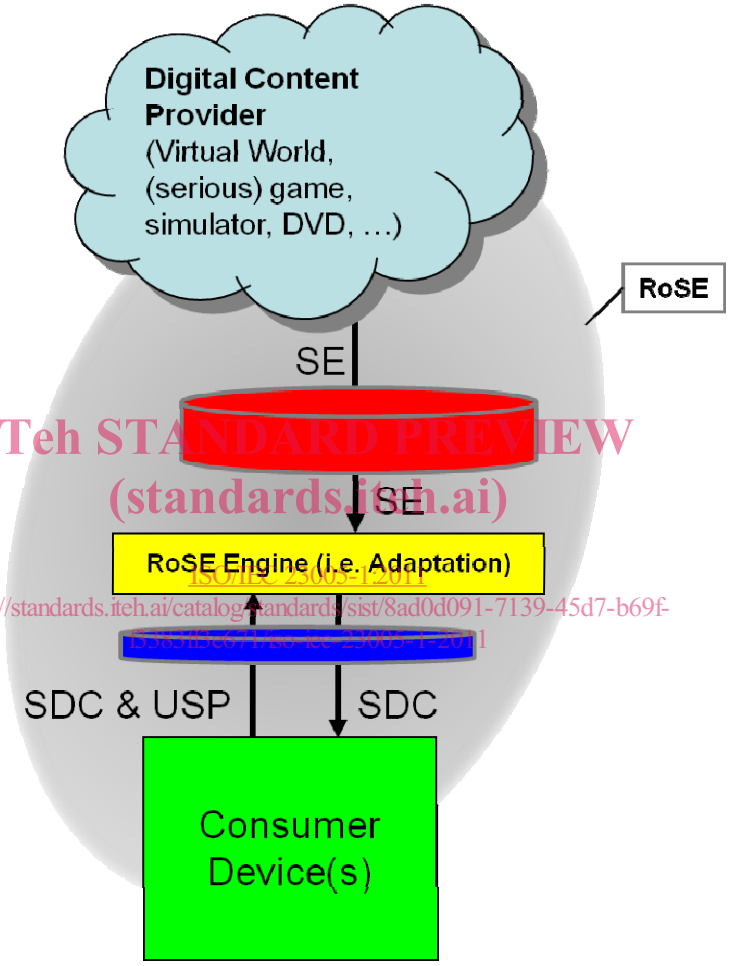

Figure 4 — RoSE System Architecture

A *provider* within the RoSE framework is referred to as an entity that acts as the source of the *sensory effect metadata* such as a broadcaster, content creator/distributor, or even a service provider. The *RoSE Engine* is an entity that takes the *sensory effect metadata*, the *sensory device capabilities* and the *user sensory preferences* as inputs and generates *sensory device commands* based those in order to control the consumer devices enabling a worthwhile, informative experience to the user.

Consumer devices are entities that act as the sink of the *sensory commands* and act as the source of *sensory device capabilities*. Additionally, entities that provide user sensory preferences towards the RoSE engine are also collectively referred to as consumer devices. Note that *sensory devices* (see below) are sub-set of consumer devices including fans, lights, scent devices, human input devices such as a TV set with a remote control (e.g., for preferences).

The actual *sensory effect metadata* provides means for representing so-called *sensory effects*, i.e., an effect to augment feeling by stimulating human sensory organs in a particular scene of a multimedia application. Examples of sensory effects are scent, wind, light, etc. The means for transporting this kind of metadata is referred to as sensory effect delivery format which, of course, could be combined with an audio/visual delivery format, e.g., MPEG-2 transport stream, a file format, or Real-time Transport Protocol (RTP) payload format, etc.

The *sensory device capabilities* define description formats to represent the characteristics of sensory devices in terms of which sensory effects they are capable to perform and how. A sensory device is a consumer device by which the corresponding sensory effect can be made (e.g., lights, fans, heater, fan, etc.). *Sensory device commands* are used to control the sensory devices. As for sensory effect metadata, also for sensory device capabilities and commands corresponding means for transporting this assets are referred to as sensory device capability/commands delivery format respectively.

Finally, the user sensory preferences allow for describing preference of the actual (end) users with respect to rendering of sensory effects for also a delivery format is provided.

3.2 Instantiation 2: Exchanges between real world and virtual world

3.2.1 Instantiation 2.1: Full motion control and navigation of avatar/object with multi-input sources

Full motion control and navigation of avatar/object with multi-input sources allows for the full motion control and navigation of 3D objects and avatars in a Virtual World. Recently, user interest in human-computer interaction has grown considerably based on large volumes of recent research. Through the development of VR technology, it has applied to various fields. Especially, the Entertainment area is commercialized such as 3D virtual online communities like *Second Life* and 3D Game station. Nintendo Wii provides new game experience using 3D input device. Especially the control of objects and avatars in 3D virtual space requires more complex methods than conventional input devices such as mouse, keyboard, joystick and etc. The Figure below shows the example picture of these systems and like this style, it is applied to home, school or other place for various purposes such as entertainment or education including digital contents of 3D virtual world.

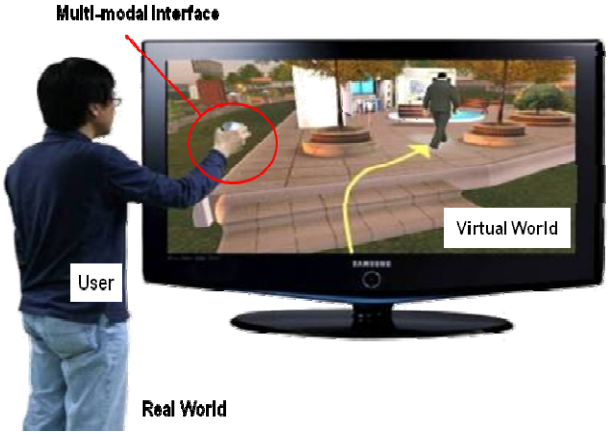

Figure 5 — Full motion control and navigation of avatar/object with multi-input sources