Construction of a haptic-enabled broadcasting system based on the MPEG-V standard

Jaeha Kim\textsuperscript{a}, Chang-Gyu Lee\textsuperscript{b}, Yeongmi Kim\textsuperscript{a}, Jeha Ryu\textsuperscript{a,*}

\textsuperscript{a} School of Mechatronics, Gwangju Institute of Science and Technology, Republic of Korea
\textsuperscript{b} Department of Medical System Engineering, Gwangju Institute of Science and Technology, Republic of Korea

\textbf{Abstract}

With rapid developments in communications technology and digital multimedia, there has been increasing demand in recent years for realistic broadcasting technology beyond conventional audio-visual media. In response to this demand, this paper presents an example of the construction of a haptic-enabled broadcasting system based on the MPEG-V standard that was established recently. The construction processes of the proposed haptic-enabled broadcasting system include various types of media acquisition, haptic content creation by modeling and authoring, transmission, rendering, and interaction. This paper illustrates the data flow within the system, from the creation of haptic contents to the rendering of these contents to the end user, and explains a method of building the system with the MPEG-V standard. The constructed haptic-enabled broadcasting system allows users to have more immersive interaction with the synthesized haptic multimedia, which is closely synchronized with audio-visual data.

\section{Introduction}

In accordance with considerable recent developments in communications technology and digital multimedia, high-tech audio-visual devices, such as ultra-high-definition (UHD) video, 3D television, and multichannel 3D sound systems, have come into wide use. These devices have been actively developed to provide viewers with more immersive and realistic interactions with multimedia. At present, some theme parks and movie theaters provide immersive and interactive entertainment to the public. For example, Disneyland and Universal Studios provide 4D contents for a more realistic experience, incorporating 3D stereoscopic pictures, multichannel audio effects, and some tactile effects such as vibrating chairs synchronized with a melody or with the context. Reflecting these developments, demand for more immersive and interactive broadcasting technology beyond conventional audio-visual media has been increasing in recent years. A haptic-enabled broadcasting system is one possible means of allowing viewers to touch and feel audio-visual media using conventional TV in the near future.

Haptics is a research field concerned with the sense of touch, allowing users to have deeper recognition and immersiveness by providing tactile or kinesthetic stimuli such as force, torque, vibration, and sensory experience of temperature, with respect to virtual, augmented, or real environments. Here, haptic sensation includes both tactile and kinesthetic sensations. Tactile sensations are invoked in the mechanoreceptors in the skin, while kinesthetic sensations are invoked by the sensors in the joints, muscles, and tendons to provide feelings of force and torque. Haptic interaction has been integrated typically within fully synthesized virtual reality (VR) worlds such as in the stand-alone system in \cite{1}.

In addition to conventional media such as images, video, and audio, haptics is now expected to play a prominent role in multimedia applications \cite{2,3}. Some pioneers have already worked on such applications.
O’Modhrain and Oakley [4] discussed a potential role that the haptic interaction could play in supporting a greater sense of immersion within broadcast content. They have proposed presentation interaction that would allow the relocation of a character’s rendered position in a scene with a small two degree-of-freedom (DOF) force feedback interface. Motivated by this work, many haptic-enabled broadcasting system concepts and basic algorithms have been suggested [5–14]. Cha et al. [13] proposed a framework for haptic broadcasting system, and presented some potential haptic interaction scenarios in broadcasting such as feeling haptic data, touching a 2.5D natural scene, and touching and manipulating 3D models. Later, Cha et al. [15] described a system for recording and annotating haptic information that is time-referenced to a movie, and then replaying the recorded haptic information for a viewer. Yamaguchi et al. [16] proposed a system that automatically generates haptic effects from 2D graphics, relying on metadata that describes the movement characteristics of the contents. Communication of haptic media has become an important issue because haptic media have different requirements in comparison with traditional audio-visual media. Eid et al. [17] proposed an adaptive application layer multiplexing framework that includes a communication protocol for multimedia applications incorporating audio-visual, haptic, and scent data. They presented an adaptive and intelligent multiplexer for multimodal input media streams based on the application requirements and network conditions. King et al. [18] proposed the use of the Session Initiation Protocol (SIP) for multimodal telepresence sessions including audio, video, and haptic data. In [19], Steinbach et al. extensively summarized recent advances in the area of haptic communications from both psychophysical and technical viewpoints, including telepresence, haptic control architecture, performance evaluation, perceptual coding, and communication protocols.

In order to systematically describe various elements related to a haptic application, some researchers have proposed a structured data format. Cha et al. [13] extended MPEG-4 BIFS to support the representation of synchronization and the transmission of haptic data and audio-visual media. El-Far et al. [20] proposed HAML, which is an XML-based language intended to describe haptic-related information, including haptic devices, haptic development APIs, virtual environments, and communication. King et al. [18] introduced a haptic codec for transcoding raw haptic device data into a common format, which provides standardized data interfaces for easy interconnection between devices, and also allows tele-operation capabilities to be added to existing multimedia applications. Recently, in 2011, the MPEG-V standard was established. The MPEG-V standard has a wide coverage; with respect to haptic application, it can describe material properties such as mass, stiffness, and damping of virtual objects, and it can also describe various effects such as tactile or kinesthetic feedback, supporting various sensors and actuators [25–30]. Because both HAML [20] and MPEG-V provide extensive support for audio, video, haptic, and scent data, they overlap to some extent. For example, both can describe haptic properties such as stiffness, damping, and static/dynamic friction. However, HAML and MPEG-V are different in some aspects. First, while HAML has transmission functionality, transmission is not within the scope of MPEG-V. Second, while HAML requires relatively thorough descriptions, MPEG-V requires relatively general descriptions and supports not only haptic-related devices and effects but also other sensors and actuators, and various effects such as light, fog, and vibration.

Several works have been published recently with respect to the MPEG-V standard [21–24]. Pyo et al. [21] suggested a device-rendered sensible media and metadata schema for representing effect and control information, and designed a service framework for device-rendered sensible media based on the UPnP framework. Choi et al. [22] introduced a new-generation media service called Single Media Multiple Devices (SMMD), which is based on Sensory Effect Metadata (SEM) as defined in the MPEG-V standard. Timmerer et al. [23] introduced the concept of sensory experience by utilizing sensory effects as another dimension contributing to the quality of the user experience (QoE). In [24], Yoon et al. proposed a framework for 4D broadcasting systems based on the MPEG-V standard, and presented guidelines for building it.

In this paper, on the basis of the previously proposed haptic-enabled broadcasting system framework [13], we present an example of the construction of a haptic-enabled broadcasting system based on the MPEG-V standard. More specifically, we discuss how haptic content is created and rendered with the MPEG-V standard. Since the MPEG-V standard focuses on data transmission is out of its scope. The transmission protocol, therefore, can be chosen suiting service providers, such as the MPEG-4 standard, an IEC/ISO standard for streaming multimedia objects in broadcasting-specific applications [38].

This article is organized as follows. In Section 2, we briefly discuss the MPEG-V standard and some topics related to the building of a haptic-enabled broadcasting system. In Section 3, we explain the construction of the haptic-enabled broadcasting system stage by stage, some open issues and research challenges are discussed in Section 4, and finally, our conclusions are presented in Section 5.

2. Organization and scope of the MPEG-V standard

Moving Picture Experts Group (MPEG) has produced numerous standards for various technologies related to audio and video compression and transmission. Geared by dramatically evolving technologies and societal changes, multimedia is emerging more and more and will be everywhere in our life. Now, many people are beginning to enjoy more immersive experience, which stimulates senses other than vision or audition; accordingly, multimodal multimedia content is coming into the spotlight, and interaction and interoperability between multimedia content and various sensors and actuators are becoming important issues. In recent years, the MPEG-V (ISO/IEC 23005) standard was established [25–30]. This standard
provides an architecture [25] (see Fig. 1) and associated information representations for interaction and interoperability between virtual worlds and real worlds through various sensors and actuators.

The MPEG-V standard is divided into seven parts. The relationships between the parts and the scope of each part are shown in Fig. 2 and Table 1, respectively [25–30]. Part 1 deals with the MPEG-V system architecture and the overall scope, and introduces various application scenarios. Part 2 provides control information for manipulating devices in the real world as well as in virtual worlds. The scope of part 2 covers the interfaces between the adaptation engine and the capability descriptions of various sensors and actuators in the real world, and preference information. Part 3 defines various sensory effects that a content creator might wish to provide to users; the number of these effects is very large and growing. For instance, they include light, temperature, wind, vibration, scent, tactile, and kinesthetic effects, and so on. Part 4 deals with various properties a virtual object or avatar can have, such as appearance, animation, scent, sound, and haptic. It also includes information related to manipulation of virtual objects (or avatars), such as scaling and setting up position and orientation, and so on. Part 5 deals with data formats for interaction devices, such as device commands and sensed information. This part provides a general data format for various sensors and actuators. Part 6 provides common data types and tools, and part 7 provides conformance and reference software. The adaptation engine is an interface that enables communication between the virtual world and real world. It takes six inputs – sensory effects, the user’s sensory effect preferences, sensory device capabilities, sensor capability, sensor adaptation preference, and sensed information – and then generates outputs with device commands to control external devices in the real world or to manipulate and feel virtual world objects. Note that the adaptation engine (RV or VR engine) is not within the scope of the MPEG-V standard. Please see [25–30] for more details.

Immersive haptic broadcasting is potentially a good example of a MPEG-V applications. Some typical scenarios of immersive haptic broadcasting are illustrated in Fig. 3. For instance, in a home shopping scenario, while a viewer sees the product (e.g., a smart phone or PDA with haptic capability) and listens to a description of it, the viewer can touch it by feeling the surface of the PDA or pressing the button with a click feeling using a force-reflecting haptic

![Fig. 1. System architecture of the MPEG-V standard [25,42]. The numbers in brackets refer to specific parts of the standard.](image1)

![Fig. 2. Interoperability between the virtual world and the real world, and the relationship between the parts of the MPEG-V standard.](image2)

---

**Table 1**

Scope of each part of the MPEG-V standard.

<table>
<thead>
<tr>
<th>Part</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part 1</td>
<td>Provides an overview and introduction to MPEG-V, and describes its main architecture, the relations between its parts, and various application scenarios.</td>
</tr>
<tr>
<td>Part 2</td>
<td>Defines control information such as capability and preference information required to provide interoperability in manipulating devices in the real world as well as in virtual worlds.</td>
</tr>
<tr>
<td>Part 3</td>
<td>Describes various sensory effects represented in the virtual world and/or real world.</td>
</tr>
<tr>
<td>Part 4</td>
<td>Characterizes virtual objects, virtual avatars, and their various properties.</td>
</tr>
<tr>
<td>Part 5</td>
<td>Defines data formats for interaction devices, i.e., device commands and sensed information.</td>
</tr>
<tr>
<td>Part 6</td>
<td>Provides common data types and tools.</td>
</tr>
<tr>
<td>Part 7</td>
<td>Provides reference software.</td>
</tr>
</tbody>
</table>
device (not shown in the figure) before actually buying the product. In a fishing experience scenario, a viewer can feel vibration with a haptic device when a fish takes bait, and may also feel tension or dragging forces when the fisherman on the screen catches the fish. These examples show some typical use cases of providing haptic sensory effects based on the captured audio-visual scenes or on the characteristics of virtual world objects. These scenarios can be implemented with parts 2, 3, 4, and 5 of the MPEG-V standard. In the case of the haptic device, the DOF, the maximum force and torque level, position, velocity, and so on, can be specified with Sensory Device Capability and Sensor Capability in part 2. In order to provide haptic-related effects, the content creator can use TactileEffect or KinestheticEffect in part 3. Haptic properties such as stiffness, damping, friction, and the virtual environment can be described with part 4. TactileType, KinestheticType, PositionSensorType, ForceSensorType, and so on, of part 5 can be used as data formats for interaction devices. This will be explained in detail in the following section.

3. Construction of a haptic-enabled broadcasting system based on MPEG-V

A haptic-enabled broadcasting system consists of four main stages: (1) acquiring various media sources; (2) editing and authoring the media sources to create haptic contents; (3) transmitting (delivering) the created haptic contents over the broadcasting channel; and (4) rendering the transmitted haptic contents [13]. Fig. 4 shows the general framework of a haptic-enabled broadcasting system that is based on the MPEG-V standard. In this framework, a haptic-enabled broadcasting system can be implemented as follows. Haptic contents are created using the MPEG-V standard, and they are then encoded/decoded with BIFS encoders/decoders and later transmitted through the MPEG-4 framework, and finally played using the MPEG-V standard in the rendering stage. The created haptic contents go through coding and multiplexing and are delivered to viewers through many transmitting channels. The transmitted haptic contents undergo several processes, such as demultiplexing, decoding, and scene composition, and are finally rendered with diverse audio/visual/haptic interfaces. With this haptically enhanced content, viewers can see, hear, and even feel the multimedia simultaneously.

Haptic contents can be defined as a data set that includes synthesized and closely synchronized information about haptic and audio-visual data, as shown in Fig. 4. Haptic data may be classified into three categories and thus represented in part 4 of the MPEG-V standard: (i) haptic material properties, such as stiffness, static/dynamic friction, damping, haptic textures, and mass (or weight); (ii) dynamic force effects, such as spring, damping, and motion data; and (iii) tactile properties, such as temperature, vibration, and tactile patterns. In order to provide haptic sensations for the virtual objects that the user touches, haptic content may include the geometry of an object or refer to its location (e.g., using a URI). Haptic data may use a 3D geometry mesh that can be created by 3D modeling tools such as 3DS-MAX, Maya, and Blender, or may use 2.5D depth images that can be obtained using Z-cam [31] or KINECT [32] to provide not only visual data but also tactile geometry data. After acquiring the geometry of a virtual object, various haptic material properties can be added to the geometry models through a series of editing (or modeling) processes to make viewers feel the given object. This process can be carried out using a haptic modeler such as HAMLAT [35] or K-Haptic modeler [33,34]. Moreover, a haptic content creator can add dynamic force effects for manipulating the virtual object. For example, putting a simple spring-damper system behind a small box model can be used as a good model for pushing or clicking a button. Haptic contents can also include motion data that can be acquired from motion capture devices; the trajectory of the captured motion data can be rendered with appropriate kinesthetic feedback. This can be applied to train skillful behaviors such as calligraphy [36] or the conducting of an orchestra [14]. The haptically modeled (or edited) data can be authorized and synchronized in a time frame with the audio-visual data to become the final haptic contents, as shown in Fig. 4.

A haptic content creator collects various media sources and tries to represent them in the MPEG-V format. In particular, the content creator can add various haptic sensory effects from part 3 of MPEG-V, such as
TactileEffect and KinestheticEffect, and can edit the scene with their appropriate use. Since haptic content consists of various media, we need a way to integrate them into one scene, and we also need a way to transmit the created content via a broadcasting system. For this purpose, a scene description representation such as MPEG-4’s BIFS (Binary Format for Scenes) [37] may be used, as presented in our previous research [13,14]. A scene descriptor describes the spatial and temporal arrangements of the media elements in a scene. Alternatively, we can consider the use of the methods introduced in Section 1 [17,18]. Note that the method of processing the created content is not in the scope of the MPEG-V standard. In other words, the MPEG-V standard is not concerned with methods of encoding/decoding, transmitting, (de)multiplexing, and rendering. Therefore, the adaptation engine must take charge of all such processes. Consequently, the implementation details may vary according to the choice of sensors and actuators, the communication method, the rendering and control algorithm for devices, and so on. Given this flexibility, it is possible to construct the system according to one’s preference.

In the terminal, to set various sensors and actuators into operation, we need a means of describing device commands, and part 5 of the MPEG-V standard plays a role in this. For example, the audio-visual data are rendered with a screen and speakers. The haptic data are properly processed in the adaptation engine in accordance with Sensory Device Capability, Sensor Capability, and User’s Sensory Effect Preference of part 2 of the MPEG-V standard [26], and it is rendered through a kinesthetic device or a tactile device. In the following section, we systematically illustrate in detail how to construct a haptic-enabled broadcasting system based on MPEG-V.

3.1. Acquiring various media sources for haptic content creation

As mentioned in the previous section, haptic content consists of various audio-visual media and haptic (kinesthetic/tactile) information. Therefore, for haptic content creation, haptic as well as audio-visual data must first be acquired. For example, in the home shopping scenario in Fig. 3(a), content creators can obtain visual information such as color, shape, and audio data for environments by using a common camcorder. The geometry data of the shopping host may also be captured as 2.5D depth images by using Z-cam [31]. Of course, a full 3D model may be reconstructed from video scenes of the shopping host. A complex 3D mesh model for a dynamic scene, however, may be time-consuming to generate and heavy for transmission because of the need to calculate the position and orientation of the object in the scene frame by frame.
On the other hand, a product on sale (e.g., the PDA in Fig. 3(a)) needs to be carefully modeled as a 3D mesh model for geometry description with visual properties. Haptic properties such as stiffness, damping, friction, and weight can also be acquired by direct physical measurements made by, for example, a force/torque sensor. Later, these haptic properties are to be mapped onto the geometry model by haptic modeling tools [33–35] in order to help provide potential customers with a comprehensive description of the product.

3.2. Editing and authoring haptic content

After collecting all ingredients to generate haptic content, both spatial and temporal editing and authoring with audiovisual media is required. Haptic content editing and authoring is the process of appropriately arranging all audio/video/haptic media sources spatiotemporally according to the director’s intention so that these sources form a unity that clearly and effectively conveys the intent of the content creator. This stage, essentially one of editing and arranging, fundamentally relies on the media format properties.

A virtual world and virtual objects in a scene with their visual and haptic material properties can be described with part 4 of the MPEG-V standard, by using Appearance, stiffness, friction, damping, mass, texture, and so on (Fig. 5). For the home shopping example, in order to make viewers feel the shape, buttons, and texture of the given virtual object (the PDA), the object should have a 3D mesh model (Appearance) and its properties (MaterialPropertyType). Therefore, the content creator should model the virtual object by using a 3D graphic modeling tool such as 3DS-MAX and Maya, and in the meantime, might also assign haptic properties such as stiffness, friction, and damping by using haptic modeling tools.

On the other hand, the haptic content creator can use TactileEffect in part 3 to provide viewers tactile feedback in a series of movie scenes. Tactile effects can be given through various physical phenomena, such as vibration, thermal energy exchange, and pressure. They can be applied directly onto some areas of human skin through many types of actuators, such as motor-based vibrators, pneumatic actuators, piezo actuators, and thermal actuators. A tactile effect may effectively be represented by an ArrayIntensity or a TactileVideo, or by audio file(s) containing the waveform of the tactile effect. Fig. 6 shows the use of TactileVideo for authoring vibrotactile effects onto some scenes (e.g., the actor jumps from one train to the other by punching the wall of the train in the movie “Ghost”). For this effect, for example, we can choose TactileVideo to use and connect a TactileResource element to the output file of the tactile authoring tool. The output file may be in a well-known format such as *.avi, containing grayscale data that include intensity information for array-type tactile effects for multiple actuators. Such fundamental attributes of Sensory Effect as id, activate, duration, fade, and so on, are used, where id is used to distinguish the effects, activate is used for checking whether the corresponding effect is activated, duration describes the total length of the effect, and fade describes the period of time within which the defined intensity shall be reached. In addition, TimeSamples, the number of times data is updated per second, is used.

The final sensory effect to be felt by the end-user might be different from the original intention of the haptic content creator, since the final effect depends on the user preference and sensor/device capability, as described in part 2 of the MPEG-V standard. For example, even though the haptic content creator intends to give users 100 °C with TemperatureType, this temperature cannot be produced or even clipped off if the maximum displayable temperature of the device is limited to 50 °C. Similarly, the original intention of the haptic content creator can be changed if the user sets the maximum or minimum by using HeatingCapabilityType in part 2 of the standard.

After creating the description of the content with the MPEG-V standard, the created haptic content is transmitted.

3.3. Transmission of haptic content

Aforementioned, the MPEG-V standard is not concerned with the issue of which protocol is going to be
used for the transmission of the haptic content. Therefore, any transmission protocol may be used, be it TCP/UDP/IP or MPEG-2 Transport Stream (ISO/IEC 13818, ATM Adaptation Layer 5 (AAL5)), for example.

In general, MPEG-4 can also be used for this stage, since MPEG-4 BIFS, an extensible format for representing audio, video, and other media as a set of objects with synchronized functionality, provides functionality for the transmission of multimedia contents [38]. However, to additionally achieve the necessary functionality and support for the synchronized representation and transmission of haptic data, MPEG-4 BIFS should be extended to cover haptic data. In other words, MPEG-4 BIFS needs to be supplemented with some newly defined nodes and rendering methods to handle haptic data. Due to the wide range and complexity of haptic information, a haptic-enabled broadcasting system needs to support a wide variety of data types. Please refer to [13] for a detailed description of how to extend BIFS nodes to be able to handle haptic data.

Note that the use of two different types of metadata might necessitate internal data conversion in the adaptation engine. MPEG-V data is not directly handled by MPEG-4 BIFS, and hence we need data conversion (or adaptation) between the MPEG-V data and the BIFS data. Fig. 7 shows some examples of how to convert MPEG-V data to MPEG-4 BIFS data.

3.4. Rendering of haptic content

In the viewing and interaction stage, viewers can enjoy either active haptic interaction or passive haptic playbacks synchronized with an audio-visual scene beyond traditional watching and listening. Note that, even though the MPEG-V standard provides sufficient description tools regarding when actuators have to be activated/deactivated (e.g., a timestamp and a flag for enabling activation of the actuator), the proper synchronization of audiovisual media and haptic data is the responsibility of the adaptation engine; otherwise, it should be arranged in the haptic content creation stage (e.g., manually in the authoring stage by the haptic content creator). For example, owing to problems with real-time data transmission in a broadcasting system, there could be a synchronization problem. In the case of passive haptic playbacks with vibrotactile arrays and a Bluetooth connection, there might be a hardware limitation, such as slow response of actuators, time delays, data losses, and disconnection. If the given actuator is activated about 100 ms after it receives the device command, for example, and if the haptic content creator wants to overcome this delay, then, he or she can forward the activation time of the device command by 100 ms. Such problems need to be appropriately handled by the content creator manually, or by the adaptation engine. In the case of active haptic interaction, the system possibly has to provide kinesthetic feedback, and this kind of feedback usually requires an update rate greater than 1 kHz. In this situation, generating and issuing a device command every millisecond might be a burden; even the system might not be able to generate device commands at such a rate. The adaptation engine or, more precisely, the haptic controller is in
charge of handling this time constraint, and it must provide suitable ways of attenuating side effects caused by this disharmonious synchronization problem. We will briefly explain the haptic controller later.

Fig. 8 shows the stage of rendering haptic content. In this stage, the transmitted haptic content is fed to a demultiplexer so that it is separated into several independent streams. These streams are decoded, and the decoded media is fed to a compositor process that has access to the BIFS scene graph. Traditionally, the compositor scans the scene graph, determines what audio and visual content should be shown, and then passes this information to the audio-visual renderers that actually handle the display of the media on an entertainment device, such as a TV. Haptic content and metadata are then processed in the adaptation engine, which refers to Sensory Device Capabilities, User’s Sensory Preferences, and so on, of part 2 of the MPEG-V standard, and finally generates Device Commands (in part 5 of the standard) to control the haptic device in rendering the haptic data.

Fig. 9 shows the structure of the haptic interaction subsystem for kinesthetic feedback. Haptic devices may be divided into two categories: kinesthetic devices and tactile devices. Fig. 10 shows examples of various haptic devices. Typically, a kinesthetic device measures the position and orientation of the device as the human operator grips the end effector and moves around. In the virtual world, there is a point or a tool object (called a proxy) mapped to the human finger or the tool grasped by the hand that is on the end effector of the haptic device. If this virtual object collides with other virtual objects or environments, then the response force and torque are computed in the haptic renderer and fed to the haptic controller to generate a stable and transparent response to be applied to the human operator as feedback. In the meantime, a tactile device can stimulate the user’s skin and tissue. As shown in Fig. 10(e)–(h), tactile devices have diverse forms. Various types of actuators can be used to create a tactile effect. For example, motor-based vibrators, piezo actuators, electrodes, pneumatic actuators, or even shape-memory alloys can be used. By adjusting the intensity and frequency of the actuators and combining them, the haptic content creator can make various tactile patterns. For instance, any type of tactile device with actuators in an array can be successfully supported by the MPEG-V standard. We have already shown that the tactile gloves shown in Fig. 10(f) work well with the MPEG-V standard. Israr et al. [40] presented a tactile surface display to be placed on the user’s back, and they also suggested a set of haptic effects to be treated as basic units for the construction of new haptic effects. Since its mechanism is similar to that of the tactile gloves, their tactile device is expected to be successfully supported by the MPEG-V standard. Moreover, the MPEG-V standard also supports the construction of complex, composite tactile patterns by combining basic patterns [42].

The task of the haptic renderer includes, for example, collision detection, dynamic simulation, application of physical constraints, processing of haptic texture, and computation of reaction force and torque. As shown in Fig. 9, haptic rendering is performed in four main stages: (1) reading of the position and orientation of the haptic
device; (2) checking whether the mapped virtual probe corresponding to the read position and orientation is in collision with the virtual environment; (3) computing the response force/torque if collision occurs; and (4) applying the computed force and torque to the haptic device. Information required for collision detection and computation of the reaction force/torque can be obtained from a 3D model (describing geometric properties) and its physical properties (stiffness, friction, damping, mass, etc.) assigned by haptic modeling tools [33–35]. Typically, by Hooke’s law, the response force is computed by multiplying the penetration depth and the stiffness of the given object.

In most haptic applications, polygon meshes are widely used to represent 3D virtual objects. In multimedia applications, however, this representation method may not be appropriate because of large amounts of data with limited network bandwidth, progressive transmission, and compression. In order to bridge the gap between image-based modeling and full 3D modeling, the use of depth image-based representation can be a good alternative. In order to deal with both 3D polygon data and 2.5D data obtained from depth image-based processing, a special depth image-based haptic rendering algorithm [12] may be used as a component of the proposed haptic broadcasting system.

The force/torque signal from the haptic renderer is the desired command force/torque to be produced through a force/torque-reflecting device such as a joystick, e.g., Sensable’s Phantom series [47], Forcedimension’s Omega series [48], and Novint’s Falcon [49] haptic device. Feeding the force/torque signal from the haptic renderer to the haptic controller has been widely accepted and has proved greatly successful in important post-processing tasks such as stabilization and smoothing of output before they are actually applied to the haptic device. The haptic controller, in Fig. 9, re-computes or shapes the response force and torque commands that satisfy transparency as much as possible while guaranteeing haptic interaction stability [43,44]. Since there is bidirectional energy exchange during haptic interaction, unstable or oscillatory behaviors deteriorate the immersiveness of interaction. Moreover, the user can even be physically injured due to instability of haptic interaction. Furthermore, the haptic device cannot provide infinitesimal impedance when the haptic probe moves in free space in the virtual environment, and it also cannot provide infinite impedance when the probe collides with a very stiff (rigid) object; it can cover only some extent of the impedance range [45]. Hence, transparency is a critical and valuable property of a haptic interaction system. This is the reason why maximizing transparency while guaranteeing stability is very important, and the haptic controller is in charge of doing this.

One of the implemented haptic broadcasting systems is shown in Figs. 11 and 12. In Fig. 11, a viewer wears tactile gloves that consist of many vibrotactile actuators, and passively feels a sense of touch as the actor on the screen moves his arms and hands around. In Fig. 12, a viewer can see, hear, and touch the content while considering purchasing the product on the screen.

4. Open issues and research challenges

Even though this paper has proposed a conceptual way to develop a successful haptic media system and has
presented sample implementations, a lot of future work needs to be carried out to realize daily-life applications. First, (semi-)automatic generation of haptic contents from existing audio-visual data seems to be very necessary for easy and economical production because manual authoring (e.g., as discussed in [14]) is very time-consuming and tedious. Fully autonomous generation is, however, not possible and even undesirable, because the creation of the most dramatic haptic sensation needs to be directed by the movie director together with the audio-visual contents. Some methods for intelligent content analysis are therefore necessary for haptic effect generation. Moreover, easy-to-use haptic modeling and editing tools (as discussed in [34]) need to be improved. At the same time, there are many research issues regarding methods of transmitting virtual objects together with audio-visual data. For example, an efficient real-time transmission (or compression) method could be investigated for a virtual object with a very fine geometry and haptic data. The virtual objects thus sent may be independently rendered for active haptic interaction, as in the case of the home shopping scenario shown in Fig. 11. Human factor studies on haptic sensation may also be required for enjoyable and entertaining content, because the haptic sensation is not objective but subjective. Some haptic sensations may bother people or may not generate the intended amusement. It is uneasy to establish a sound quality metric for the sensory experience in short order. In that respect, assessing the quality of sensory experience for haptic content is a challenging task. A different valuation basis may be needed since haptic content has different requirements in comparison with traditional audio-visual media, due to the bidirectional nature in haptic interaction. More research will be carried out by developing quality metric and evaluating the proposed framework. In general, building correlation mapping between sensory (including haptic) effects through the proposed framework and user experiences can provide a common tendency and it can be used as quality model for the proposed framework. Concretely, for instance, evaluating the proposed framework may include accuracy measurement on device output (e.g., feedback force or tactile patterns) compared with sensory information considering user’s sensory preference and device capability in various conditions. Moreover, synchronization of the proposed framework can be evaluated by presenting various sensory effects simultaneously. Then, user perception of asynchrony can be rated. Finally, effectiveness of sensory device capability will be investigated through user evaluation when the same sensory information is transmitted to the adaptation engine. These research challenges demand inter- and trans-disciplinary collaboration, not only with haptic/visual/audio engineers but also with content creators, directors, psychologists, educators, and so on.

5. Conclusion

In this paper, we have briefly explained the MPEG-V standard, which includes defining standardized interfaces between virtual worlds and the real world, and have comprehensively illustrated how to construct a haptic broadcasting system based on the MPEG-V standard stage by stage. A haptic broadcasting system using the MPEG-V standard can provide multimodal interactions; synthesizing and incorporating haptic feedback with conventional audio-visual content can provide the viewer with a diverse, rich, and immersive experience. Even though the concept and framework of the presented haptic broadcasting system may not be perfect, this paper provides a good starting point for the development of a haptic-enabled broadcasting system.

Since our proposed concept and framework are at an early stage of development, an in-depth study is required to examine its scope and advantages, and to discover potential practical and commercial applications of this emerging technology. In order to achieve this, there is a great need to train content creators such as producers, filmmakers, educators, and artists, and to develop intuitive and productive editing and authoring (modeling) tools to support and accelerate the widespread creation of haptic contents. More research is also needed to develop tools for properly evaluating the proposed framework and assessing user experience.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0009234).