Beginning of a New Standard: Internet of Media Things

Sang-Kyun Kim[1], Nevadita Sahu[2], Marius Preda[3]
[1] College of Convergence Software, Myongji University
34 Geobukgol-ro Seodaemun-gu Seoul, 03674, Korea
[e-mail: goldmunt@gmail.com]
[2] Computer Engineering Department, Myongji University
116 Myongji-ro Choin-gu Yongin, 17058, Korea
[e-mail: nevadita.sahu@gmail.com]
[3] INT Télécom Ecole de Management
Paris Area, France
[e-mail: marius.preda@it-sudparis.eu]
*Corresponding author: Sang-Kyun Kim

Received May 5, 2017; revised July 5, 2017; accepted July 24, 2017; published November 30, 2017

Abstract

Recently, Internet of Things (IoT) drives a large variety of research, development, and new type of markets. All type of devices and sensors will be part of the Internet of Things and will be able to communicate not only plain data, but also audio-visual, olfactory, and haptic media data. In addition, as the devices and sensors getting smarter, it is highly probable that they can process acquired media and metadata to extract higher level of information (e.g., semantics). To support such enhanced functionalities, ISO/IEC SC29 WG11 (MPEG) starts a new standard project, ISO/IEC 23093, called Internet of Media Things (IoMT) to provide standard data formats and APIs for media things. This paper presents the standardization activities of IoMT focusing on explaining terms, standard scopes, and major media things with their use cases. One of the use cases, an IoT system for a blind pedestrian navigation assistance, is evaluated to prove its effectiveness.

Keywords: Internet of Things, media things, Internet of Media Things, IoT standardization, blind pedestrian navigation

A preliminary version of this paper appeared in 3rd EEECS 2017, Jan 9-11, Okinawa, Japan. This version includes the detailed IoT standadization activities behind the proposed methods of a blind person assistant system. The experimental results of times, speeds, and subjective tests with the implemented system are presented as well. This work was supported by the the national standard technology improvement program (10053655, R&D and International Standardization of IoT cameras, sensors, and services) funded by Korean Agency for Technology and Standards (KATS), Ministry of Trade, Industry and Energy (MOTIE).

https://doi.org/10.3837/tiis.2017.11.001 **ISSN : 1976-7277**

1. Introduction

Since 2014, the MPEG International Standardization Group (ISO/IEC SC29 WG11) started a new Ad Hoc group activity called "Internet of Media Things" to define interfaces exchanged between media things and humans. This standard supports the interoperability of data formats between media things including sensors, actuators, media analyzers, and humans (e.g., system designers). Recently, a call for proposal for this new standard has been issued [1] and the standard group receives its first response from the market and industry on April of 2017. The group expects its final draft of International Standards (FDIS) on Jan. of 2018.

The rationales of IoMT standardization activities in MPEG are as follows:

- Existing media standards – MPEG has been successfully developed renowned media standards for compressing, streaming, storing, rendering, and describing media such as MPEG-1/2/4/H, MPEG-A, DASH and MMT, MPEG-7/21, and MPEG-V/U. These existing standards can be adopted and utilized in the IoT system,
- Future MPEG standards – MPEG is developing for future media coding (e.g., point clouds, plenoptics, future video/audio coding) and media-related systems/tools (e.g., CDVA, Big Media, Media Orchestration, 360 VR). These upcoming standards shall be adopted in the IoT environment,
- Limitation of IoT services – commercial IoT services available in the market support limited functionalities such as simple on/off type actuations and sensed data readings. When standard interfaces and data formats are provided for supporting media acquisition, exchanges, analysis, controls, and consumptions in IoT systems, there will be richer IoT media services,
- Cooperate with existing network protocols and standards – the new standard will be on top of any type of network protocols such as TCP/IP, UDP, RTP, MQTT, and CoAP. In addition, the standard will be liaised with other IoT standard groups (e.g., UPnP, ISO WG10, IEC TC100) to provide interfaces to connect existing IoT systems.

The preliminary report on media-centric IoT standardization activities was presented in [2]. A media-centric IoT camera system [3] was introduced with data exchange flows, which explain how to exchange media thing's characteristics, to connect, and to perform the given mission. A brief introduction of IoMT standardization along with its relationship with surrealistic worlds and 3D scent technologies was presented in [4].

The interface standardization for sensors and actuators bridging between the virtual and real world has been standardized in ISO/IEC 23005 (MPEG-V) [5]. The content authoring for sensorial effects was researched in [6][7][8][9], and the interaction between the real and virtual worlds using sensor data was presented in [10]. The characteristics and data formats for sensors and actuators are presented in ISO/IEC 23005-2 and ISO/IEC 23005-5, respectively. The data formats for actuation commands are presented in ISO/IEC 23005-5. The sensors and actuators specified in MPEG-V will be utilized to describe the media things and extended to fulfil the use cases of IoMT.

There are studies about media content sharing between Device-to-Device (D2D) [11][12] and Base-station-to-Device (B2D) [13]. In [11][12], they propose the efficient algorithms and models to match between potential content providers and content demanders. In [13], a practical content sharing system was proposed to share and stream media content recorded in a

home media server to a personal mobile device. Both studies, however, did not show the standard formats and APIs to connect and exchange media data between devices.

In this paper, we present the standardization activities of IoMT focusing on explaining terms, standard scopes, and major media things with their use cases. This paper explains what kind of essential standard data shall be exchanged between media things. This paper also contributes to explain how the multiple media things can be composed in the IoT environment to provide enhanced media services in near future. In addition, one of the use cases, the blind person assistant system using Internet of Media Things [14][15][16], is tested to prove its effectiveness.

This paper is organized as follows. The terms and definitions of IoMT is described in section 2. Section 3 presents the standardization scopes of IoMT. Section 4 exemplifies major media things followed by their use cases in section 5. Section 6 shows experimental results of the blind person assistant system. Finally, the conclusion is presented in section 7.

2. TERMS AND DEFINITIONS [17]

The Internet of Media Things (IoMT) is the collection of interfaces, protocols and associated media-related information representations that enable advanced services and applications based on human to device and device to device interaction, in physical and virtual environments. Information refers to data sensed and processed by a device, and/or communicated to a human or another device. The major terms for IoMT include:

- Audio: anything related to sound in terms of receiving, transmitting or reproducing or its specific frequency,
- Camera: a special form of an image capture device that senses and captures photo-optical signals,
- Display: the visual representation of the output of an electronic device; the portion of an electronic device that shows this representation, as a screen, lens, or reticle,
- Gesture: a movement or position of the hand, arm, body, head, or face that is expressive of an idea, opinion, emotion, etc.,
- Haptic: an input or output device that senses the body's movements by means of physical contact with the user.
- Image capture device: a device which is capable of sensing and capturing acoustic, electrical or photo-optical signals of a physical entity that can be converted into an image,
- Internet of Media Things (IoMT): a special subset of IoT where information resources are limited to media.
- IoMT Device: an IoT Device that contains MThings,
- IoMT System (MSystem): an IoT system whose main functionality is related to media processing,
- Loudspeaker: an electroacoustic device, that is connected as a component in an audio system, generating audible acoustic wave,
- Media: data that can be rendered, including audio, video, text, graphics, images, haptic and tactile information; these data can be timed or non-timed,
- Media Thing (MThing): a Thing capable of sensing, acquiring, actuating, or processing of media or metadata,
- Microphone: an entity capable of capture and transform acoustic waves into changes in electric currents or voltage, used in recording or transmitting sound,

- Media Wearable (MWearable): an MThing intended to be located near, on or in an organism,
- Motion: the action or process of moving or of changing place or position; movement,
- Natural User Interface (NUI): a system for human-computer interaction that the user operates through intuitive actions related to natural, everyday human behavior,
- Presentation: an act of producing human recognizable output of rendered media.

3. STANDARDIZATION SCOPE OF IoMT [18]

The scope of the **IoMT** is to standardize a set of interfaces, protocols and associated media-related information representations related to:

- User commands (setup info.) between a system manager and a MThing, cf. Interface 1 in **Fig. 1**,
- User commands (Setup info.) forwarded by an MThing to another MThing, possibly in a modified form (e.g., subset of 1), cf. Interface 1' in **Fig. 1**,
- Sensed data (Raw or processed data) (compressed or semantic extraction) and actuation information, cf. Interface 2 in **Fig. 1**,
- Wrapped interface 2 (e.g. for transmission), cf. Interface 2' in **Fig. 1**,
- MThing characteristics, discovery, cf. Interface 3 in **Fig. 1.**

(2)

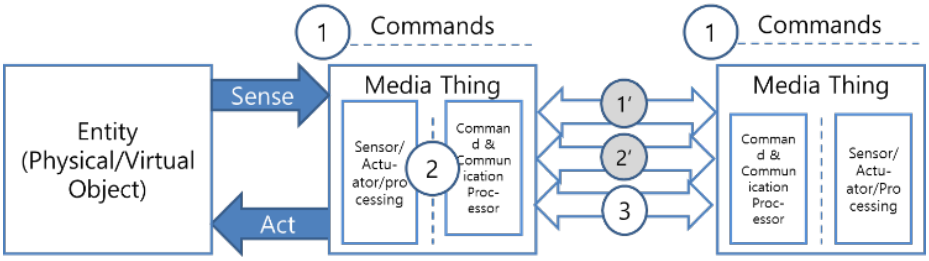

Fig. 1. Conceptual architecture of IoMT standardization

Each media thing can contain sub-modules like sensors and/or actuators, and/or media processing modules as well as command & communication processing modules. The sensors can collect media data about entities from either physical or virtual world. The actuators can affect the status of entities. The media processing modules can analyze media data collected from sensors and extract some meaningful information to decide what to do. The command & communication processing modules can interpret commands or communication messages from other media things.

As devices, IoTs have strong constraints with respect to the processing, storage and communication capabilities. This implies the careful design of algorithms at all levels: an IoMT camera should be able to analyze the captured video partly in order to transmit only meaningful (and probably compressed) information to the environment. Probably the most important aspect of IoMT (and IoTs in general) is that they are part of larger systems composed by a multitude of similar or various IoTs. Such a modular approach allows a big variety of services and applications; some systems may be of large scale (e.g. a city video-surveillance system) or of small scale (e.g. a person bio-signals recorder system).

4. EXAMPLE OF MTHINGS [1]

MThings may be classified based on their main functionalities. This section provides description of (some of) particular MThings and their functionalities, which should be interpreted as particular instantiation of the requirements introduced in [17].

4.1 IoMT Digital camera

IoMT Digital camera is an IoMT device of which the main functionality is that of a camera. Additional functionalities of the IoMT Digital camera may include compression, storage, transmission or streaming of acquired visual data. Some IoMT Digital camera may be able to manage audio data.

More advanced IoMT Digital camera may have functionalities such as transmitting or streaming intermediate visual related descriptors.

IoMT Digital camera may receive information to control its settings and be able to generate metadata related to capabilities and characteristics, as well as be able to provide actuation commands to control other things.

IoMT Digital camera may be able to process the acquired visual data locally, able to decide actions to take, able to receive and interpret task descriptions, able to process multimedia (audio, video, image, etc.).

IoMT Digital camera may be able to recognize and interpret gestures, to extract contours and regions, to provide low-level spatial and temporal A/V descriptions. The IoMT camera may also be able to recognize and interpret semantic descriptors related to events (e.g., fire, collision, fight, intrusion), to provide metadata related to events, and to detect entities of interest (e.g., person, car, logos) and provide descriptors for the detected entities. Finally, the IoMT camera may be able to compare the descriptors from the detected entities with the descriptors in the local storage and to generate semantic metadata of the detected entities (e.g., genders, ages, car types, license plates).

4.2 IoMT Smart microphone

IoMT Smart Microphone is an IoMT device of which the main functionality is that of a microphone. Additional functionalities of the IoMT Smart Microphone may include compression, storage, transmission or streaming of acquired acoustic data. Some IoMT Smart Microphone may be able to manage audio data.

4.3. Media Processing Unit

Media Processing Unit is an IoMT device of which the main functionality is processing of sensed or acquired media or metadata (e.g., audio/video/graphic/text/sensor/metadata).

Media Processing Unit may have additional functionalities such as recognize and interpret gestures, recognize and/or interpret sound, extract contours and regions, provide spatial and temporal A/V descriptions, recognize and interpret semantic descriptors related to particular events (e.g., fire, collision, fight, intrusion, etc.), to provide metadata related to events, to detect objects of interest (e.g., person, car, logos) and provide descriptors for them, to compare the descriptors from the detected objects with the descriptors in the local storage, to generate semantic metadata of the detected objects (e.g., genders, ages, car types, license plates), to synthesize sounds, to translate languages, generate natural language responses.

4.4. Smart glasses

Smart glasses are MWearables of which the main functionality is that of providing a multisensory interface between the user and the physical/virtual world.

Smart glasses are a range of functionalities including those associated with particular devices such as camera, microphone, media processing, display and loudspeaker.

In addition, smart glasses should be able to provide Natural User Interface (NUI) (e.g., hand gesture, head motion, body gesture, voice, marker, eye tracking), and to provide NUI in a combined way (multimodal) as well as in an individual way (e.g., hand gestures or voices). The smart glasses may also be able to provide NUI for controlling the smart glass and sensors/actuators equipped in the smart glass. The smart glasses may also be able to transmit the user description (e.g., gender, nationality, age) of the smart glass user to a processing unit and to provide NUI to control processing (e.g., image analysis) in the smart glass.

5. MAJOR USE CASES [18]

Use cases for IoMT can be enumerable because any combination of media things are feasible to support a new type of IoT media services. In this section, major use cases of IoMT are explained with the categories of smart spaces: monitoring, smart spaces: navigation, and smart environment.

5.1. Smart spaces: Monitoring and control with network of audio-video cameras

5.1.1. Face recognition to evoke sensorial actuations

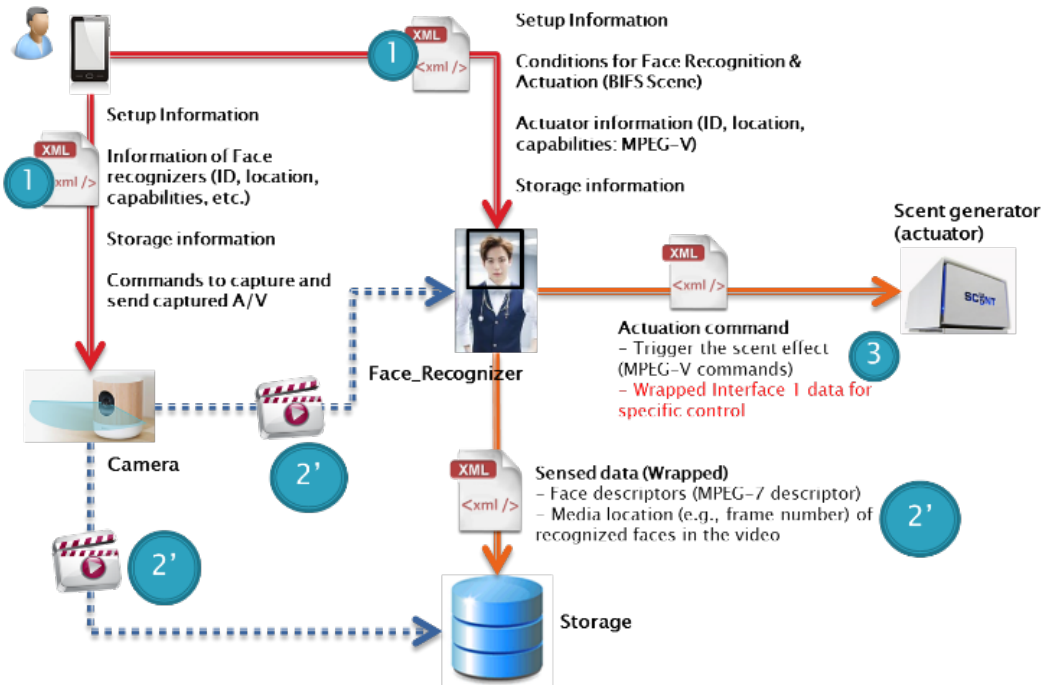

Fig. 2. Data flow schematics evoking sensorial actuations using the face detection

An IP surveillance camera (IoMT camera) captures A/V data and sends them to both a storage (IoMT Storage) and a face recognizer unit (Media Processing Unit). When the face recognizer detects and recognizes the face of a pre-registered person, it activates a scent generator to spray some specific scent. The specific descriptors (e.g., detected face locations, face descriptors, media locations of detected moments) can be extracted and sent to a storage (**Fig. 2**). In this use case, the scent generator can be replaced by any type of actuators (e.g., light bulbs, displays, music players). The user (e.g., a system designer) can setup either all the MThing in the system (i.e., in a centralized manner) or only an IoMT camera that can exchange necessary information with other MThings to achieve the mission (i.e., in a distributed manner). The numbers in the figs (from this point) denote the interface number described in **Fig. 1**.

5.1.2. Human tracking with multiple network cameras

Because urban growth is today accompanied by an increase in crimes (e.g., theft, vandalism), many towns are willing to put in place video surveillance systems that would help reduce urban crimes. Such a city video surveillance system is an IoMT system that includes a set of IP surveillance cameras (IoMT camera), a storage (IoMT Storage) and a human tracker unit (Media Processing Unit) (**Fig. 3**).

A particular IP surveillance camera captures A/V data and sends them to both the storage and the human tracker unit. When the human tracker detects a person in the visible area, it traces the person and extracts the moving trajectory.

If the person gets out of the visual scope of the first IP camera but stays in the area protected by the city video surveillance system, another IP camera from this system in the vicinity takes over the control and keeps capturing A/V data of the corresponding person.

If the person gets out of the area protected by the city video surveillance system, for example the person enters into a commercial center, then the city system search if this commercial center is also equipped with a video surveillance system. If this is the case, the city video surveillance system sets up a communication with the commercial center video surveillance system in order to allow another IP camera from the commercial center video surveillance center to keep capturing A/V data of the corresponding person. In this case, the interfaces between the city video surveillance system and the commercial center video surveillance system must be standardized to communicate and exchange data interoperable.

In both cases, the specific descriptors (e.g., moving trajectory information, appearance information, media locations of detected moments) can be extracted and sent to the storage.

In this use case, the human tracker module can control the activation and deactivation of cameras in the area.

The user (e.g., a system designer) can setup either all the MThings in a particular video surveillance system (i.e., in a centralized manner) or only an IoMT camera that can exchange necessary information with other MThings to achieve the mission (i.e., in a distributed manner).

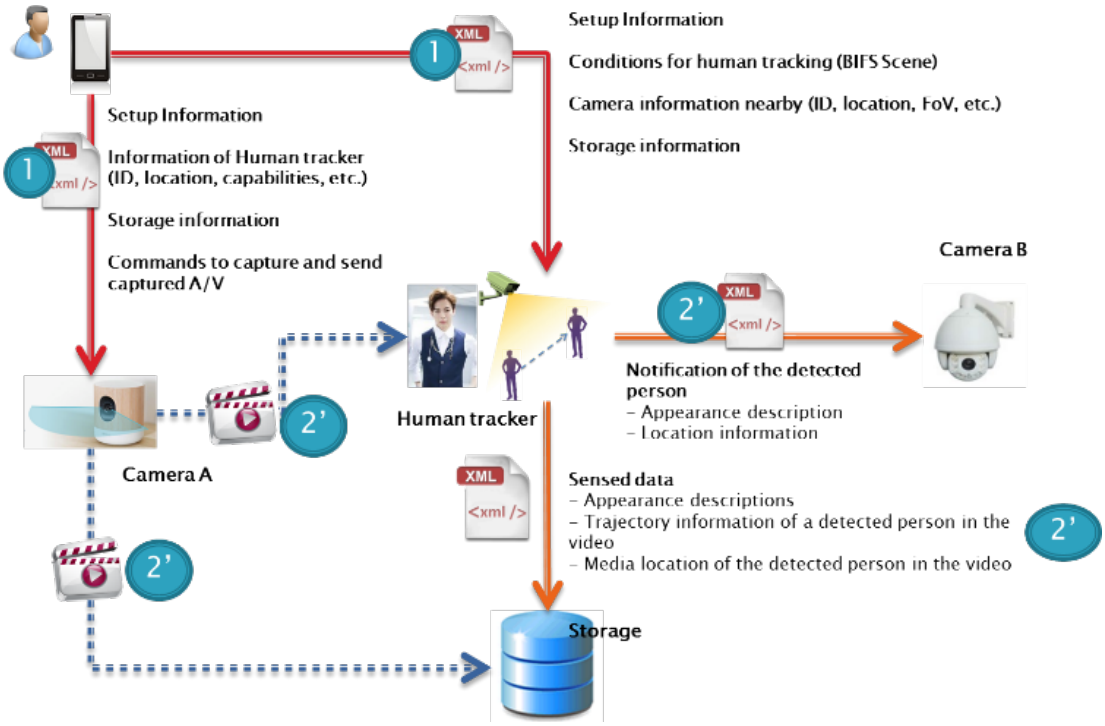

Fig. 3. Data flow schematics of a city video surveillance system

5.2. Smart spaces: Navigation

5.2.1. Collision warning

Fig. 4. Blind person assistant system to avoid obstacles

Assume that a blind pedestrian tries to avoid any obstacles in front (**Fig. 4**). A blind person carries a smart cane, a vibration band, a smart phone, and a networked headphone. The smart cane equipped with distance sensors (e.g., an ultrasonic sensor, an infrared sensor) can

measure the distance between the cane and obstacles in front. A collision coordinator (Media Processing Unit) receives the distance data and decides what actions to take. If the distance is reasonably far, an alarming text data of the corresponding distance (e.g., “5 meters before colliding obstacles ahead.”) is produced by the collision coordinator and sent to a Text-to-Speech generating unit (Media Processing Unit). The Text-to-Speech generator creates the corresponding audio file and sends its URL to a networked headphone. The headphone plays the corresponding audio files to the blind person. If the distance is close enough, the collision coordinator activates either a wristband to vibrate or the headphone to create beeping sounds (**Fig. 5**) [5].

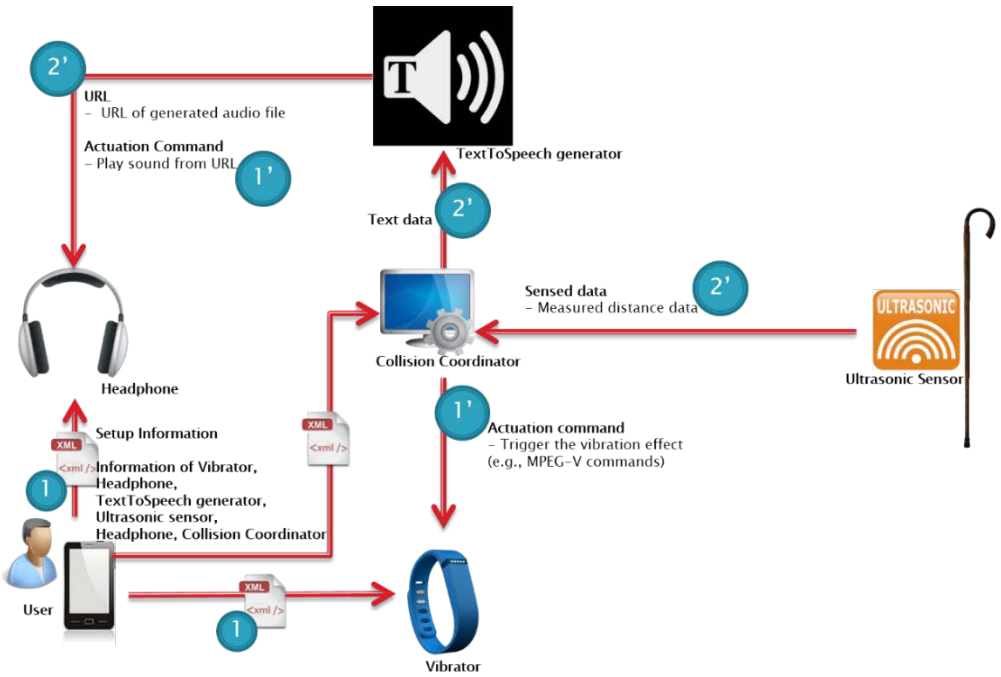

Fig. 5. Data flow schematics of a blind pedestrian collision aviodance system

5.2.2. Guiding direction

Assume that a blind pedestrian travels to a destination (**Fig. 6**). The global navigation can be provided by any web service. However, RFID tags (or beacons) that contain the exact location coordination can enhance the local navigation. The exact location of the RFIDs (or beacons) can be registered and retrieved from the server using their unique IDs. The RFID tags, therefore, can be embedded in every corners of the streets. The blind person carries a smart cane, a smart phone, and a networked headphone. The smart cane is equipped with a RFID reader, some inertia sensors (e.g., a gyro, a compass). The RFID reader can read the RFID tags embedded in every street corners. A direction guider (Media Processing Unit) receives the RFID tag data (or beacon data) and retrieves the current location of the blind person. Combining with the other inertia information, the direction guider creates directional guidance (e.g., “Turn left”, “Turn left a little more”, “OK, go straight”) and sends it to a Text-to-Speech generating unit (Media Processing Unit). The Text-to-Speech generator creates the corresponding audio file and sends its URL to a networked headphone. The headphone plays the corresponding audio files to the blind person (**Fig. 7**) [14][15][16].

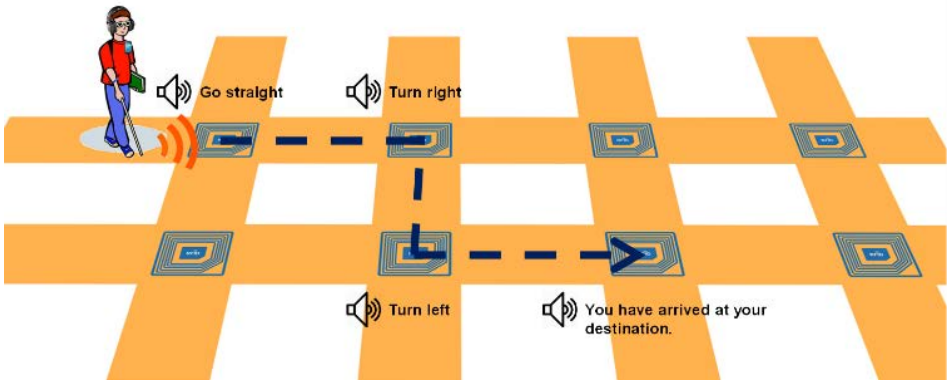

Fig. 6. Usage scenario of a blind pedestrian navigation system

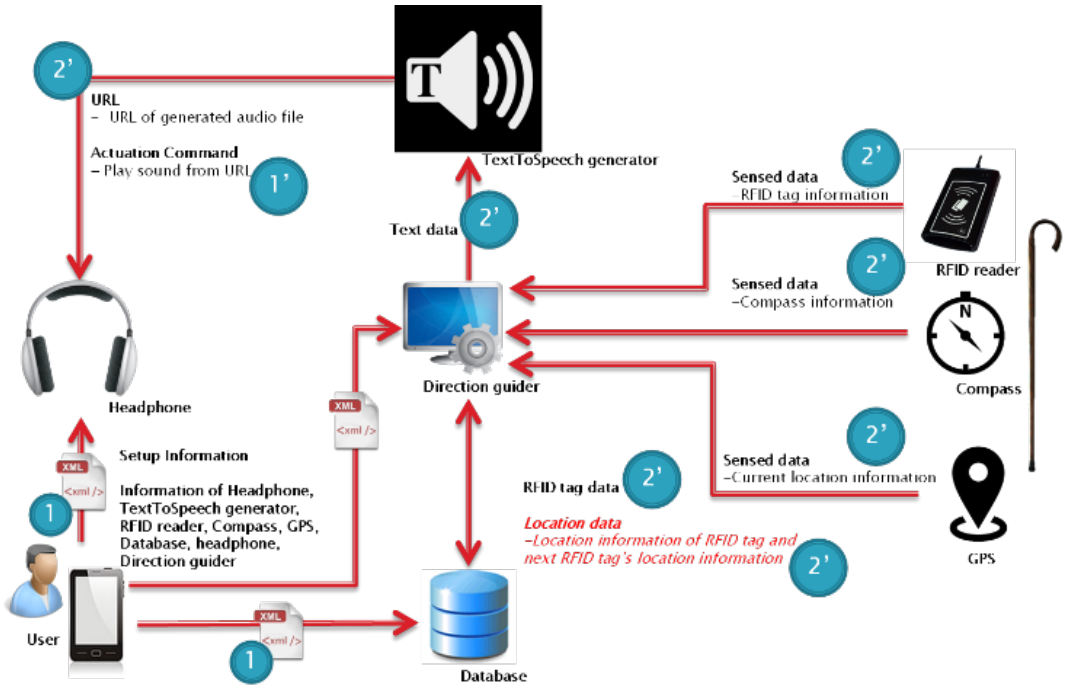

Fig. 7. Data flow schematics of a blind pedestrian navigation system

5.2.3. Personalized tourist navigation with natural language functionalities

Natural language is a very convenient interface between a human user and the computer. The user can control the machine with a speech command, find information, and ask questions to the intelligent agent inside the smart-phone or a server. With the wearable devices and the natural language communication, the user can have the freedom to participate in an activity without using a keyboard or any other terminal devices.

Wearable devices with natural language interface can provide accessibility functionalities for people with disabilities. Smart watch/glasses can guide people with low or no vision for the direction using location, visual information and speech interface. Smart devices can provide sound information to the people with hearing problems using vibration of the watch or some

light signals from the smart-glasses when some emergency situation occurs. People of low intelligence or reading problems can also benefit from wearable devices through reading software embedded in or connected to the devices.

Speech translation for people of different languages is a very convenient service in the multi-cultural, multi-lingual society and in a global environment. Evolving from being delivered on a PC through laptop and tablets to smart-phone, speech translation systems are getting more usable with wearable devices. When a user speaks to the microphone embedded in the smart watch or headphone in one language in a conversation with another user with different languages, it will be translated to the target language.

The result of the translation can be heard by the user of the target language through the wearable device. The translation engine is either in the remote server (remote translation system) or in the smart-phone (stand-alone translation system) which is connected to the wearable device. With the wearable translation service, the user is able to use his hands freely while the conversation is translated. The wearable device is also used for finding someone automatically who can speak one of the languages which the embedded translation system handles in a travelling situation (**Fig. 8**).

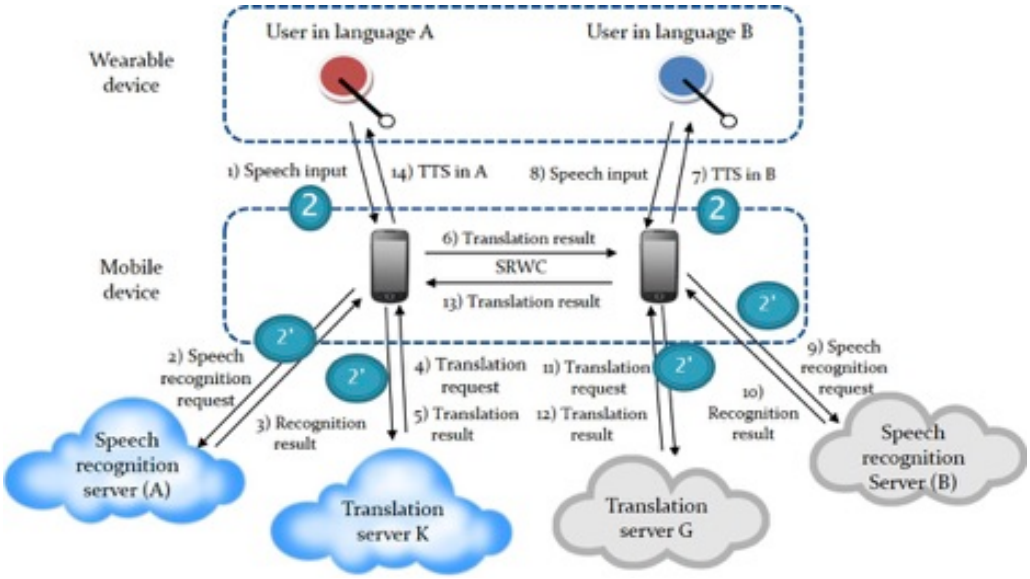

Fig. 8. Data flow schematics of a personalized tourist navigation system

5.3. Smart environments in smart cities

5.3.1. Smart factory: car maintenance assistant system using smart glasses

Fig. 9 illustrates the use case of smart glasses for the car maintenance system. It is assumed that a technician wearing smart glasses is working on maintenance of a car. The smart glasses automatically provide a list of maintenance manuals related to a specific part to be checked on the display, then a user select and read the necessary manual by using hand gestures. In this way, a user can perform maintenance works while using both hands freely.

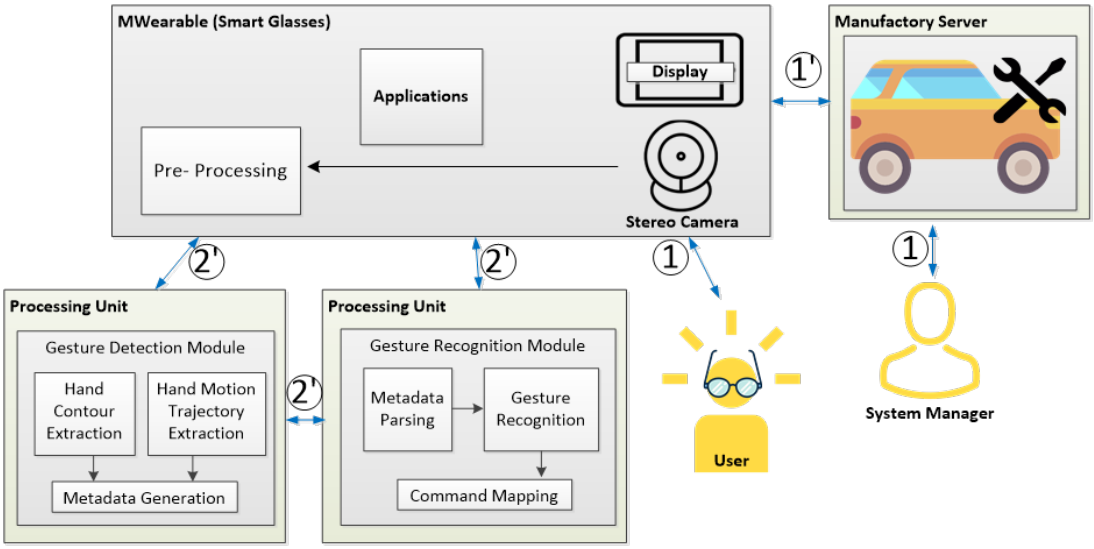

Fig. 9. Data flow schematics of a car maintenance assistant system

5.3.2. Smart museum: Augmented education using smart glasses

Fig. 10 illustrates a use case of augmented education with smart glasses in a museum in which augmented information such as a narrative explanation about a modern work of art and a video clip showing the painter's interview can be provided according to a user's request invoked by hand gesture. In this way, a user enjoys the museum tour with rich information presented by smart glasses without any guide brochure and/or the help of guiders.

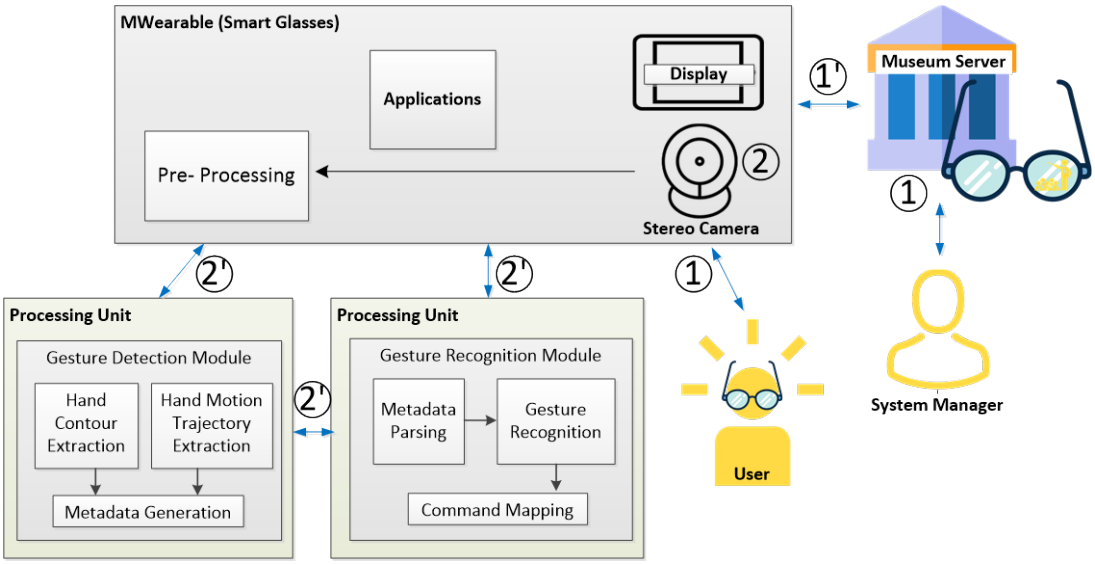

Fig. 10. Data flow schematics of an augmented reality tour in a museum

6. Experimental results

An IoMT system for the blind pedestrian use cases [14][15][16] presented in Section 5.2.1 and 5.2.2 was implemented and tested to prove its effectiveness in this section.

6.1. Experimental Settings and Procedure

For testing the navigation system different routes are set in the campus by putting the beacons on the wall and few obstacles on the way. Distance from source to destination point is set as 25 meters. The blind user has to hold the Smart Cane, which contains the ultrasonic sensor and an android application is deployed in the user's smartphone, which will guide the user. **Fig. 11** shows the experimental setup for a blind folded person holding the smart cane and listening to the instructions with the headphone.

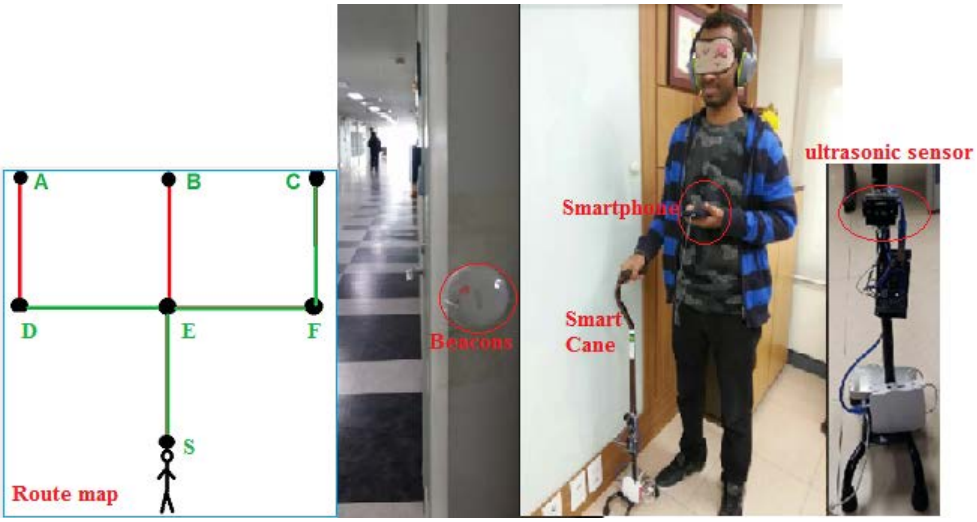

Fig. 11. Experimental setup

The experiment is performed by blindfolding 75 undergraduate engineering students with the age between 20 and 30 years old. The test subjects are composed of 49 males and 34 females, who are new to the location. Each subject should reach his/her desired destination by the support of navigation system by avoiding obstacles in between. A few students took the test in both ways with and without the navigation support. The navigation times are recorded.

The evaluation process of the IoMT navigation system is based on three criteria as follows: (1) blind person's safety navigation through real time obstacle detections, (2) the navigation system accuracy and usability by measuring a mean opinion score of evaluation questions, (3) the navigation system efficiency by calculating the average walking speed.

In order to meet these criteria, the experimental procedure is as follows:

(1) All the participants are blindfolded and navigate to the destination using the Smart Cane and the smartphone, by listening to the incoming direction messages,

(2) For each of these participants, the navigation time is recorded. Some of the participants are asked to navigate without the navigation support and the navigation times are recorded as well.

After finishing the experiment, each participant assesses the navigation system with the evaluation questions.

Considering different ways to evaluate the navigation system, we came up with ten evaluation questions. Through these questions, we tried to capture the participant's quality of experience regarding the system.

The evaluation questions presented to the participants are as follows:

(Q.1) From the scale of 1 to 5, where 1 is "hard" and 5 refers to "easy", how do you feel about the system usability?

(Q.2) From the scale of 1 to 5, where 1 refers “incomprehensible” and 5 is “clear”, how understandable is the message delivered by the system?
(Q.3) From the scale of 1 to 5, where 1 is “useless” and 5 corresponds to “very useful”, how useful is the information provided by the system?
(Q.4) From the scale of 1 to 5, where 1 corresponds to “delayed” and 5 is “immediate”, how fast is the delivery of message from the system?
(Q.5) From the scale of 1 to 5, where 1 is “very imprecise” and 5 is “exact”, how precise was the contextual information provided by the system?
(Q.6) From the scale of 1 to 5, where 1 corresponds to “difficult” and 5 corresponds to “very easy”, how clear is the user interface of the application?
(Q.7) From the scale of 1 to 5, where 1 corresponds to “hard” and 5 corresponds to “comfortable”, how comfortable is choosing the destination through voice command?
(Q.8) From the scale of 1 to 5, where 1 corresponds to “imprecise” and 5 corresponds to “exact”, how precisely the system can lead to destination?
(Q.9) From the scale of 1 to 5, where 1 corresponds to “erroneous” and 5 corresponds to “very accurate”, how accurate is the voice navigation?
(Q.10) From the scale of 1 to 5, where 1 corresponds to “ineffective” and 5 corresponds to “very effective”, effectiveness of the application compared to not using the application for a visually impaired person?

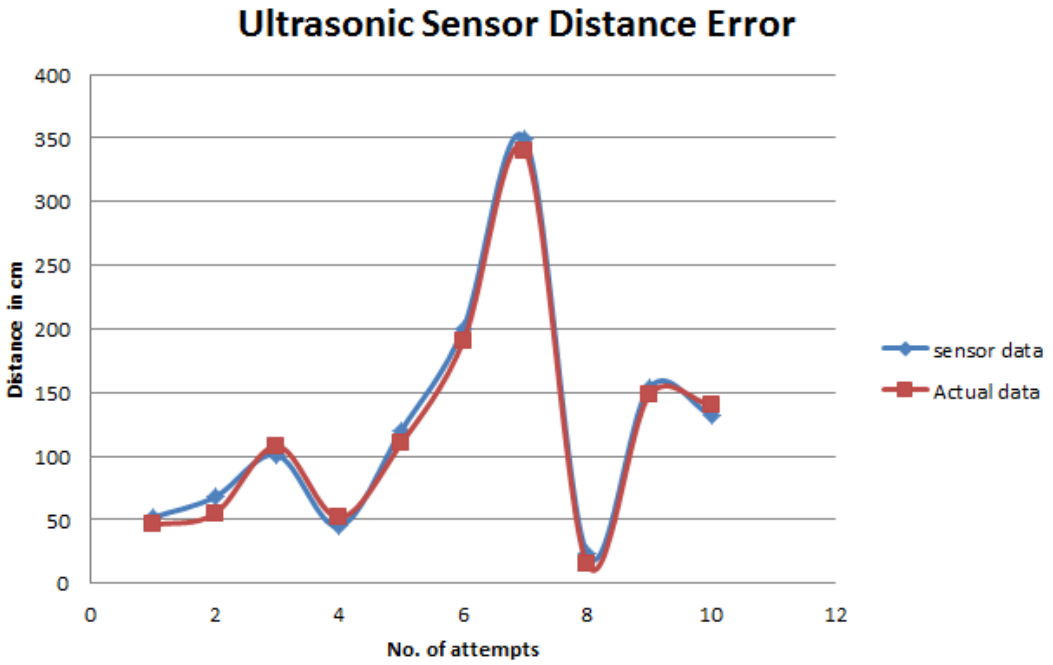

Fig. 12. Distance error measured with an ultrasonic sensor

While the blind folded users were moving towards the obstacles, the distance is measured by the ultrasonic sensor. The actual distance shall be measured and compared to the sensed distance in order to check the accuracy of the system [19]. This process is repeated for the test subjects. The distance errors between actual distance and the sensed distance are shown in **Fig. 12.** The distances to the obstacles measured by the ultrasonic sensor are depicted in diamond

dots with blue lines, and the actual distances are depicted in rectangular dots with red lines. The distances were measured 10 times for difference distant cases. As shown in the figure, the distance error falls in the range of 10cm, which is insubstantial, so that the ultrasonic sensor is effective to detect the obstacle.

When the user faces any obstacle, the messages do not intend to tell what movement the user should act. The system tries to provide only contextual information, which will help the users to make their own decision about which way to go, whether to proceed further or stop at that point, etc. The alert message, for example, is like "An obstacle is detected one meter ahead. Please be careful and walk slowly". This message is re-configurable and it can be changed as per the user's preferences. The warning distance of the system is also re-configurable. In the experimental scenarios, the warning distance is set to one meter. Other parameters like a time delay between two messages is set to be two seconds, which can also be corrected as per user's convenience.

The scores of evaluation questions mentioned above are analyzed as shown in **Table 1**.

Table 1. Statistics of quality assessment for the blind pedestrian navigation system

No. of participants	Min. score	Max. score	Mean score
75	27	50	39.16

The average score for the navigation system is 39.16 out of total score 50. The minimum score is 27 and the maximum score is 50. The mean score 39.16 is higher than the midpoint (i.e., 30) of the total score 50. This shows an acceptable opinion about the navigation system.

Apart from this, the tests and retrieved results for each question are analyzed further. Based on this analysis, it is assumed that the minimum threshold acceptance score is 3, out of the score range 1 to 5. **Fig. 13** shows the average of scores given by the participants for each evaluation question.

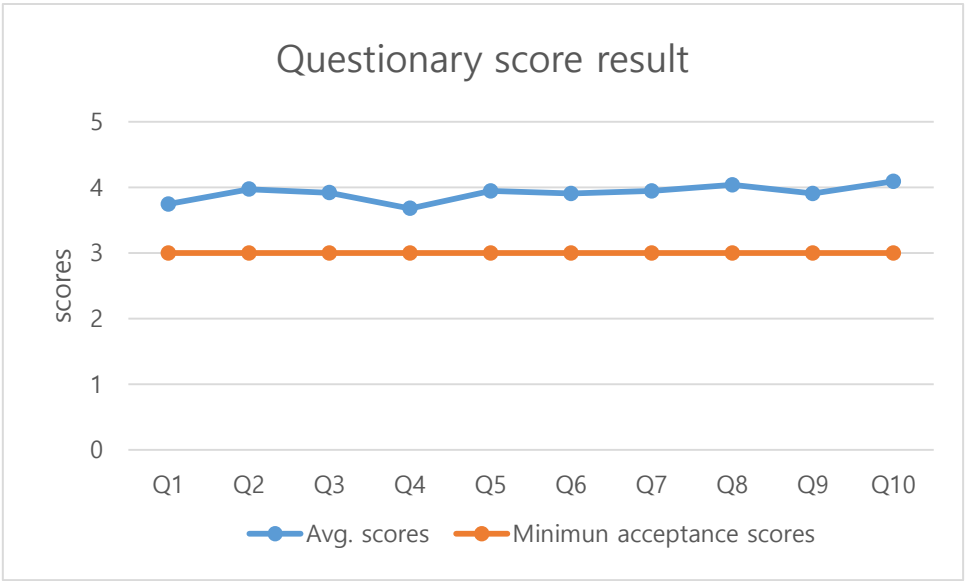

Fig. 13. Average score of each evaluation question

The result highlights a favorable evaluation, showing the scores above the minimal threshold acceptance level. Most of questions except question (Q.1) and question (Q.4) are evaluated more than score 3.9. The system usability (Q.1) seems a bit hard (3.74) normal

people who had no experience of the blind pedestrian navigation. The warning or direction messages seems a bit delayed (Q.4: 3.68) while the message seems clear (Q.2: 3.97) to people.

In case of question (Q.5:3.95), i.e., how precise was the contextual information provided by the system, there was a minor complaint about contextual direction. This was the situation, when the user goes out of the threshold angle, the message "you are out of the line. Turn slightly left/right" was given as a direction. The users are confused with the term "slightly" so that they are hesitate to make a next move immediately.

User interface of the system seems reasonable (Q.6: 3.91), and the system leads subjects to the destination well (Q.8: 4.0). The subjects satisfy with the voice navigation system (Q.9: 3.91) and feel the system useful to a visually impaired person (Q.10: 4.1).

There was a minor comment on question (Q.7: 3.95), i.e., how comfortable is choosing the destination through voice command that the google voice recognition API is not very efficient in capturing the accent of each person. For example, when the user speaks "d" as destination, the application recognizes it as "di" or "dee". In this case, the system promptly asks the subject to try again for entering the destination.

Table 2 summarizes the average walking speeds with and without the navigation support. As the users receive the direction messages, e.g., his/her movement angles and the obstacle alerts, they are able to navigate with a good walking speed. The navigation system helps the blind folded people reach their destination in a safer and faster way.

Table 2. Walking speed comparison

Type	**Distance** (meter)	**Ave. time** (seconds)	**Walking speed**
With navigation support	25	123.52	0.20m/s
Without navigation support	25	357.14	0.07m/s

7. Conclusion

In this paper, the standardization activities of IoMT was presented which are focusing on explaining terms, standard scopes, and major media things with their use cases. The experimental results of one of the IoMT use cases, the blind pedestrian navigation system, were presented to show the feasibility of the IoMT application using the IoMT standards. In the future, standardized data formats and APIs of IoMT along with actual system implement results with statistical analysis will be reported as the standardization activities progress.

References

[1] ISO/IEC JTC1 SC29 WG11 N16535, "Call for Proposals on Internet of Media Things and Wearables," *116th MPEG Chengdu meeting*, Oct. 2016.
[2] S.-K. Kim, "Standardization of Media-centric Internet of Things," *IWAIT 2016*, Jan. 2016.
[3] S.-K. Kim, "Media-centric Internet of Things Camera System," in *Proc. of The 4th International Conference on Smart Media and Application*, vol. 4(1), pp. 133-136, Jan. 2016.
[4] S.-K. Kim, "Internet of media things towards surrealistic worlds with 3D scent technology," *3rd World congress on Digital Olfaction*, vol. 4, pp. 10, Dec. 2016.
[5] Kyoungro Yoon, Sang-Kyun Kim, Jae Joon Han, Seungju Han, Marius Preda, "MPEG-V: Bridging the virtual and real world," *Elsevier*, ISBN: 978-0-12-420140-8.
[6] Yong-Soo Joo, Sang-Kyun Kim, "Sensory Effect Authoring Tool for Sensible Media," *Journal of Broadcast Engineering*, Vol. 16, No. 5, pp. 693-893, Sept. 2011 (in Korean).

Article (CrossRef Link).
[7] Sang-Kyun Kim, "Authoring Multisensorial Content," *Signal Processing: Image Communication*, vol. 28, Issue 2, pp. 162-167, Feb. 2013. Article (CrossRef Link).
[8] Sang-Kyun Kim, Yong-Soo Joo, YongMi Lee, "Sensible Media Simulation in an Automobile Application and Human Responses to Sensory Effects," *ETRI Journal*, Vol. 35, No. 6, pp. 1001-1010, Dec. 2013. Article (CrossRef Link).
[9] Sang-Kyun Kim, Seung-Jun Yang, Chung Hyun AHN, Yong Soo Joo, "Sensorial Information Extraction and Mapping to Generate Temperature Sensory Effects," *ETRI Journal*, Vol. 36, No. 2, pp. 224-231, Apr. 2014. Article (CrossRef Link).
[10] Sang-Kyun Kim, Jae Joon Han, Seungju Han, Yong Soo Joo, "Virtual world control system using sensed information and adaptation engine," *Signal Processing: Image Communication*, vol. 28, Issue 2, pp. 87-96, Feb. 2013. Article (CrossRef Link).
[11] Dan Wu, Liang Zhou, Yueming Cai, "Social-Aware Rate Based Content Sharing Mode Selection for D2D Content Sharing Scenarios," *IEEE Transactions on Multimedia*, vol. PP, Issue 99, 03 May 2017.
[12] "Mobile Device-to-Device Video Distribution: Theory and Application," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 12, no.3, pp. 1253-1271, 2015.
[13] Sang-Kyun Kim, Jinguk Jeong, Hyoung-Gook Kim, and Min Gyo Chung, "A Personal Videocasting System with Intelligent TV Browsing for a Practical Video Application Environment," *ETRI Journal*, vol.31, no.1, pp.10-20, Feb. 2009. Article (CrossRef Link).
[14] Nevadita Sahu, "A Navigation System for Visually Challenged People Implementing IoT (Internet of Things)," Maters degree thesis.
[15] Nevadita Sahu, Jonghoon Chun, and Sang-Kyun Kim, "EVALUATION OF IOT BASED NAVIGATION SYSTEM FOR VISUALLY IMPAIRED," *3rd EEECS 2017*, Jan. 2017.
[16] Nevadita Sahu, Min Hyuk Jeong, Jonghoon Chun, and Sang-Kyun Kim, "A Vision Disabled-Aid using the Context of Internet of Things," *Journal of broadcast engineering*, Vol. 22(1), pp. 78-86, Jan. 2017.
[17] ISO/IEC JTC1 SC29 WG11 N16534, "Requirements for Internet of Media Things and Wearables," *116th MPEG Chengdu meeting*, Oct. 2016.
[18] ISO/IEC JTC1 SC29 WG11 N16533, "Use cases for Internet of Media Things and Wearables," *116th MPEG Chengdu meeting*, Oct. 2016.
[19] A. Singh, A. Thakur and A. Taparia, "Analysis of computer vision and sensor technologies to assist the visually impaired," in *Proc. of Green Computing and Internet of Things (ICGCIoT), International Conference on*, pp. 135-138, 2015. Article (CrossRef Link).

Sang-Kyun Kim

- 1997: Computer Science in U of Iowa, BS(1991), MS(1995), PhD
- 1997.03. ~ 2007.02.: Multimedia Lab in Samsung Advanced Institute of Technology
- 2007.03. ~ 2016.02.: Professor of Computer Engineering in Myongji University
- 2017.03. ~ Current : Department of Convergent Software, Myongji University
- AhG Chair and Project Editor of ISO/IEC JTC1 WG11 (MPEG: MPEG-7, MPEG-A, MPEG-V, MPEG-IoMT)
- ORCID : http://orcid.org/0000-0002-2359-8709
- Research interest: digital content (image, video, and music) analysis and management, fast image search and indexing, color adaptation, 4D media, sensors and actuators, VR, Internet of Things, and multimedia standardization

Navadita Sahu

- 2006: Bachelor's Engineering(B.E) in computer Science & Technology, BPUT, India
- 2006 ~ 2010: Embeded Software Engineer, Wipro Technology, Bangalore, India
- 2017: Master's in computer engineering, Myongji University, South Korea
- Research interest : Internet of Things, Big data analysis, Database Applications

Marius Preda is Associate Professor at Institut MINES-Telecom and Chairman of the 3D Graphics group of ISO's MPEG (Moving Picture Expert Group). He contributes to various ISO standards with technologies in the fields of 3D graphics, virtual worlds and augmented reality and has received several ISO Certifications of Appreciation. Marius received a Degree in Engineering from Politehnica Bucharest in 1998, a PhD in Mathematics and Informatics from University Paris V in 2002 and an eMBA from Telecom Business School, Paris in 2014.