INTERNATIONAL STANDARD

ISO/IEC 23093-1

First edition
2020-02

Information technology — Internet of media things —

Part 1: Architecture

Technologies de l'information — Internet des objets media —

Partie 1: L'architecture IoMT

Reference number
ISO/IEC 23093-1:2020(E)

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

Published in Switzerland

Contents

Page

Foreword

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

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

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

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

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

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

A list of all parts in the ISO/IEC 23093 series can be found on the ISO website.

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

Introduction

The ISO/IEC 23093 series provides an architecture and specifies application programming interfaces (APIs) and compressed representation of data flowing between media things.

The APIs for the media things facilitate discovering other media things in the network, connecting and efficiently exchanging data between media things. The APIs also provide means for supporting transaction tokens in order to access valuable functionalities, resources, and data from media things.

Media things related information consists of characteristics and discovery data, setup information from a system designer, raw and processed sensed data, and actuation information. The ISO/IEC 23093 series specifies data formats of input and output for media sensors, media actuators, media storages, media analysers, etc. Sensed data from media sensors can be processed by media analysers to produce analysed data, and the media analysers can be cascaded in order to extract semantic information.

This document does not specify how the process of sensing and analysing is carried out but specifies the interfaces between the media things. This document describes the architecture of systems for the internet of media things.

The International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) draw attention to the fact that it is claimed that compliance with this document may involve the use of a patent.

ISO and IEC take no position concerning the evidence, validity and scope of this patent right. The holder of this patent right has assured ISO and IEC that he/she is willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statement of the holder of this patent right is registered with ISO and IEC. Information may be obtained from the patent database available at www.iso.org/patents.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights other than those in the patent database. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Information technology — Internet of media things —

Part 1:
Architecture

1 Scope

This document describes the architecture of systems for the internet of media things.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

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

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

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

3.1 Internet of media things terms

3.1.1
audio
anything related to sound in terms of receiving, transmitting or reproducing it or of its specific frequency

3.1.2
camera
special form of an image capture device that senses and captures photo-optical signals

3.1.3
display
visual representation of the output of an electronic device or the portion of an electronic device that shows this representation, as a screen, lens or reticle

3.1.4
gesture
movement or position of the hand, arm, body, head or face that is expressive of an idea, opinion, emotion, etc.

3.1.5
haptics
input or output device that senses the body's movements by means of physical contact with the user

3.1.6
image capture device
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

3.1.7
internet of media things
IoMT
special subset of *IoT* (3.2.9) whose main functionalities are related to media processing

3.1.8
IoMT device
IoT (3.2.9) device that contains more than one *MThing* (3.1.12)

3.1.9
IoMT system
MSystem
IoT (3.2.9) system whose main functionality is related to media processing

3.1.10
loudspeaker
electroacoustic device, connected as a component in an audio system, generating audible acoustic waves

3.1.11
media
data that can be rendered, including audio, video, text, graphics, images, haptic and tactile information

Note 1 to entry: These data can be timed or non-timed.

3.1.12
media thing
MThing
thing (3.2.20) capable of sensing, acquiring, actuating, or processing of media or metadata

3.1.13
media token
virtual token for accessing functionalities, resources and data of media things

3.1.14
microphone
entity capable of capture and transform acoustic waves into changes in electric currents or voltage, used in recording or transmitting sound

3.1.15
media wearable
MWearable
MThing (3.1.12) intended to be located near, on or in an organism

3.1.16
motion
action or process of changing place or position

3.1.17
natural user interface
NUI
system for human-computer interaction that the user operates through intuitive actions related to natural, everyday human behaviour

3.1.18
presentation
act of producing human recognizable output of rendered media

3.2 Internet of things terms

3.2.1
actuator
component which conveys digital information to effect a change of some property of a physical entity

3.2.2
capability
characteristic or property of an entity that can be used to describe its state, appearance or other aspects

EXAMPLE An entity type, address information, telephone number, a privilege, a MAC address, a domain name are possible attributes, see Reference [1].

3.2.3
component
modular, deployable and replaceable part of a system that encapsulates implementations

Note 1 to entry: A component may expose or use interfaces (local or on a network) to interact with other entities, see Reference [2]. A component which exposes or uses network interfaces is called an endpoint.

3.2.4
digital entity
any computational or data element of an IT-based system

Note 1 to entry: It may exist as a service based in a data centre or cloud, or a network element or a gateway.

3.2.5
discovery
service to find unknown resources/entities/services based on a rough specification of the desired result

Note 1 to entry: It may be utilized by a human or another service; credentials for authorization are considered when executing the discovery, see Reference [4].

3.2.6
entity
anything (physical or non-physical) having a distinct existence

3.2.7
identifier
information that unambiguously distinguishes one *entity* (3.2.6) from another one in a given identity context

3.2.8
identity
characteristics determining who or what a person or thing is

3.2.9
internet of things
IoT
infrastructure of interconnected objects, people, systems and information resources together with intelligent services to allow them to process information of the physical and the virtual world and to react

3.2.10
interface
shared boundary between two functional components, defined by various characteristics pertaining to the functions, physical interconnections, signal exchanges, and other characteristics, as appropriate

Note 1 to entry: See Reference [5].

3.2.11
IoT system
system that is comprised of functions that provide the system the capabilities for identification, sensing, actuation, communication and management, and applications and services to a user

Note 1 to entry: See Reference [7].

3.2.12
network
entity that connects endpoints, sources to destinations, and may itself act as a value-added element in the IoT system or services

3.2.13
process
procedure to carry out operations on data

3.2.14
physical entity
thing (3.2.20) that is discrete, identifiable and observable, and having material existence in real world

3.2.15
reference architecture
description of common features, common vocabulary, guidelines, interrelations and interactions among the entities, and a template for an IoT architecture

3.2.16
resource
any element of a data processing system needed to perform required operations

Note 1 to entry: See Reference [8].

3.2.17
sensor
device that observes and measures a physical property of a natural phenomenon or man-made process and converts that measurement into a signal

Note 1 to entry: A signal can be electrical, chemical, etc., see Reference [9].

3.2.18
service
distinct part of the functionality that is provided by an entity through interfaces

Note 1 to entry: See Reference [10].

3.2.19
storage
capacity of a digital entity to store information subject to recall or the components of a digital entity in which such information is stored

3.2.20
thing
any entity that can communicate with other entities

3.2.21
user
human or any digital entity that is interested in interacting with a particular physical object

3.2.23
visual
any object perceptible by the sense of sight

4 Architecture

The global IoMT architecture is presented in Figure 1, which identifies a set of interfaces, protocols and associated media-related information representations related to:

— user commands (setup information) between a system manager and an MThing, with reference to interface 1.

— user commands (setup information) forwarded by an MThing to another MThing, possibly in a modified form (e.g., subset of 1), with reference to interface 1'.

— sensed data (raw or processed data) (compressed or semantic extraction) and actuation information, with reference to Interface 2.

— wrapped interface 2 (e.g., for transmission), with reference to interface 2'.

— MThing characteristics, discovery, with reference to interface 3.

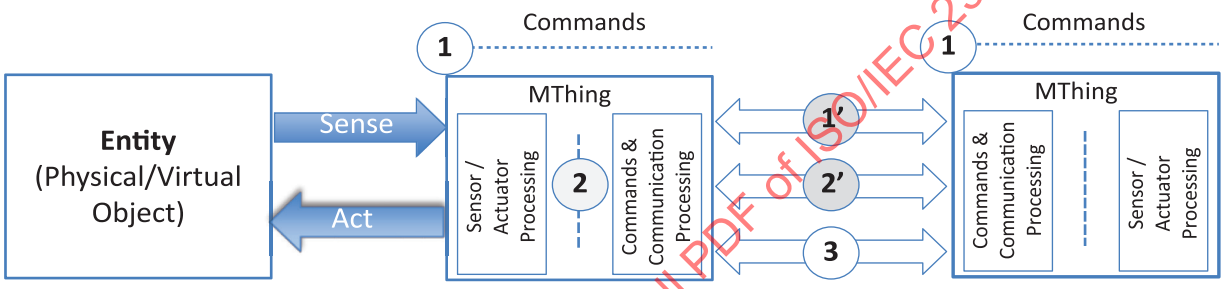


Figure 1 — IoMT architecture

This IoMT architecture can be mapped to the IoT reference architecture (Reference [4]) as shown in Annex A.

5 Use cases

5.1 General

MPEG identified 27 use-cases for IoMT; they are structured in the following five main categories:

— **Smart spaces: Monitoring and control with network of audio-video cameras (see 5.2)**

  — human tracking with multiple network cameras

  — automatic title generation

  — intelligent firefighting with IP surveillance cameras

  — networked digital signs for customized advertisement

  — digital signage and second screen use

  — self-adaptive quality of experience for multimedia applications

  — ultra-wide viewing video composition

  — face recognition to evoke sensorial actuations

  — automatic video clip generation by detecting event information

  — temporal synchronization of multiple videos for creating 360° or multiple view video

  - intelligent similar content recommendations using information from IoMT devices
- **Smart spaces: Multi-modal guided navigation (see 5.3)**
  - blind person assistant system
  - personalized navigation by visual communication
  - personalized tourist navigation with natural language functionalities
  - smart identifier: face recognition on smart glasses
  - smart advertisement: QR code recognition on smart glasses
- **Smart audio/video environments in smart cities (see 5.4)**
  - smart factory: car maintenance assistance A/V system using smart glasses
  - smart museum: augmented visit museum using smart glasses
  - smart house: light control, vibrating subtitle, olfaction media content consumption
  - smart car: head-light adjustment and speed monitoring to provide automatic volume control
- **Smart multi-modal collaborative health (see 5.5)**
  - increasing patient autonomy by remote control of left-ventricular assisted devices
  - diabetic coma prevention by monitoring networks of in-body/near body sensors
  - enhanced physical activity with smart fabrics networks
  - medical assistance with smart glasses
  - managing healthcare information for smart glass
- **Blockchain usage for IoMT transactions authentication and monetizing (see 5.6)**
  - reward function in IoMT by using blockchains
  - content authentication with blockchains

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

5.2.1 General

The large variety of sensors, actuators, displays and computational elements acting in our day-by-day professional and private space in order to provide us with better and easier accessible services lead to 11 use cases of interest for IoMT, mainly related to the processing of video information.

5.2.2 Human tracking with multiple network cameras

Because urban growth is today accompanied by an increase in crimes rate (e.g., theft, vandalism), many local authorities consider surveillance systems as a possible tool to fight this phenomenon. A city video surveillance system is an IoMT system that includes a set of IP surveillance cameras, a storage unit and a human tracker unit.

A particular IP surveillance camera captures audio-video data and send 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 extract the moving trajectory.

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

If the person gets out of the protected area, for example the person enters into a commercial centre, then the city system searches whether this commercial centre is also equipped with a video surveillance system. Should this be the case, the city video surveillance system sets up a communication with the commercial centre video surveillance system in order to allow another IP camera from the commercial centre video surveillance centre to keep capturing A/V data of the corresponding person.

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.

5.2.3 Automatic title generation

In the sustainable smart city of Seoul, IoMT cameras (smart CCTV) are deployed around the city. These cameras are continuously capturing video (24 hours/7 days). When unusual events such as a violent scene, crowd scene, theft scene or busking scene occurs, the title generator (event description generator) generates a title for the video clip with time and place information in real-time. The generated title is stored with the video clip in MStorage. As an example scenario, consider a CCTV capturing videos (visual data), with time and GPS information. The title generator analyses the video stream, selects a keyframe and combines time, GPS and keyframe to generate a formatted title. The captured video with the generated title is sent to storage.

5.2.4 Intelligent firefighting with IP surveillance cameras

Figure 2 illustrates an example use-case of intelligent firefighting with IP surveillance cameras. In this case, the fire station and the security manager can rapidly receive the fire/smoke detection alert, thereby averting a potential fire hazard. Unlike conventional security systems, the outdoor scene captured by intelligent IP surveillance cameras is immediately analysed and the fire/smoke incident is automatically alerted to the fire station based on the analysed results of the captured scene.

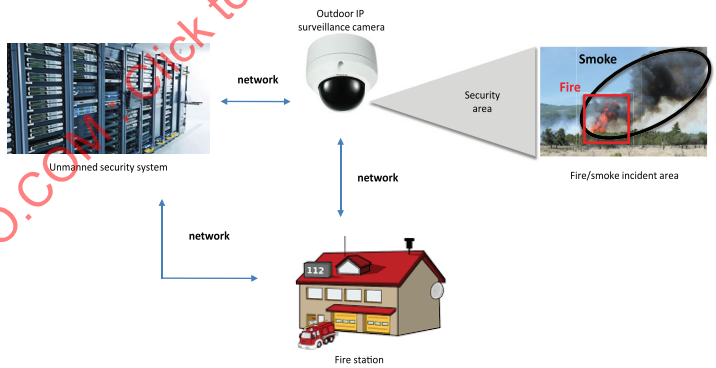


Figure 2 — Example use-case of intelligent firefighting

5.2.5 Networked digital signs for customized advertisement

A camera can be either attached to or embedded in a digital screen displaying advertising content, so as to be able to capture A/V data and send them to both a storage unit and a gaze tracking/ROI analysing unit. When the gaze tracking/ROI analyser detects a person in front of the corresponding digital sign, it starts to trace the eye position, calculates the corresponding region of interest on the currently played advertisement, and deduces the person's current interest (e.g., goods) on the advertisement. When the person moves to the other digital sign, that new sign starts playing relevant advertisement according to the estimated person's interest data.

5.2.6 Digital signage and second screen use

This use case addresses the pedestrians who want to get additional information (e.g., product information, characters, places) of content displayed on digital signs with his/her mobile phones (i.e., second screens), as illustrated in Figure 3.

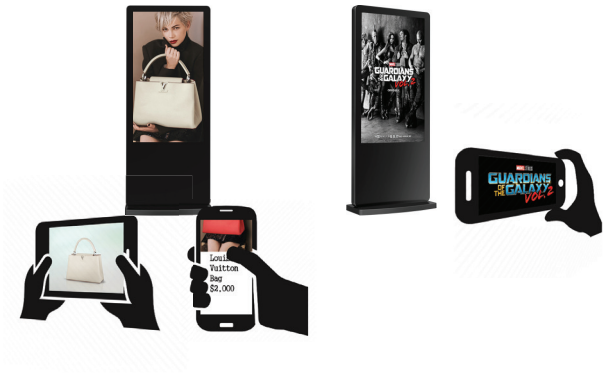


Figure 3 — Display signage and second screen use-case

5.2.7 Self-adaptive quality of experience for multimedia applications

The self-adaptive multimedia application is an application working on wearable device with a middleware providing optimal quality of services (QoS) performance for each application, according to the static/dynamic status of the application and/or system resources.

The user initially starts the self-adaptive multimedia application and updates the initial setup to guarantee the application's performance quality in a wearable device. The self-adaptive application needs the static/dynamic status information between the wearable device and processing unit. And then the self-adaptive application is normally running on wearable devices until a status change/update event is generated. These events happen at the moment of detection of a performance level decrease and then the status information request is sent to the processing unit.

The processing unit can support a heterogeneous type of wearable devices and it includes static/dynamic system manager to optimize computing performance. The processing unit performs resource management optimally, based on the performance requirement of self-adaptive application.

5.2.8 Ultra-wide viewing video composition

The ultra-wide viewing video composition is possible thanks to the videos captured from multiple cameras equipped with multiple sensors (time, accelerator, gyro, GPS, and compass) along with a video composer, storages and display devices as MThings.

5.2.9 Face recognition to evoke sensorial actuations

An IP surveillance camera captures audio-video data and send them to both a storage unit and a face recognizer 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 alternatively extracted and sent to a storage unit. In this use case, the scent generator can by replaced by any type of actuators (e.g., light bulbs, displays, music players).

5.2.10 Automatic video clip generation by detecting event information

This use case describes automatic video clip generation by detecting event information from audio/video streaming feed from a video camera. Usually, family or friends hold many events such as birthday

parties, wedding anniversaries or pyjama parties. By using surveillance cameras, these events can be detected and pictures or videos taken at the event can be used to make a time-lapse video.

5.2.11 Temporal synchronization of multiple videos for creating 360° or multiple view video

A new video can be created by using videos captured by multiple cameras. Any camera has its own local clock with various sensors and can record the shooting time based on the local clock. As each camera has a different timeline, when creating a new video (e.g. 360° video) using time information (e.g., stitching) from two different devices, some errors are likely to occur. The time-offset information between individual videos can be cancelled by performing temporal synchronization using visual and/or audio information with sensor data, thus obtaining a natural-looking video.

Moreover, if individual videos are transmitted through the network, people can watch the videos taken from various viewpoints of the event. This means someone can watch just one video whilst another watches multiple videos at the same time, and someone else can alternately watch videos.

5.2.12 Intelligent similar content recommendations using information from IoMT devices

Currently, video content taken by individuals with unprofessional cameras, smartphones, etc. is commonly found on various internet resources, from social networks to video sharing systems. Such content is very heterogeneous: concerts, sports games, unboxing videos of new products, etc. While, for a person, it is practically impossible to provide precise and rich recommendations of video content, with intelligent similar content recommendation systems, users can easily have the choice on a large variety of content related to the content he/she already posted.

To recommend similar content, metadata of the video content is needed. The metadata of the video content is generated using the position (GPS) captured by a specific individual, visual, auditory and time information of that video.

5.3 Smart spaces: Multi-modal guided navigation

5.3.1 General

This clause regroups 5 use cases to illustrate the way in which multimodal information can be processed and fused inside IoMT systems in order to provide the user with an enhanced navigation experience.

5.3.2 Blind person assistant system

The navigation in smart spaces can help blind and visually-impaired persons in many ways, for instance by providing them with information about possible collisions, with guiding directions or the position of local landmarks.

Collision warning: 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 coordinating unit receives the distance data and decides the actions to be taken. If the distance is reasonably far, an alarming text data of the corresponding distance (e.g., “5 metres before colliding obstacles ahead.”) is produced by the collision coordinator and sent to a text-to-speech generating 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 really close, the collision coordinator activates either a wrist band to vibrate or the headphone to create beeping sounds.

Guiding direction: Assume that a blind person travels to a destination. The global navigation can be provided by any web service. However, the local navigation can be enhanced by RFID tags that contain exact location coordinates. The RFID tags can be embedded in every street corner. The blind person carries a smart cane, a smart phone and a networked headphone. The smart cane is equipped with an RFID reader, some inertia sensors (e.g., a gyro, a compass). The RFID reader can read the RFID tags

embedded in every street corner. A direction guiding unit receives the RFID tag 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. 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.

Informing local landmarks: Assume that a blind person arrives at a destination. The blind person wears a smart glass equipped with a camera, a smart phone and a networked headphone. The camera (MThing camera) takes an image shot in front of the person and sends it to a visual feature extracting module. The visual feature extractor extracts feature data from the image and sends again to a landmark matching unit. The landmark matcher compares the feature data from the database and retrieves the name of which he/she is watching. Upon the retrieved name, the landmark matcher creates landmark name guidance (e.g., “you are in front of the burger restaurant”) and sends it to a text-to-speech generating 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.

5.3.3 Personalized navigation by visual communication

Visual messages can improve the efficiency of the interaction between the user and the wearable device when the display resources are very restricted. In a visual communication everyone can intuitively understand a pictogram, so that people can easily express an implicit meaning or an ambiguous emotion.

The rough map program on wearable devices is executed by a user who is travelling abroad. He/she is locating visual objects such as his/her characters, restaurants, and attractions on a rough map and presses the button to take a tourist route which is recommended by a processing unit. The wearable device is transmitting data (which consists of information related to a visual object reflecting user intentions and context sensed by sensors, e.g., location, weather, time and temperature) to processing units or servers to request recommendations. The processing unit makes a recommendation including a visual object, service information and a tourist route based on the processing data received from the wearable device and sends a recommendation to the wearable device. The wearable device displays the recommended tourist route according to the processed information by using visual objects.

5.3.4 Personalized tourist navigation with natural language functionalities

The natural language functionalities can serve as a precious tool in improving the comfort of a tourist travelling abroad. The present use-case illustrates the usage of speech translation, questioning-answering and multimodal interaction.

Speech translation: 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, laptop or tablet to smartphone, speech translation systems are getting even more usable with wearable devices. When a user speaks to the microphone embedded in the smart watch or headphone in one language, an automatic translator can be activated to enable a conversation with a person speaking a different 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 smartphone (stand-alone translation system) which is connected to the wearable device. With the wearable translation service, the user is able to use their hands freely while the conversation is translated. The wearable device is also used for automatically finding someone who can speak one of the languages which is embedded in the translation system in a travelling situation.

Question-answering: QA is an advanced function to generate answers for the user’s question in a natural language. More systems in the future are expected to be equipped with QA functions for advanced user experience. Consider the case of Michael, who visits Milan, Italy during his holiday. It is his first time in Italy and he does not have much knowledge on history or location of the various

attractions. Using his wearable device, a smart headphone, he asks all the questions in his natural language via the speech interface and receives answers conveniently through the intelligent QA service. He can travel around Milan easily without help of a tour guide.

Multimodal interaction: The user interfaces to various devices have been enhanced in the direction of improving convenience of communication and providing rich user experiences. Multimodal user interfaces combine independent modalities such as speech, gestures, text and touch depending on situations to achieve maximum convenience to use devices. Multimodal interaction with wearable devices would compensate with the failure cases where single modality is used. For example, when the speech recognition does not work very well with complex sentences, gesture command can support the missing information. Depending on the situation of the users, one modality is better than others and most times combined modalities, e.g. speech plus gestures, work better.

The first example can be a user pointing to the building on the right side and asking "Is there a bank in that building?"

The second example relates to facial expressions, which can be used to help people understand the speaker's emotion and intention and help communication among people go smoothly. They are used not only for everyday conversation but also for people with disabilities. Deaf people use facial expressions in addition to lip reading to communicate with other people. Considering that wearable devices could be good tools to improve accessibility for people with disabilities or specific needs, facial expressions that transmit the user's intention and emotion to the communication partner should be supported to enhance user interface.

5.3.5 Smart identifier: Face recognition on smart glasses

The use case of face recognition using smart glasses is conceptually represented in Figure 4. First, a face region is detected from the incoming image sequence; then the detected face is recognized; and finally the identification information associated with the recognized face is presented to a user.

The processing required for face recognition can be shared by several processing units in an efficient manner. For instance, a rectangular region, from which a face is supposed to be detected, is extracted in the processing unit embedded in the smart glasses, by relatively simple pre-processing. Then, the extracted region is transmitted to the main processing unit in which face detection and recognition are performed with the required computational power. Finally, the associated identification information is delivered from the main processing unit to smart glasses.

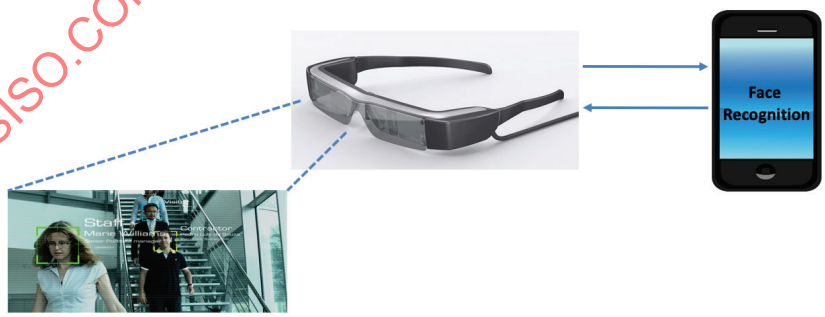


Figure 4 — Face recognition on smart glasses

5.3.6 Smart advertisement: QR code recognition on smart glasses

A QR code is a standardized 2D barcode which is represented in the image sequence and can contain a lot of information. Its applications include the exchange of information, product tracking, item identification and general marketing.

QR code recognition can be achieved by using smart glasses. First, a region including the QR code is detected from the incoming image sequence, and then the QR code is recognized by analysing the

detected region. Finally the identification information associated with the recognized QR code is presented to a user.

The processing required for QR code recognition can be shared by several processing units, as explained in the previous use-case.

5.4 Smart audio/video environments in smart cities

5.4.1 General

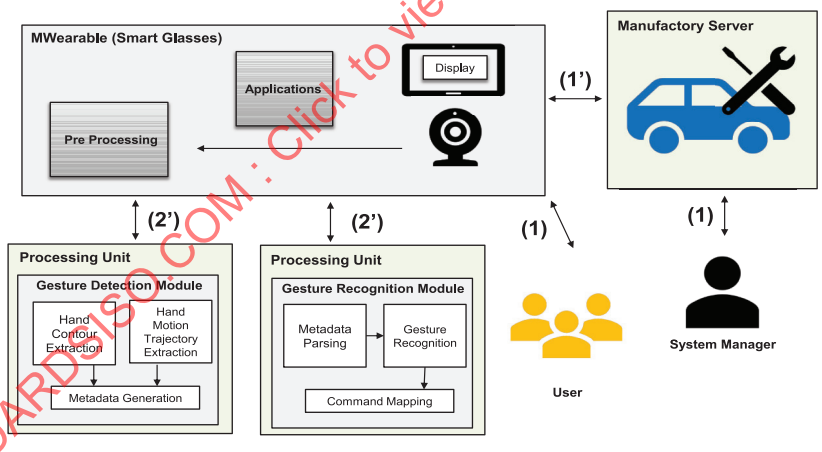
A city becomes smart when its traditional infrastructures are combined with disruptive technologies to improve the life of its citizen and business activities in a sustainable way. The realization of smart cities involves aggregating operation of different subsystems (smart spaces) that need to retain their primary private function but must interact with each other in order to fulfil more global objectives.

This aggregation of subsystems could be made between homogeneous subsystems or between heterogeneous subsystems. Aggregation could also be made either through communications between functional systems operating inside the same area or operating in spatially correlated area, from the smart buildings to smart cities, and ultimately to smart territories.

The use cases of interest for IoMT are presented in 5.4.2 through 5.4.5.

5.4.2 Smart factory: Car maintenance assistance A/V system using smart glasses

Figure 5 illustrates the use case of smart glasses for the car maintenance system. It is assumed that a technician wearing smart glasses is performing the car maintenance. 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 gesture. This way, a user efficiently does maintenance work, freely using both hands.



NOTE The numbers relate to the IoMT interfaces presented in Clause 4.

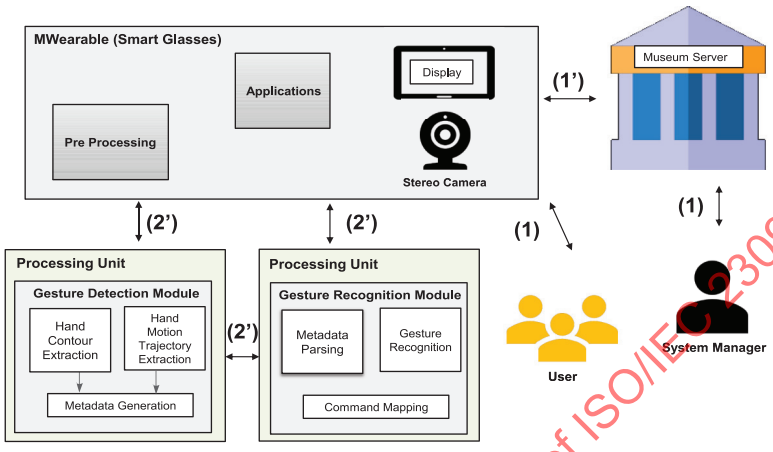
Figure 5 — Car maintenance assistance A/V system using smart glasses

5.4.3 Smart museum: Augmented visit using smart glasses

Figure 6 illustrates a use case of an augmented museum visit with smart glasses, in which augmented information such as a narrative explanation about a modern work of art and a video clip showing the painter interview can be provided according to a user's request invoked by hand gesture. This way, a user enjoys a museum tour with rich information presented by smart glasses without any brochure and/or the help of a guide.

When a user wants to know the details of a specific picture, the picture is identified by recognition of the picture identification number displayed under the picture by smart glasses. Then, the available

information associated with the identified picture, such as video clips explaining the artists of the painter, are listed on the smart glasses. Then, a user selects one of the clips and plays the clip. Basic types of play controls, such as play, stop, pause, fast forward, and random access are enabled. In this use case, such interaction with smart glasses is enabled by hand gesture. In order to do that, the hand gesture is recognized and is mapped into a command to control the smart glasses and an application is available in the smart glasses. Much more diverse use cases using gesture-based commands with enabled smart glasses are possible in a museum visit.



NOTE The numbers relate to the IoMT interfaces presented in Clause 4.

Figure 6 — Augmented museum visit using smart glasses

5.4.4 Smart house: Light control, vibrating subtitle, olfaction media content consumption, odour image recognizer

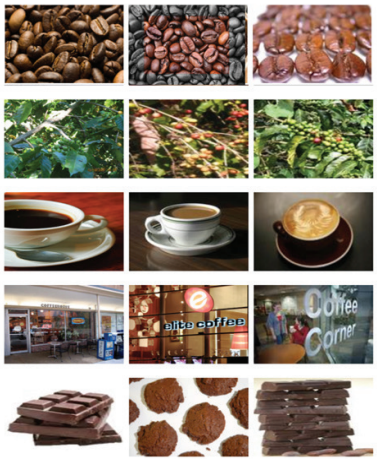
The use of the IoMT in smart houses is illustrated through the following four use cases.

Light control with respect to the characteristics of the music being played: The smart lights provide APIs to change brightness, hue and saturation. By using these APIs, the user can remotely control light characteristics; for instance, the light can be automatically adjusted with respect to the sound near to the light: if the music being played contains a high frequency, the light turns to red, etc.

Vibration subtitle for movie by using wearable device: The user experience in movies theatres is continuously evolving: you can watch the movie with 3D glasses or a vibrating chair in order to enhance the fun. You can also enjoy this vibration by using a smartphone, smartwatch and Bluetooth earphones at home.

Olfaction media content consumption: An image is taken by a camera jointly with the sensing of the ambient smell by an e-nose sensor (e.g. the image corresponds to a fire and the smell to the grilled meat). The captured image and sensed scent metadata are combined in olfaction-enhanced media form and then transferred to a database or to some other device capable to consume the olfaction-enhanced media.

Odour image recognizer: The images/video taken by a camera are input to an image analyser that is designed to detect several pre-defined objects, as illustrated in Figure 7. According to the output of the image analyser and to some additional logic, a scent sensation is created. For instance, if some coffee beans are detected as an image, a coffee scent is produced.



No.	OdorImageLabelCS	ScentCS
1	Coffee bean	Coffee
2	Coffee tree	
3	A cup of coffee	
4	Coffee shop	
5	Coffee chocolate	

Figure 7 — Odour image recognizer synopsis

5.4.5 Smart car: Head-light adjustment and speed monitoring to provide automatic volume control

IoMT can be used to allow drivers to automatically avoid glare, encountered when a driver chooses to turn on his/her vehicle headlight, such as high-beams or high intensity discharged (HID) lamps and to automatically adjust the vehicle headlight (in terms of both brightness and direction).

IoMT can also help by providing automatic volume control for in-car audio systems. While driving, noise from the engine to cabin varies with respect to acceleration and speed. The noise increases when the revolutions per minute (RPM) and/or speed increases and interferes with audio listening. To provide a more comfortable listening environment, volume is adjusted with respect to the current speed or RPM of a vehicle.

5.5 Smart multi-modal collaborative health

5.5.1 General

It is currently accepted that the cost of healthcare can be reduced while improving the quality of life of patients by integrating professional and user-created data. To illustrate this trend, the 5 uses cases in 5.5.2 through 5.5.6 are under the scope of IoMT.

5.5.2 Increasing patient autonomy by remote control of left-ventricular assisted devices

Current day left-ventricular assist devices are generally managed by some parameters set before the patient leaves the hospital. However, in order to increase the life autonomy of the patient, solutions for ensuring the possibility of bi-direction communication between the heart and the physician, in the sense of distant control and monitoring of its state and the state of its host (the human), are searched for.

A patient with a left-ventricular assist device is having breathing problems, and an ambulance is taking them to the hospital. While the patient is transported, he is connected with a range of sensors: blood pressure, body temperature, breathing, etc. and monitored with a real-time camera. All these sensors (sources of precious information) are connected to a processing engine, which gathers all data and transmits them to the hospital where the patient is heading. At the same time, the physician remotely accesses the sensors, reads the data and actually interacts with them, especially with the patient artificial heart, by controlling the tempo, the level of compression, etc.

5.5.3 Diabetic coma prevention by monitoring networks of in-body/near body sensors

The possibility of deploying reliable services around in-body devices is currently under investigation. These devices are battery-free, wireless, intelligent sensor modules implanted in the body to enable continuous health monitoring in the future. Tiny sensor-devices gather physiological information and communicate this to the outside world if deviations in physiological properties are detected.

John Smith, 55 years old, suffers from diabetes: so, physical activity would be of much benefit to him but might also cause him a sever coma situation. John visits his doctor for implanting the latest glucometer, which is the size of a rice grain. Now, he can go for a walk or a mild jogging every day. The in-body glucometer is connected to a processing engine, which also receives relevant data from his wrist device measuring his body temperature, skin humidity and heart rate. His training schedule automatically adjusts to his physical performance, through an intelligence medical service. Two years later, the sensor registers abnormal patterns in the received sensors signals. A visit to his doctor is scheduled automatically.

5.5.4 Enhanced physical activity with smart fabrics networks

A digital t-shirt is also called a d-shirt. It is made of intelligent fabrics integrating sensors in its design for monitoring and gathering data (audio/video and a semantic description about them) from the person wearing it and the environment where the person is. D-shirts also send/apply some information/ feedback to the wearer by vibration and enable haptic interaction.

A person is engaging in smart sports by wearing a smart digital shirt (d-shirt) for optimizing their activity. While running, the d-shirt is capable of recognizing the type of activity and the person's state, and provides feedback on the person's skin. If the person is running in a new environment and needs guidance for the route, by following the feedback from the d-shirt they are able to make the course as planned, with a tempo well adapted for their age, weight and condition. While running, they can also record the environment (audio/video).

By allowing the d-shirt to communicate with the processing unit (an intelligent and more powerful device), the user is capable of seeing their results at the end of their activity, and can let the processing unit make a recommendation for the next activity.

5.5.5 Medical assistance with smart glasses

Medical application is one of the most promising areas in which smart glasses can play a continuously increasing role, thanks to the possibility of combining voice and gesture commands to display properties.

In general, it is assumed that there are two available processing units in wearable smart glasses for medical applications: one is a basic processing unit with low computational power and the other is a main processing unit with high computational power, which can be provided as a separate device such as a smartphone, a desktop PC, etc. Additionally, a database (DB) in a server may be used for providing specific information such as patient information including medical treatment records.

While patient care has become increasingly data-driven, doctors need a way to receive heterogeneous data such as vital signs, medical images and patient records while remaining hands-on with the patient. A real-time app that is running in smart glass can present associated information on the equipped display without interruption for the doctor during surgery or treatment. In this way, smart glasses allow doctor to better concentrate on their main operations. There are many such use cases of smart glasses for medical applications:

Surgery-oriented usage of smart-glasses:

— Training and remote education.

— Image/video based decision-making.

Emergency support (save time and life):

— A rescue crew's video is transmitted to an emergency room in the hospital.

— A rescue crew can use smart glasses in the field to access the patient's identification information and medical record.

Telemedicine and smart interaction with electronic medical record (EMR) (dictation, advance information access)

When a paramedic team equipped with smart glasses arrives at an accident spot, images or video sequences can be taken, see Figure 8. This information can be transmitted in real-time under their voice and/or gesture control to a hospital server; thus, a doctor in the hospital can view and/or record the state of the patient and have first-hand information about the circumstances of the accident.

Moreover, if a paramedic team wants to obtain expert advice during patient transport, a direct call can be established. Inside the hospital, a signal of emergency call from the paramedic will be presented on an emergency doctor's smart glasses. Then, the doctor can view the video streaming to check the patient's state. At this moment, if necessary, the doctor can request more information, such as the vital signs of the patient, to perform proper treatments. By using gesture/voice commands, the paramedic can transmit the requested information, such as heart rate, oxygen levels, blood pressure, etc. The doctor can then guide the paramedics to perform the right actions on the patient.

This use case clearly states the need for supporting bi-directional conversion of image/video data captured by smart glasses inside IoMT and the DICOM-based PACS systems.

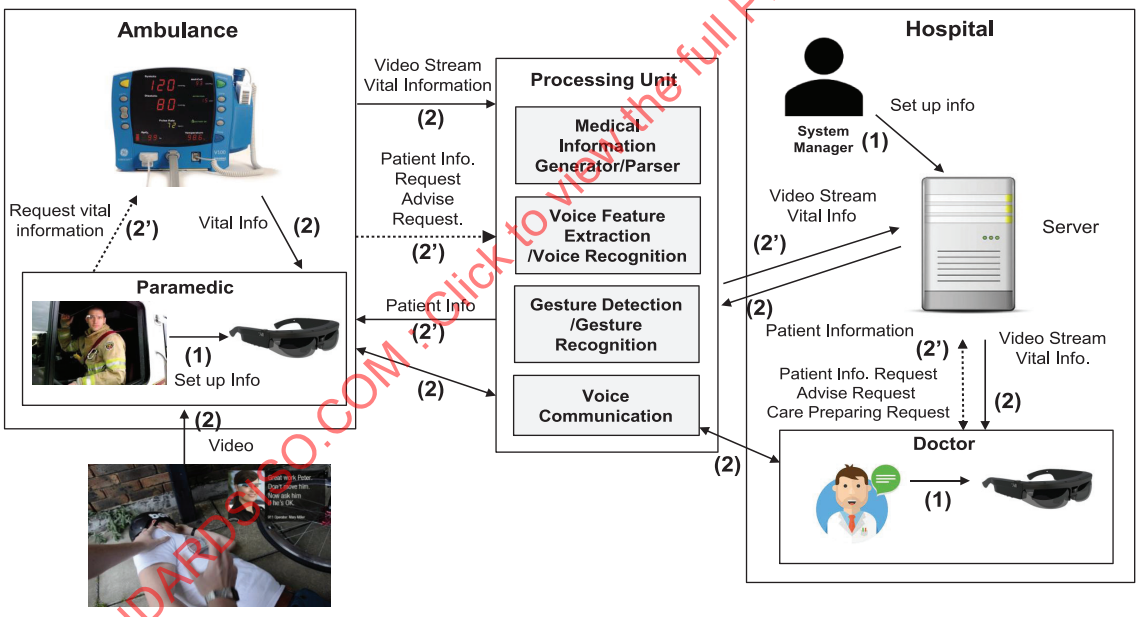


Figure 8 — Smart glasses media consumption control for emergency support (save time and life)

5.5.6 Managing healthcare information for smart glasses

Medical information captured by smart glasses can be conveniently managed by using blockchain technologies.

As described in the use case in 5.5.5, the video and audio information related to an accident is transmitted to the hospital by paramedics. Then, the medical staff can prepare the appropriate medical