**[[1]](#footnote-1)INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG2011/N15997**

**San Diego, California, February 2016**

|  |  |
| --- | --- |
| **Source** | **Telecom SudParis** |
| **Title** | **White Paper on the 2nd edition of ARAF** |
| **Author** | Traian LAVRIC, Marius PREDA |

# Introduction

Augmented Reality (AR) applications refer to a view of a real-world environment (RWE) whose elements are augmented by content, such as graphics and sound, in a computer driven process.

The Augmented Reality Application Format (ARAF) is an ISO standard published by MPEG and can be used to formalize a full **Mixed and Augmented Reality (MAR) experience**. It consists of an extension of a subset of MPEG-4 Part 11 (Scene Description and Application Engine) standard, combined with other relevant MPEG standards (MPEG-4 Part 1, MPEG-4 Part 16, MPEG-V) and it is designed to enable the consumption of 2D and 3D multimedia, dynamic, interactive, natural and virtual content. About two hundred elements (nodes) are standardized in MPEG-4 Part 11 allowing various kinds of MAR experiences to be created. ARAF is based on a subset of these elements and introduces new ones in order to facilitate the creation of augmented reality applications and to support features that are generally required in the creation of such experiences.

ARAF allows the creation of any MAR experience, from basic “Hello World” applications to very complex scenes that can integrate a full set of features provided by the ARAF technology. Among all, here's a list of the most important ones: interoperability between real and virtual worlds by connecting to sensors (e.g. device orientation, GPS, camera, etc.), commanding actuators, dynamic content (interpolators, scripting), user interaction (touch sensors) and server communication. Moreover, a set of elements specifically designed for augmented reality experiences have been introduced in ARAF in order to allow the content creator to easily add specific functionalities in his AR experience: image recognition and tracking, audio recognition and synchronization, map support and camera calibration.

Looking from another perspective, by using ARAF, content creators can design MAR experiences from location-based applications to image-based augmentation, from local to cloud-assisted processing, from single-player to multiplayer.

# AR nowadays

There is an increasing demand of AR services and applications around the world nowadays, especially because the current technology evolved to the point where all the required information can be collected, processed in real-time and presented in such a way that can improve our everyday life. Augmented reality has many applications. First used in military, industrial, and medical applications, it has also been applied to commercial and entertainment areas.

The following table presents domains where AR has already started to be used.

|  |  |
| --- | --- |
| **User Type** | **Domains** |
| **Professional user** | Archaeology, Architecture, Construction, Engineering Design, Industrial design |
|  | Manufacturing, Maintenance, Repair, Task support |
|  | Medical |
|  | Emergency management / search and rescue |
|  | Robotics and Telerobotics |
|  | Collaborative environements, Spatial immersion and interaction |
| **End-user** | Commerce (try before buy): Beauty, Consumer Design |
|  | Education, Training |
|  | Entertainment, Gaming, Sports |
|  | Outdoor & Indoor Navigation, Tourism and sightseeing |
|  | Television |
|  | Translation |

Below there are presented screenshots of some AR applications that have been developed.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | \\psf\Home\Desktop\mpeg\113 - Geneva\whitepapaer_images\architecture-2.png | | Figure1. Virtual visit to prefabricated house (copyright Augment). | | **\\psf\Home\Desktop\mpeg\113 - Geneva\whitepapaer_images\TWNKLS.jpg** | | Figure 3. Maintenance and repair in Maritime Industry (copyright Twnkls). | | **\\psf\Home\Desktop\mpeg\113 - Geneva\whitepapaer_images\veinviewer.jpg** | | Figure 5. AR used to display the patient vasculature (copyright VeinViewer). | |  | | Figure 7. AR dressing room : virtually try on clothes before buying (copyright TopShop). | | **\\psf\Home\Desktop\mpeg\113 - Geneva\whitepapaer_images\National-Geographic-Realidad-Aumentada-300x180.jpg** | | Figure 9. Interaction with virtual animals in public areas (copyright Appshaker). | | |  | | --- | | **\\psf\Home\Desktop\mpeg\113 - Geneva\whitepapaer_images\Boeing-augmented-reality_olkby1.jpg** | | Figure 2. AR in an assembly line (copyright Boeing). | | \\psf\Home\Desktop\mpeg\113 - Geneva\whitepapaer_images\volkswagens-marta.jpg | | Figure 4. Using AR to highlight body car parts to be investigated (copyright Volkswagen). | | \\psf\Home\Desktop\mpeg\113 - Geneva\whitepapaer_images\phote-for-the-marker.jpg | | Figure 6. Markerless 3D model superimposed in the real world. | | \\psf\Home\Desktop\mpeg\113 - Geneva\whitepapaer_images\DuffettAR_Detail_Image.jpg | | Figure 8. Using AR to preview a full-scale, mockup of feature stairs (copyright BNBuilders). | | **\\psf\Home\Desktop\mpeg\113 - Geneva\whitepapaer_images\iSkull.jpg** | | Figure 10. AR used to display interactive 3D model of a skull (copyright iSkull). | |  | |  | |

# ARAF

Augmented Reality Application Format focuses not on the client or server procedures but on the data format used to provide an augmented reality presentation. ARAF specifies scene description elements for representing AR content, mechanisms to connect to local and remote sensors and actuators, mechanisms to integrated compressed media (image, audio, video, graphics) mechanisms to connect to remote resources such as maps and compressed media.

MPEG already provided data type representations for all kinds of media, form static image, video, audio to 3D graphics and complex dynamic scenes. Additionally, MPEG developed a set of standards related to sensors and actuators. By bringing these two components together into an application format called ARAF, MPEG enables interoperability when used to build AR applications and services in a standard manner.

The following table synthesizes the set of existing MPEG tools that are used for addressing AR applications.

|  |  |
| --- | --- |
| **Feature** | **Standard** |
| Image (video) & audio capturing | MPEG-V |
| Capture real camera position and orientation | MPEG-V & CDVS |
| Detection and tracking of visual objects | CDVS |
| Transmission of media assets | MPEG-4 Part 1, 2, 3, 10, 11, 16, 25 |
| Image & video rendering as a background | MPEG-4 Systems |
| Control the virtual camera parameters | MPEG-4 Part 11 |
| Synthetic content representation | MPEG-4 Part 11 and Part 16 |

Therefore, by using these MPEG technologies it is possible to envision a generic AR browser. Instead of downloading a new and heavy application for every context, the users could simply point to an URL. The ARAF browser retrieves a scenario from the Internet, starts the video acquisition, tracks images and objects, recognizes them from visual signatures, computes the camera position, downloads 2D/3D graphics, composes a new scene, gets input from various sensors and constantly adapts the interaction in order to offer an optimal ARAF experience.

Instead of developing a new application for each use case and smart phone platform, the industry could rely on MPEG-compliant Authoring Tools and MPEG-compliant browsers to reach a maximum number of customers. The schema below presents the MPEG vision on AR:

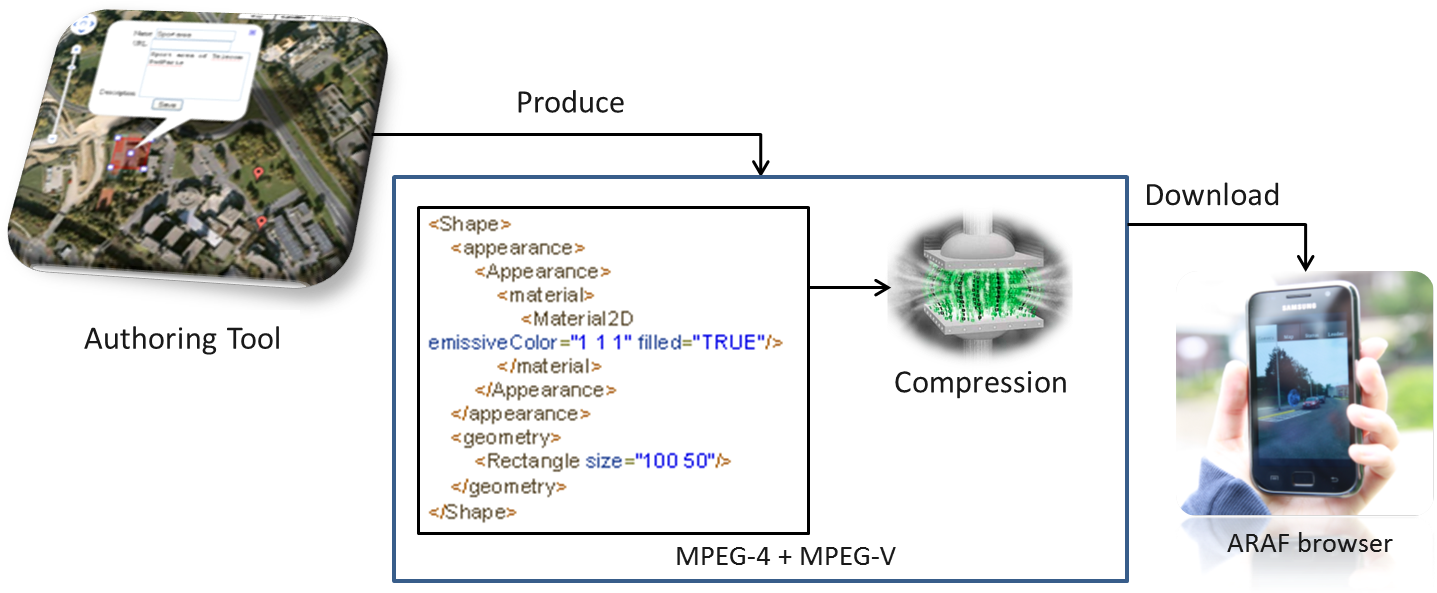


Figure 11. MPEG vision on AR

The creation of ARAF content is fast and easy because the AR experience creator does not necessarily need to have any ARAF prior knowledge. Instead of designing the application from scratch, he can choose to create generic template-based AR applications in a matter of hours by using authoring tools. The authoring tool can be a user-friendly web interface where the user is allowed to design and define the behavior of his AR experience. The generated result, an XML representation, and the media which is linked into the application, are compressed together in a single file which contains the full ARAF content. At the end of the chain, the content is downloaded and consumed by an ARAF browser.

On the other hand, the ARAF model provides full control to the designers who want to create brand new AR experiences. If one decides to create an AR experience from scratch, he can simply start writing the ARAF content by himself, as one would create an HTML page.

Some key elements differentiating ARAF from other available technologies are the following:

- It can be easily used by designers without a programming background.

- It is an open standard provided by ISO.

- The software implementation of ARAF browser is available as open source.

## ARAF principle and context

Figure 17 presents a generic AR architecture including ARAF. The **ARAF content**, available as a file or stream, is interpreted by a device, called **ARAF device**. The elements (nodes) of the ARAF scene point to different sources of multimedia content such as 2D/3D image, audio, video and graphics, and sensor/sensory information sources/sinks that are either local or remote.

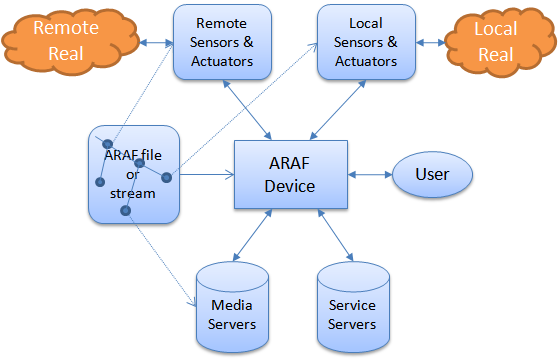


Figure 12. The ARAF context

## ARAF Scene Description

In order to design a multimedia scene ARAF extends MPEG-4 Part 11 BIFS (ISO/IEC 14496-11) which is based on VRML97 (ISO/IEC 14772-1:1997). About two hundreds elements are standardized in MPEG-4 BIFS and VRML, giving the possibility of a content creator to create any AR experience that he can imagine.

MPEG-4 Part 11 describes a scene with a hierarchical structure that can be represented as a graph. The nodes (elements) of the graph build up various types of objects, such as audio, video, image, graphics and text. Furthermore, to ensure flexibility, a new user-defined type of node can be defined on demand by using the prototyping (PROTO) method.

In general, the nodes expose a set of parameters, through which aspects of their appearance and behavior can be controlled. By setting these values, the content creators (designers) have the possibility to force a scene reconstruction or scene update at clients’ terminals to adhere to their intention in a predefined manner. In more advanced scenarios, the structure of the BIFS nodes is not necessarily static; nodes can be added or removed from the scene graph arbitrarily and dynamically.

Certain types of nodes called sensors can interact with users and trigger appropriate actions, which are transmitted to other nodes through a routing mechanism, causing changes in the state of these receiving nodes. They are one of the bases for dynamic behavior of multimedia content supported by MPEG-4 part 11.

The maximum flexibility in the programmable feature of MPEG-4 scene is carried out with the Script node. By using the routing mechanism to a Script node, an associated function can be called. The implementation of this function is designed by the MAR experience creator, i.e. the creator can freely process some computations, and then output the result to the scene graph. In other words, by using the scripting feature of ARAF, the MAR experience creator can achieve a set of functionalities which are not supported natively by the nodes that are already part of the technology. This way the scene can be programmatically updated based on time events, user actions, or a mixture of the two. It is therefore provided the possibility of manipulating everything that is defined in the scene graph, at any moment in time and based on events that are triggered automatically or by the end user.

ARAF supports the definition and reusability of complex elements by using the MPEG-4 PROTO mechanism. The PROTO statement creates new nodes by defining a configurable object prototype; it can integrate any other node from the scene graph and it basically provides a method of creating new functionality that can be re-used within the scene graph.

Furthermore, a new functionality (a PROTO) that was designed by a MAR experience creator can be re-used in different scene graphs by using the EXTERNPROTO mechanism. This feature allows an object prototype to be described in a separate file and imported from there into any other scene graph.

## Sensors and actuators management

The data that is captured from sensors or used to command actuators in ARAF is based on ISO/IEC 23005-5 data formats for interaction devices (MPEG-V Part 5).

MPEG-V provides architecture and specifies associated information representations to enable interoperability between virtual worlds. Concerning mixed and augmented reality, MPEG-V specifies the interaction between virtual and real worlds by implementing support for accessing different input/output devices, e.g., sensors, actuators, vision and rendering and robotics.

ARAF is capable of connecting to all sensors within the general set of sensors provided by smart phones today: orientation, position, acceleration, angular velocity, GPS, altitude, geomagnetic, camera and microphone.

In addition to reading data from sensors, ARAF can also transmit data to actuators, controlling therefore physical devices directly from the scene graph. Here's the list of actuators that are supported in ARAF: light, vibration, tactile, flash, heating, cooling, wind, sprayer, scent, fog, rigid body motion, and kinaesthetic.

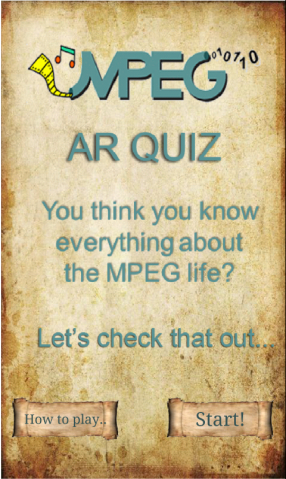
## List of current AR features supported by ARAF

* Elementary media :
  + Audio
  + Image and video
  + Textual information
  + Graphics
* Programming information
* User interactivity
  + Time sensors
  + Touch sensors
  + Media sensors
  + Access to sensors and actuators of physical devices
* Scene graph related information
  + Organization of scene elements
  + Navigation information
  + Layouting
  + Visual identification and tracking, local and remote
  + Remote visual registration and composition
  + Audio identification and synchronisation, local and remote
* User localization
  + GPS
* Map support
* Dynamic and animated scene
  + Interpolators and valuators
  + Scripting
  + Sensors
* Communication and compression
  + Media control
  + Map support
* Terminal capabilities

## ARAF Examples

### ARQuiz

ARQuiz is an augmented reality location-based game for smart phones and tablets, based on the MPEG ARAF standard. The game consists of answering questions related to physical places that are being visited by players. Each question has an associated hint. A hint is a short story that eventually helps the player to find the correct answer to a given question. All hints are placed on a specific route that the user has to follow in order to advance in the game. The hints are visible on a map (MAP view). As every hint is represented by the same icon on the map, the player does not know which hint corresponds to his current question. This is the reason why the player has to use the AR view where the hints are represented by numbered 3D spheres. Once the correct number has been identified (it is visible in the AR view), the player has to “catch” that sphere by getting closer to it (five meters or less). When close enough to the hint, the corresponding story is displayed on the device’s screen. By reading the story, the player eventually finds the correct answer of the question but may also discover more details about the place he or she is visiting at that specific moment.

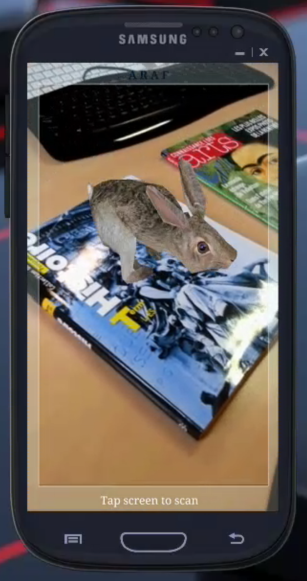
  

**Figure 13. Views from AR Quiz game**

A smartphone on which an ARAF Browser is installed needs to be used to open and run the game file. A GPS signal and an Internet connection are needed in order to play the game. The game is an MP4 file that encapsulates the textual data (ARAF-compliant file), as well as additional media that is linked in the game file. The ARAF-compliant file consists of BIFS content in which the interface and the logic of the game are described. Being an AR location-based game, the ARAF Browser needs to access GPS data, as well as the orientation sensors of the device in order to place the Digital Objects at the correct positions in the augmented reality (camera) view.

### Augmented Printed Material

Augmented Printed Material is an AR application that enriches (physical) printed material with any digital media-like videos, images, sounds or 3D graphics. The application presents the user additional information related to the printed material that he or she is reading. An Augmented Printed Material application can enrich anything from a simple book to a tourism guide, a city map or a newspaper.



**Figure 14. Examples of augmenting printed content**

The application uses ARAF-compliant content that can be read by any ARAF Browser. The user opens the application and starts pointing the camera to the pages of the printed material that he or she is reading. Digital representations (e.g., screenshots) of the printed material pages to be augmented and their corresponding augmentation media are used as input to the application. As long as the application’s scanning mode is active, camera frames are processed and matched against target images submitted by the Content Creator. Once an image has been recognized, a tracking algorithm is then used to compute the relative position (pose matrix) of the recognized target image in the real world. The application is programmed to overlay the associated augmentation media on top of the recognized physical material as long as the physical printed material is being tracked.

# Conclusion

AR is a set of technologies that starts to be used in various domains. While is this incipient phase the need of standards was not a strong requirement, nowadays, in order to take AR to the next level, the interoperability is absolutely necessary. MPEG, by providing ARAF, enables this interoperability at the level of the data format supporting exchanges between AR system components and enabling creation of authoring tools to be used by AR experience creators.s

1. [↑](#footnote-ref-1)