|  |  |
| --- | --- |
|  | Moving Picture, Audio and Data Coding by Artificial Intelligencewww.mpai.community |

|  |  |
| --- | --- |
| **N307** | 2021/07/19 |
| **Source** | Adam Sobieski (Phoster) |
| **Title** | Generation, Streaming and Recording of Biomedical, Scientific and Industrial Sensor Data for Mixed-Reality Collaborative Spaces (MCS) |
| **Target** | MPAI Members |

# Proponent

Adam Sobieski (Phoster)

# Description

*Mixed-reality collaborative spaces* (MCS) are virtual environments where participants can work together on shared tasks.

Modern MCS software support collaboration across platforms and devices. Participants can, for example, utilize HoloLens, Quest, as well as their PCs and mobile devices while working together.

This MPAI proposal suggests that new standards and recommendations can equip MCS participants with access to live streams and recordings from biomedical, scientific, and industrial sensors and devices and also recommends that new standards and recommendations can equip MCS participants with access to such data as processed by artificial intelligence.

Providing participants with access to sensor and device data and to AI processing of such data in MCS environments will accelerate scientific progress and advance STEM education.

## Discussion: Current MCS Platforms

Microsoft Mesh is an example of the state of the art in MCS. Microsoft Mesh “provides a cross-platform developer SDK so that developers can create apps targeting their choice of platform and devices – whether AR, VR, PCs, or phones. Today it supports Unity alongside native C++ and C#, but in the coming months, Mesh will also have support for Unreal, Babylon, and React Native” [1].

Microsoft Mesh “supports most 3D file formats to natively render in Mesh-enabled apps, solving the challenge of bringing in users’ existing 3D models for collaboration” [1].

Alternatives to Microsoft Mesh include Adobe Aero, ApertusVR, Campfire, Cesium, GatherInVR, Lobaki, STRIVR, and Vectory Web AR.

# Comments

Under the MCS title we can consider three strands

## Artificial Intelligence Supporting Mixed-reality Collaborative Spaces

Artificial intelligence technologies can be utilized to provide digital avatars in MCS applications. These avatars can simulate or mimic the physical appearances of participants. These avatars are animated models aligned to the physical movements of participants. These avatars can have face animations and mouth animations which can be enhanced by processing speech audio.

Computer vision and artificial intelligence technologies can be utilized to scan physical objects and scenes and to produce virtual objects and scenes. These 3D virtual object models and scenes can then be utilized in MCS software applications.

Artificial intelligence technologies can be utilized to create 3D spatial maps of end-users’ environments for augmented-reality (AR) scenarios.

glTF [4] is a popular format for 3D resources. The topic of streaming 3D animations was recently discussed in a glTF GitHub issue [5]. There, Don McCurdy stated that “one case that is already available is to put different animations into different .bin files, downloading each animation when it is needed. Could be used for breaking an animation into chronological chunks, or lazy-loading animations in a game that aren't needed until the player unlocks them.” He continued, indicating that one would “need something considerably more complex than glTF to have one application reading from a file at the same time as another application is writing arbitrary data into it. That feels more like the role of an interchange format, perhaps. But I imagine someone could define an extension that allows open-ended buffers for animation samplers, allowing animation to stream in without fundamentally changing the structure of the scene.” [6]

Streaming of 3D data is being considered in several environments.

## Artificial Intelligence Using Mixed-reality Collaborative Spaces

Artificial intelligence technologies can

1. detect and analyze the facial expressions, speech, and the emotions of participants.
2. support speech recognition and natural language understanding for conversational user interfaces or transcripts.
3. support hand tracking, gesture recognition, and the identification of things pointed to in virtual environments.
4. support multimodal input recognition, recognizing combinations of speech and hand gestures.

## Artificial Intelligence and the Processing of Sensor and Device Data

With computer vision and artificial intelligence, the contents of 2D and 3D images and video can be recognized, per semantic segmentation, object recognition, event recognition, and activity recognition.

Metadata, such as information used to explain the scale of what is being viewed, can enhance live streams and recordings from biomedical, scientific, and industrial sensors and devices. Such uses of metadata could increase the quality of training data for, and, theoretically, improve the performance of, these algorithms.

Semantic segments in 2D and 3D images and video should have unique identifiers so that ancillary metadata tracks can describe recognized objects, events, and activities. This semantic metadata has many uses, e.g., facilitating indexing and searching for recordings by their contents.

# Examples

## Education

1. A science teacher utilizes a digital microscope to stream photorealistic 3D digital content to students in an MCS environment.
	1. This digital content could include content descriptors added by AI.
		1. In the example of a living cell, such content descriptors might include the cell nucleus, ribosomes, Golgi body, and mitochondria.
	2. The teacher can adjust physical and software controls of the digital microscope while immersed, without having to physically touch the digital microscope which may be at a different location.
2. One or more students in an MCS environment browse and interact with a large collection of photorealistic recordings from digital microscopes and other scientific sensors and devices.

## Biomedicine

1. A doctor analyzes medical data in their office, assisted by artificial intelligence.
2. Personnel at multiple medical laboratories collaborate in real-time using a MCS environment.

## Science

1. A scientist analyzes scientific data at their facility, assisted by artificial intelligence.
2. Scientists at multiple locations collaborate in real-time using a MCS environment.

## Industry

1. Personnel involved in training with computer vision algorithms in an MCS environment for industrial inspection scenarios view data from multiple sensors and algorithms as foods, parts, or products glide on a conveyor belt.

# Requirements

1. Capture, identify and stream digital representation of biomedical, scientific, and industrial objects, preserving their 3D nature.
2. Extract and stream descriptors of such objects.
3. Extract, identify and stream speech and moving picture information from humans that allows representation of animated avatars in an MCS environment.
4. Present the object(s) and their content descriptors in a MCS environments allowing users to interact with, zoom, rotate, and move (virtually or physically) the object(s) and to create and view sections of objects’ interiors by intersecting with planes.
5. Store the objects and content descriptors so that users may later be able to perform the actions with the recorded data.
6. Formats for streams and recordings should be suitable for input to and output from artificial intelligence components and pipelines.
7. Formats for streams and recordings should support semantic and metadata ancillary tracks.
8. Identified and described objects should have unique identifiers, be URI-addressable, for metadata.

# Object of Standard

1. Data formats for
	1. raw sensor data.
	2. processed/compressed sensor data (i.e., suitable for streaming and storage).
	3. content descriptors of sensor data in a form suitable for streaming and storage, e.g., compressed.
	4. two-way data to enable users to remotely control sensor devices.
2. Formats for human-related data
3. Streaming protocols (TBD)
4. Raw data types include, but are not limited to:
	1. imagery, video.
	2. light-field imagery, light-field video.
	3. RGB-D imagery, RGB-D video.
	4. point-cloud imagery, point-cloud video.
	5. 3D meshes, and mesh-based animations.
	6. volumetric data.

The proposed standard will enable interoperability between (1) biomedical, scientific, and industrial sensors and devices, (2) live streams and recordings from such sensors and devices, and (3) their presentation in MCS systems.

# Benefits

With new standards and recommendations, manufacturers of biomedical, scientific, and industrial sensors and devices and developers of software for their interoperability will have a clear view of how to make their sensors, devices, and systems interoperable with MCS systems.

With new standards and recommendations, developers of MCS systems will have a clear view of how to make their systems interoperable with live-streaming and recorded sensor data.

# Bottlenecks

TBD.

# Social Aspects

Providing participants with access to sensor and device data and to AI processing of such data in MCS environments will accelerate scientific progress and advance STEM education.

# Success Criteria

A success criterion is the utilization of new standards and recommendations by biomedical, scientific, and industrial sensor and device manufacturers, developers of interoperability software for them, and the developers of MCS systems.

# References

[1] <https://techcommunity.microsoft.com/t5/mixed-reality-blog/microsoft-mesh-a-technical-overview/ba-p/2176004>

[2] <https://www.khronos.org/openxr/>

[3] <https://immersiveweb.dev/>

[4] <https://www.khronos.org/gltf/>

[5] <https://github.com/KhronosGroup/glTF/issues/1238>

[6] <https://github.com/KhronosGroup/glTF/issues/1238#issuecomment-736221220>