**3GPP TSG-SA WG4 Meeting #127-bis-eS4-240707**

**Online, 8 – 12 April 2024**

**Source: InterDigital Canada**

**Title: Mapping MPEG-I Scene Description Technologies to Reference Architecture**

**Agenda item: 9.8**

**Document for: Discussion and Agreement**

1 Introduction

The technical report for the FS\_AVATAR study item (TR 26.813) documents a reference architecture for avatars in clause 7 and describes the main functional blocks in this architecture. The architecture is shown in Figure 1 below for completeness. This contribution describes how the reference architecture for avatars can be implemented using existing technologies from Khronos and MPEG for the use cases defined in the FS\_AVATAR study. More specifically, the contribution demonstrates how glTF 2.0 and the extensions defined in ISO/IEC 23090-14 (MPEG-I Scene Description) can be utilized by the different functional blocks of the reference architecture.



Figure 1 – Avatar reference architecture in TR 26.813.

2. Reference Architecture Implementation Based on MPEG-I Scene Description

The reference architecture for avatars described in clause 7 of TR 26.813 can be implemented using glTF 2.0 and some of the vendor extensions defined in the MPEG-I Scene Description specification (ISO/IEC 23090-14 [1]). These extensions define additional glTF 2.0 nodes that can be included in a glTF scene description document to allow the usage of avatars.

The following example demonstrates how the technologies defined in ISO/IEC 23090-14 can be mapped to the reference architecture for avatars, where the reference avatar defined in clause 6.2.1 (MORGAN) of TR 26.813 is used as a template avatar/model to illustrate the process of creating a user base avatar model for the use cases identified in TR 26.813. Figure 1 illustrates a potential implementation of the 3GPP reference architecture using the MPEG reference avatar model as a base model template.

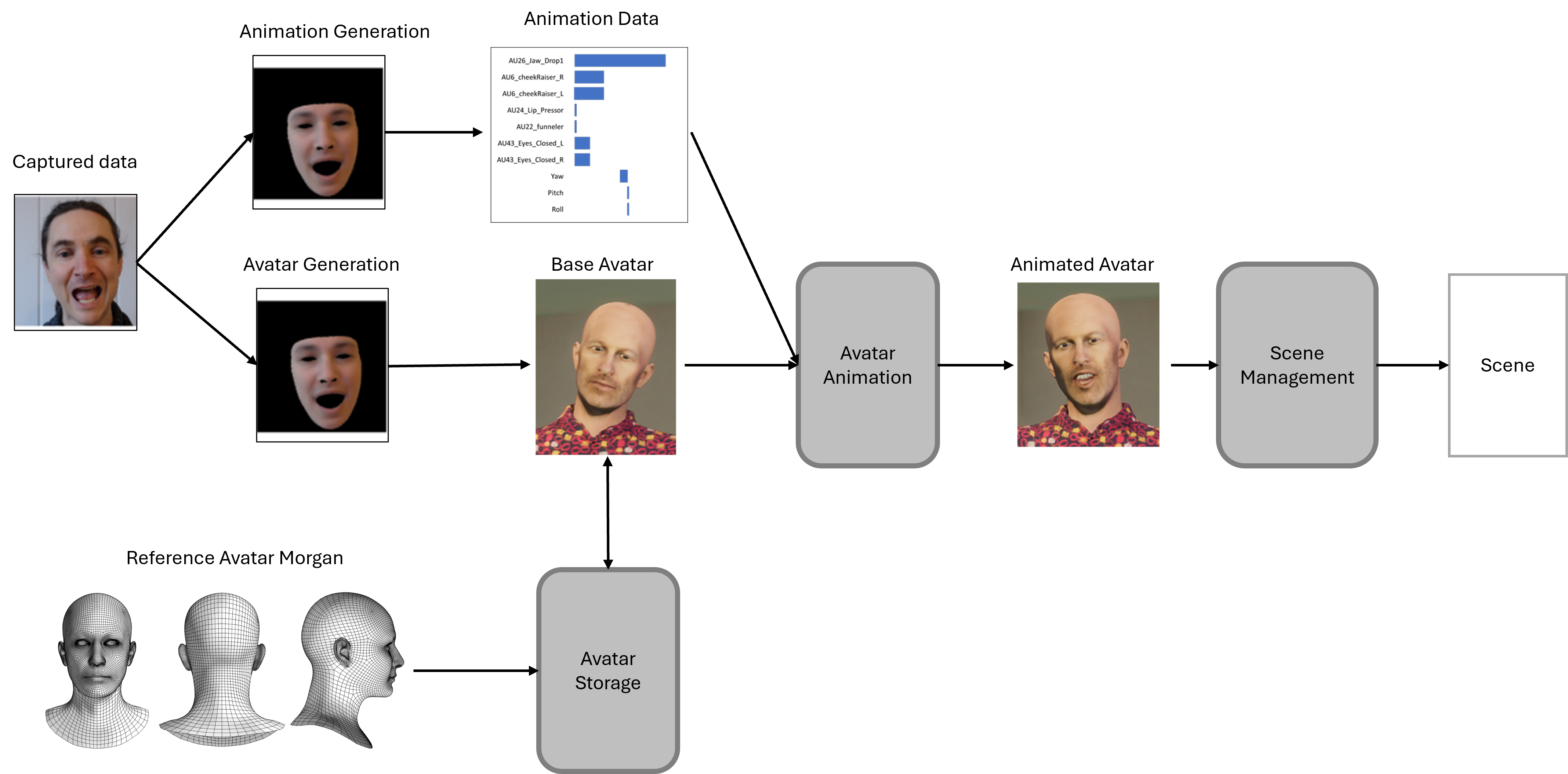


Figure 2 - Reference architecture using existing technologies in ISO/IEC 23090-14.

* **Base Avatar Generation:** The MPEG reference avatar representation (MPEG MORGAN) in clause 6.2.1 of TR 26.813 is retrieved from the Avatar Storage and used along with captured data to generate the base avatar representation. Traditional morphing or displacement algorithms can be used to convert the captured morphology to the Morgan template. The base avatar inherits all the meta information from MORGAN (semantic description, blendshapes specification, skeleton specification, etc.).

**Avatar Storage**: The reference avatar (i.e., MORGAN), which provides a standard template geometry with associated metadata (e.g., semantic description, blendshapes, skeleton) is stored in the avatar storage along with the generated base avatar.

* **Avatar Animation:** The animation stream is handled with glTF 2.0 and MPEG-I SD extensions to exploit the already existing functionalities that support basic streaming of animation data (e.g., vertex correspondence). Dynamic blenshapes’ weights and skeleton joints’ positions can be accessed through the MPEG\_accessor\_timed and MPEG\_buffer\_circular extensions and the MAF framework defined in ISO/IEC 23090-14.
* **Scene Management**: MPEG-I SD provides a solution for 3D scene management, including support for trackables and dynamic updates of scene objects. Using relevant extensions from ISO/IEC 23090-14, an application running on the UE or the AS is able position and update an avatar object within a scene.
* **Animation data generation:** The captured signal (here the video of the user) can be converted and translated into animation parameters (e.g., blendshape weights, skeleton joints position, and head pose). Keypoints detection and mapping to blendshapes and/or skeleton can be used for this purpose.

The following figure illustrates the existing technologies considered to answer the technical specification of the reference architecture.

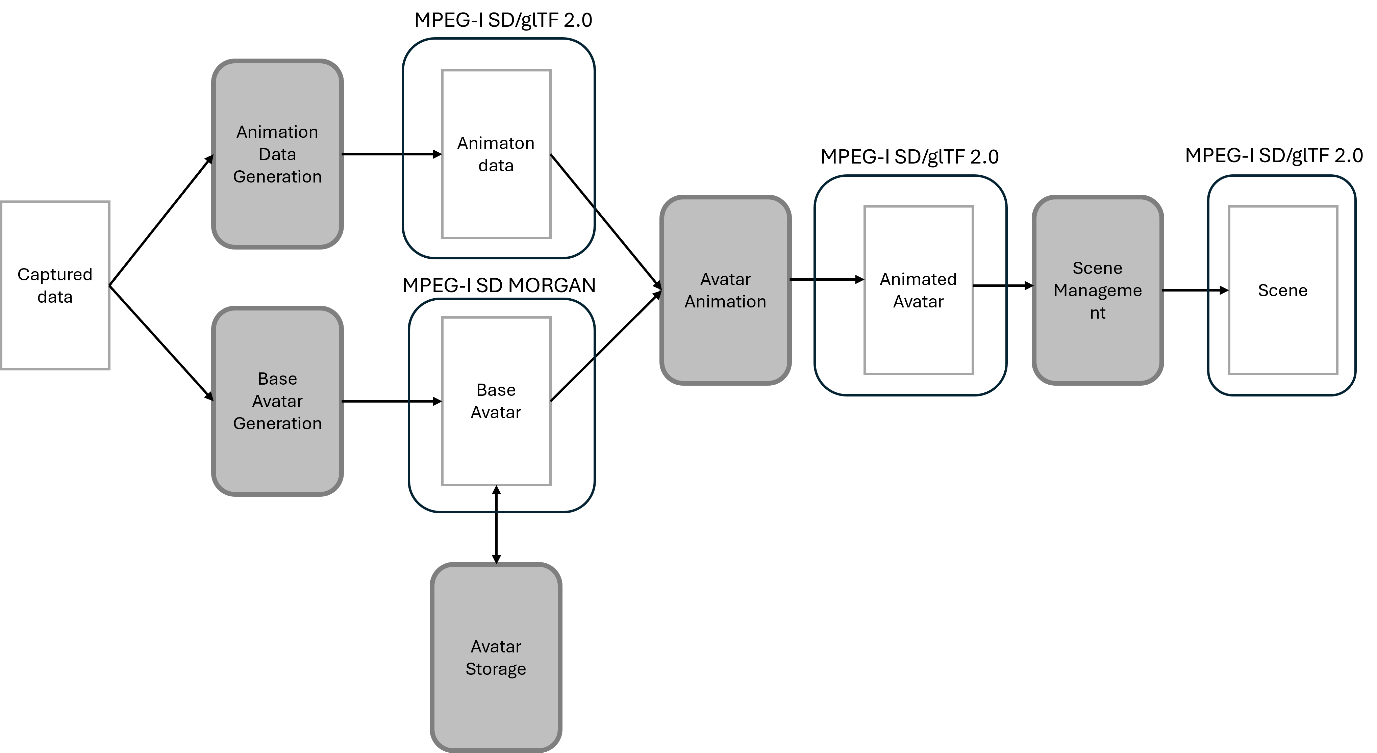


Figure 3 – Mapping of MPEG-I Scene Description technologies to the reference architecture.

3. Interoperability

We consider a use case where a user wishes to use their avatar model in use cases UC1 to UC3 in clause 5 of TR 26.813. It is assumed that a negotiation step takes place during the session establishment phase of the service and participating endpoints negotiate and configure the avatar model to be used. There are two possible scenarios:

* In a first scenario, the application running on the receiver’s UE has the same avatar model as the user’s and is therefore able to render and animate this model. In this scenario, the application on the sender UE simply sends the animation data corresponding to their avatar model to the application running on the receiver’s UE and this data can directly be applied to the model.
* In a second scenario, the remote application does not have access to the user’s base avatar model or may not be able to use that model. This may be because of resource requirements (e.g., memory and CPU), policy (e.g., to prevent the display of offensive avatars), or simply because it does not support the format of that model. In this scenario, a mapping is needed between the user’s base avatar model and that supported by the receiver application. Depending on the use case, this mapping may be obtained from the application on the user’s UE or loaded by a network function responsible for media conversion (e.g., MR or MRF). The mapping should be sufficient for supporting all features expected by the remote application. Using this mapping information, animation data corresponding to user’s base avatar model can be converted to the base avatar animation data used by the remove application.

The following figure illustrates the process described above for the two possible scenarios:



Figure 4 - Avatar model interoperability. UE1 is the sender and UE2 is the receiver (remote application).

4. Proposal

We propose to adopt the contents of sections 2 and 3 to the permanent document of FS\_AVATAR.

5. References

[1] ISO/IEC 23090-14:2023, Draft Text of ISO/IEC 23090-14 2nd edition. Online: https://www.mpeg.org/wp-content/uploads/mpeg\_meetings/144\_Hannover/w23220.zip