**3GPP TSG-SA WG4 Meeting #132 S4-250885**

**Fukuoka, JP, 19 – 23 May 2025**

**Source: China Mobile Com. Corporation**

**Title: [FS\_Beyond2D] AI-Generated Dynamic Mesh**

**Agenda item: 9.7**

**Document for: Agreement**

**1. Introduction**

This proposal introduces AI-generated 3D mesh as an emerging AIGC technology. However, it is not intended for evaluation in the Rel-19 FS\_Beyond2D study.

**2. Proposal**

It is proposed to agree the following changes to the 3GPP draft TR 26.956 V0.4.0.

\* \* \* First Change (All New) \* \* \*

#### 4.4.1 AI-Generated Dynamic Mesh

A growing number of AI-generated mesh tools now enable the direct generation of mesh models and textures from inputs such as text or images. Compared to traditional mesh production workflows, these tools offer significant advantages in terms of time efficiency. The examples of commercial services are provided below:

- AssetGen 2.0™: Meta's AI-powered 3D mesh generation system that produces models with "geometric consistency and fine-grained details" <https://developers.meta.com/horizon/blog/AssetGen2/>

- Hunyuan 3DTM: Tencent's 3D Mesh generation platform <https://3d.hunyuan.tencent.com/>

- MeshyTM: <https://www.meshy.ai/>

As the technology continues to advance, the quality and efficiency of AI-generated meshes are improving. However, there are still common issues that need to be addressed. including: excessively high polygon counts, poor topology, fragmented or irregular UV layouts, coarse texture details, baked-in lighting information in the textures, and insufficient accuracy in complex scenarios (e.g., clothing wrinkle simulation errors exceeding 15%).

#### 4.4.1.1 Image-Generated Dynamic Mesh

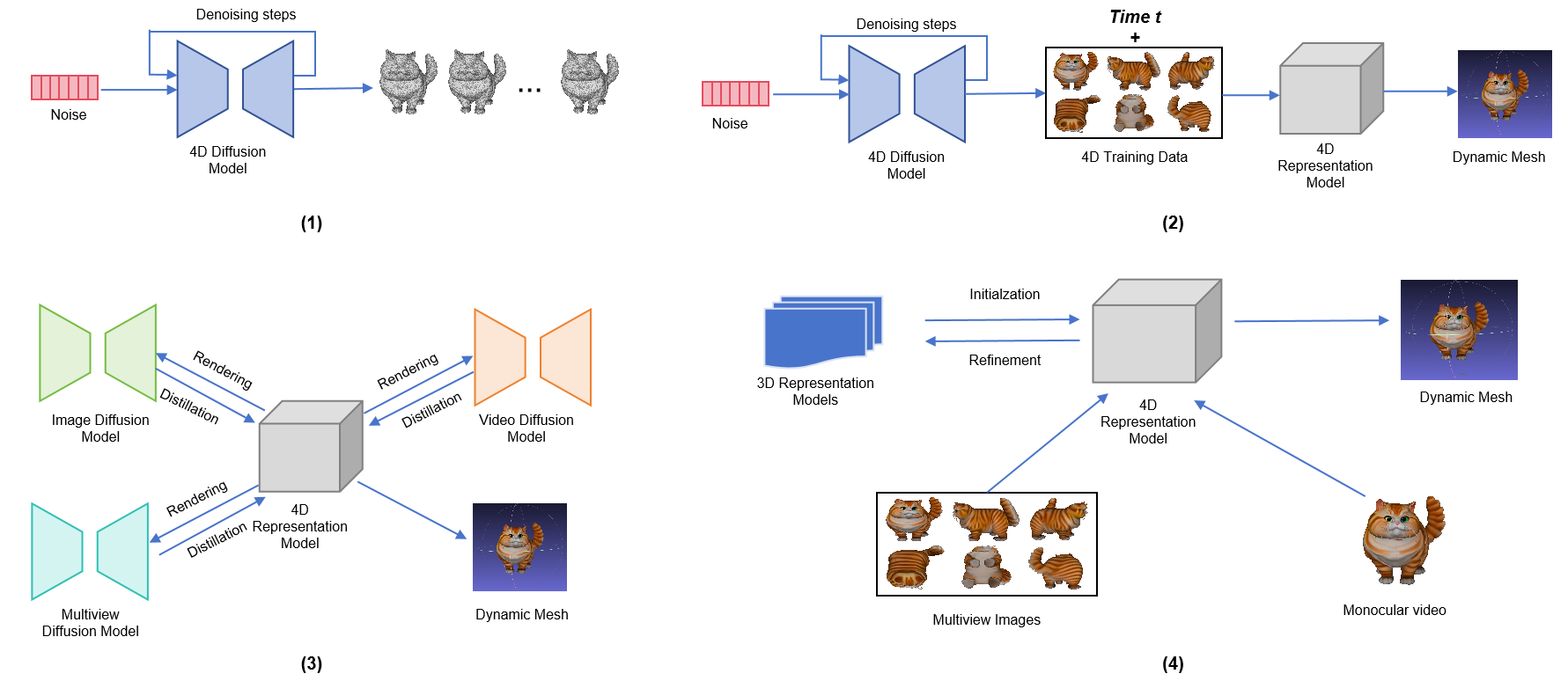
The task of generating dynamic meshes from images demands not only the creation of multiview geometric models based on the input image but also the extension into the temporal dimension to produce dynamic spatio-temporal content (4D). There are two main approaches for generating dynamic meshes, inference-based and optimization-based methods. As shown in Figure 7.2.4.1.1, the pipelines for these approaches include:

- Direct Generation: Directly generating dynamic meshes from input parameters without intermediate steps.

- Indirect Generation: Leverages diffusion models to produce multi-temporal and multi-view training data.

- Implicit Distillation: The process generates dynamic meshes through a multi-stage training framework, which combines multiple diffusion models via implicit distillation to derive generative priors.

- Explicit Supervision: Uses multi-modal data to provide explicit supervisory signals for dynamic mesh generation.



#### 4.4.1.2 Text-Generated Dynamic Mesh

Text-generated dynamic mesh requires both precise alignment between the object’s geometry and texture semantics, and accurate synchronization of its motion dynamics with describe actions or movements (4D). For example, a typical workflow may involve the following steps:

- Text Prompt Generation: Using large language models (e.g., GPT-4) to generate text prompts.

- Image Generation: Using diffusion models to generate single-view images based on these text prompts.

- Multi-view Synthesis:Levering video or multi-view diffusion models generate multi-view images from single-view images rendered from different angles.

- Dynamic Mesh Animation: Reconstruct 3D mesh and create a dynamic mesh (4D) by animating the vertices over time.

\* \* \* End of Changes \* \* \*