仅仅对第一帧生成其他3个角度的正交的图像，然后复制给其他帧。相当于组成一个BTVHWC维的向量，其中TV是帧数和视角数。

这篇文章有点没看懂他在干嘛，反正就是并没有对所有帧都用了多视图扩散模型，以及给非对称unet添加了两种注意力机制。这个unet是lgm中的，用来生成固定三维模型的。

所以相当于这篇工作的动态高斯的动态部分直接使用每一帧一个模型来弄的。