Lightweight Key Technologies for Massive Multi-player Online

WebVR Conferencing

单位信息：

作者信息：

摘要

关键字

1. **引言**

2019年底爆发的新冠病毒疫情在多个维度上改变了人们生活、工作、学习以及协作的模式，多种传统的线下行为在疫情隔离的需要之下迁移到线上环境中。此次疫情对社会运行与协作的模式的影响不仅暂时性的，同时也是持久性的，线上新的生活、工作、学习模式将在此次疫情之后得到重大的推进与普及。同时，基于疫情的影响各地政府部门也相继提出在线新经济发展模式，其归根结底是将传统的线下规划/设计/工程/审核/运维/商务/销售/管理等经济模式，经过轻量化处理后移植移动网页端，设计创新商业闭环后形成线上轻量级新经济模式，以适应用户激增后管理难度大、异地沟通成本高、大规模疫情爆发需自然隔离等社会新问题，进一步刺激并提升互联网经济的潜在活力，促进5G时代经济的新增长。

在疫情的影响之下，远程在线会议的需求激增，ZOOM会议、腾讯会议、钉钉会议等在线会议平台的使用量增幅巨大，但是目前的在线会议系统具有一定的限制性，主要有以下几点：

1. 当前的在线会议系统不具有可视化与社交特性。基于传统的音视频的方式无法生动的还原会议的实际情境，代入感较弱，没有整合VR、AR等新型的互动与展示方式。音视频的方式主要采用传统的图像与声音的信息传播方式，沉浸感比VR、AR相比大大减弱。
2. 当前会议系统的便利性及用户友好性较差。用户需额外安装插件，且人数受限，通常不能够很好地支持大规模的人群参会，部分在线会议系统较为重量级，对用户设备及网络环境要求高，会议加载速度慢，响应不及时等。

基于后疫情时代下的在线远程会议需求，想要解决当前远程会议系统存在的上述问题，主要有以下几个技术瓶颈：

1. 会议情景模型数据量大。为满足用户对在线会议的真实性、便利性等需求，给用户的提供较为良好的与会体验，势必需要大量精细的、高数据量的会议场景模型、人物模型、人物与场景动画等等，并且数据量会随着用户对场景美观度、交互方式的丰富度等的提高而进一步提高。
2. 会议情景中网络传输压力大。由于美观的场景模型、精细的人物动作带来的模型数据量的问题，会导致这些模型的传输出现比较大的网络延迟，从而对用户的与会体验造成负面影响。
3. 网页端渲染压力大。网页浏览器相对于桌面应用来说，受制于有限的内存空间与弱计算能力，不仅要渲染大规模、高数据量的会议场景、人物模型及动画等，还要时刻对用户的交互请求做出及时的响应。对渲染速度、渲染精度、资源占用等都有着极高的要求。

本文的研究目的是提出一套大规模轻量级WebVR在线会议技术解决方案，该方案采用虚拟现实的方式进行远程会议环境的呈现，给体验者带来更强的参与感与沉浸感，使得线上与线下的距离进一步拉近，采用Web+VR的方式进一步减低VR的体验与使用门槛，用户无需安装任何插件，即可快速上会，并且支持大规模的人群在线参会，同时给用户提供丰富的、友好的、快速的与会交互操作，提升用户使用的友好度与易用性。

1. **相关研究工作**

**2.1 VR在线会议系统研究现状**

2020年，可以说是各大厂商开始频繁地将VR相关技术应用在展览展会、线上会议中的一年。从谷歌、苹果、Facebook、MWC、E3、Unity等等每年备受关注的大会纷纷宣布改为线上举办后，利用科技手段在线上举办发布会逐渐形成一种潮流。

当下，对VR在线会议系统的研究主要分为两大类。第一类是对基于PC客户端的VR会议平台的研究，如VSWork[1]、Engage[2]等。第二类是对基于Web端的在线社交会议平台的研究，如Mozilia hubs[3]。

VSWork是国内的一款VR虚拟会议系统，其实现了丰富的用户交互操作，如自定义虚拟角色，人物模型与场景物体的互动等。在软件架构方面，VSWORK的全部内容都采用云端部署，终端设备的本地负担很小，不同客户的内容都对应独立的服务器，保证了用户文件与信息的安全性与私密性。Engage由VR教育公司IVRE开发，同样支持与会者一对一的丰富交互操作，同时，其对VR设备也有着较好的支持。

Mozilla hubs是运行在浏览器中的虚拟协作平台。其技术优势在于无需安装其他插件，通过链接邀请他人进入自创的虚拟空间中进行相关的交互操作，并提供对VR设备的支持。并且，其拥有着高自由度的化身定制、场景定制功能，以及高自由度的交互方式。不过在部分场景下，场景的美观度、加载速度上仍然较慢。

总的来说，基于PC的VR在线会议系统，大多是采用Unity进行开发，较为重量级；发布模式仍是PC端运行的可执行程序，用户使用门槛较高；支持有限度的互动与远程协助即使是目前像Mozilla hubs这样的WebVR会议平台，虽然有着更加轻量级、用户交互操作更多等特点，但是在某些场景的美观性和精度不足、加载速度上较慢，卡顿较为频繁，仍然不足以为用户提供一个较为良好的与会体验。

2.2 大规模轻量级WebVR在线会议关键技术

**大规模会议情景的轻量化预处理**

大规模在线会议由于参会人数众多，会议场景、人物模型数据量非常庞大，巨量数据会对网络条件、用户配置及浏览器渲染速度造成非常大的压力。因此，降低会议情景中的数据量方法就变得十分必要。刘[4]等人提出了一种基于WebBIM场景的数据轻量化方法。孙[5]等人提出了一种基于内容的可比较压缩算法，还有一些方法着重考虑模型数据之间的相似性，并将这种相似性与差异性进行度量，从而达到压缩数据的目的[6][7].特别地，许多学者将语义信息引入了模型检索过程，提出了一些基于语义的检索与冗余数据检测的间接性方案[8][9]。还有一些学者，在语义信息之外，还将几何处理方法纳入了他们的解决方案中，如laga[10]等人与zheng[11]等人，在几何识别中加入了一些语义规则.

由于模型组件之间的相似性，我们可以进一步利用它来提高效率。模型的预处理当然可以缓解问题，例如3D压缩和渐进式网格（PM）。但是他们都没有考虑模型组件之间的相似性，因此我们可以利用它来进一步提高效率。Wen[12]等人提出了一种类似感知的数据约简方法，称为轻量级渐进网格（LPM）。 LPM旨在挖掘模型中的相似组件，在删除冗余组件后生成剩余每个组件的PM表示，并使用称为轻量级场景图的结构组织所有处理后的数据。这种方法具有四个显着优势。首先，它可以极大地减小模型的文件大小，而几乎不会造成任何精度损失。其次，PM使传递变得渐进，即所谓的流式传输。第三，在客户端进行渲染时，由于使用了轻量级的场景图，因此无需进行解压缩，并且可以充分执行实例渲染。第四，它在非常有限的带宽下非常有效，尤其是在交付大型场景时。

**大规模人群轻量化预处理**

当下，大规模人群的可视化技术通常是通过基于代理的模拟技术实现[13]。

使用单一模型的缺点是加载时间可能很长，一旦人物数量增加，帧率就会显著下降。之前的方法着重于简化模型本身或提高硬件的性能。通常，这些方法会忽略大规模人群仿真的多样性和真实性。 目前，模型参数化的概念已应用于Web3D建筑物信息建模（BIM）中的建筑物简化中[14]。并且，GPU实例机制和LOD（详细级别）技术已经可以用于实现数千个化身的实时渲染，但模拟化身是同质的[15]。大多数研究都没有考虑多样性和现实因素。因此，一种在保证多样性和真实性要求的情况下，进一步扩大人群规模来灵活提高效率的方法是非常有必要的。朴[16]等人提出了一种基于形状空间的参数化技术来对化身的外观进行多样化。同时使用异步传输渲染元素，助于减少带宽压力。其提出的多级克隆实例化方法可以在短时间内生成大量的异构化身。

**细粒度化渐进式传输调度**

由于在线会议情景需传输大量模型及动画数据，为了避免出现比较大的网络延迟，减轻对用户的与会体验所造成的负面影响。有学者尝试使用边缘计算方式来代替集中式云计算，如李[17]等人提出的一种新的云端边缘浏览器计算框架与大型场景数据自适应传输调度算法。该算法通过综合用户视角的重复度、填充程度与所注意的场景来定义兴趣度，基于计算出的用户兴趣来优化传输机制、进行路径预测，以实现网络传输的最佳效果。与云计算的集中式结构不同，边缘计算采用分布式结构[18][19][20]，通过在本地设备端附近部署几个计算节点。来自本地设备的数据直接传输到边缘节点进行计算和处理。与集中式云计算相比，这种转变提高了服务器的响应速度和可靠性。侯[21]等人实现了轻量级VR眼镜及边缘云设备之间的无线连接，并在这些设备之间实现了远程渲染，并将各种解决方案在延迟和其他方面进行了比较。最终证明，通过其分散的结构，它对于Web3D可视化的实时计算和处理更加有效。

1. **总体技术路线**

志成和恩旸在这里补充一个技术路线图（突出几大关键技术），尝试画一下（用ppt来画图），大家一起来补充完善。

轻量化处理的核心是资源重用，我们使用该技术必须首先明白哪些资源是可重用的，并且尽可以提高资源的重用度。另外为了效果要保留丰富的多样性，参数化调整每个对象，另外可以通过对不同资源进行搭配组合来提高多样性。

首先将模型划分为多个区域（在这个问题中是划分了头部、上身、下身三个区域），每个区域可以匹配不同贴图（在这个问题中使用了男性32套贴图，女性16套贴图），这样就可以通过不同贴图搭配各种效果（男性有32^3=32768种组合，女性有16^3=4086种组合，共36864种组合）。另外，可以通过对人物对象的高矮胖瘦进行设置，还可以对对象的色调进行编辑（这个问题中主要是对下身的裤子颜色进行设置），在骨骼动画方面，可以为每个人物对象设置不同的动画播放速度。

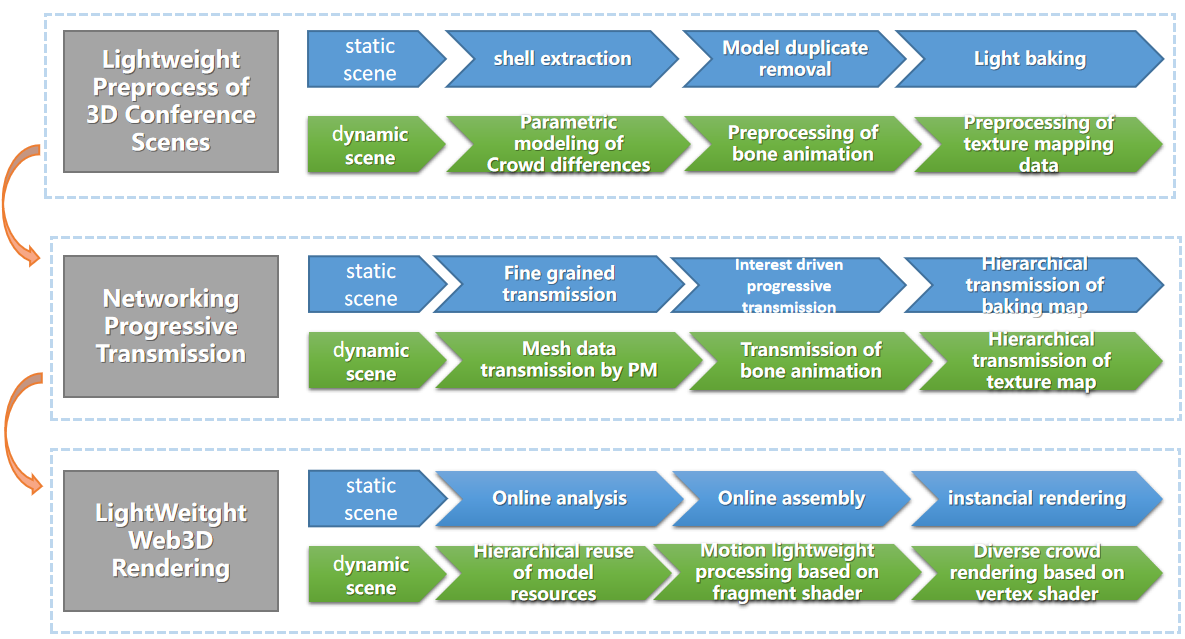


图 1 总体技术路线

1. **关键技术之一：大规模会议情境的轻量化预处理**

志成写一段引文

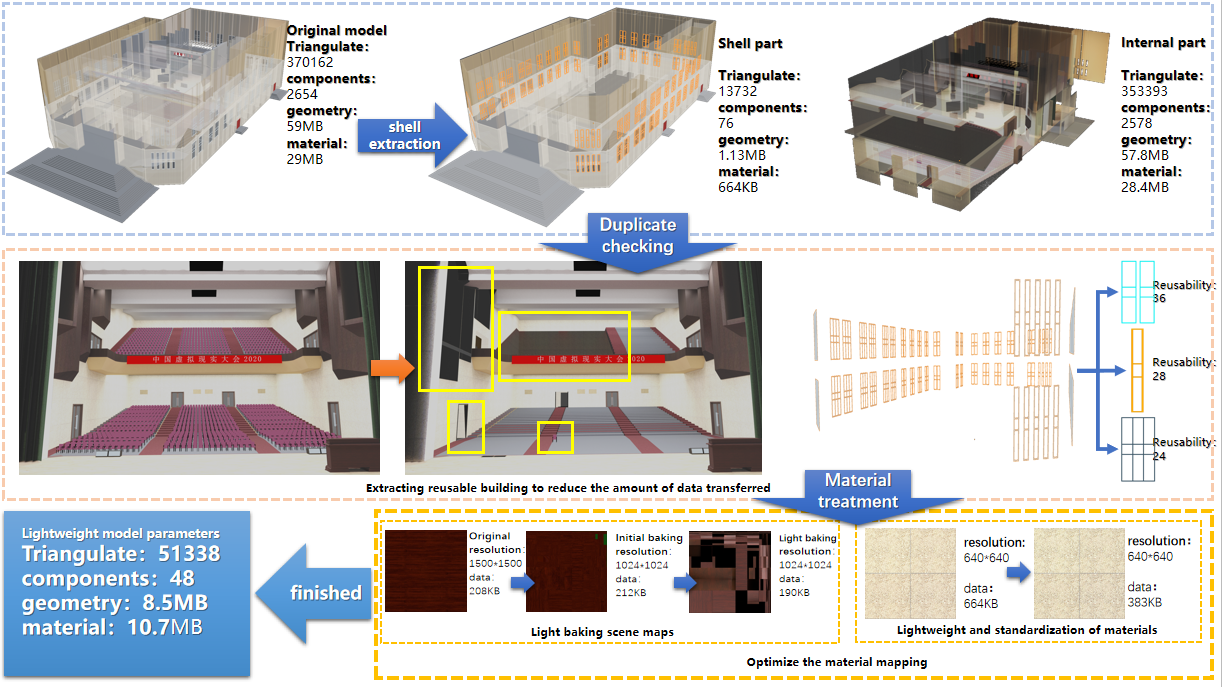


图 3 由多个小部件组合成的复合模型的轻量化处理思路

4.1**静态会场场景的轻量化预处理**

志成完善下列文字

大规模人群渲染需要存储的信息可以分为三级。第一级，所有对象都共用的数据，如模型的网格信息（网格点位置，UV等信息）；第二级，选择性使用的数据（如纹理贴图），这类数据每个对象只需要其中的一部分数据（如模型贴图有多套可供选择，但每个对象只使用其中的一套）；第二级，描述对象的参数信息（如，贴图类型，动画播放速度，高矮胖瘦，色调），这些信息每个对象都可以不同，所以每个对象的这些参数信息都要单独存放。

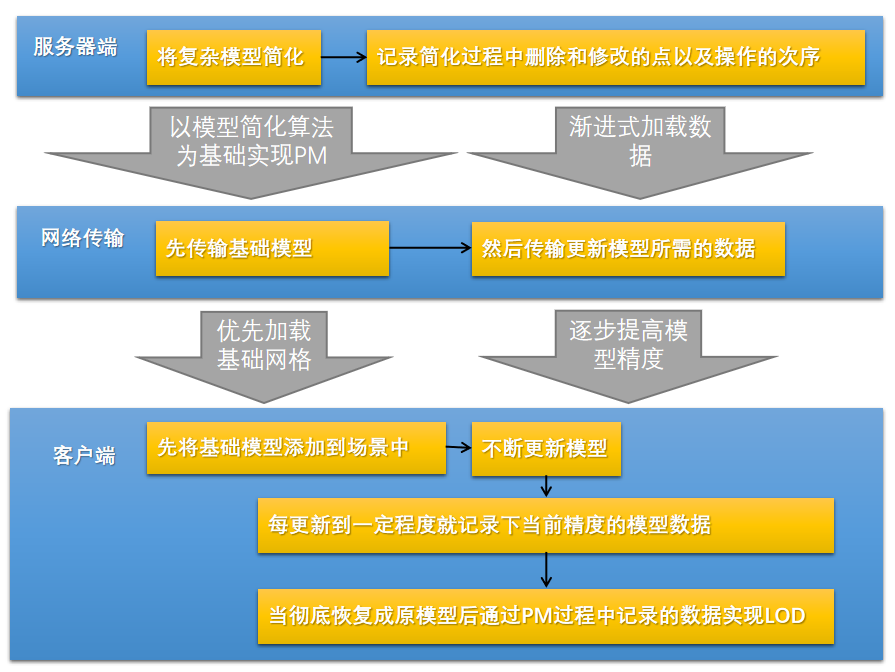


图 4 高精度模型处理（PM+LOD）

4.2**大规模参会人群的轻量化预处理（参考朴雪论文中相关内容）**

志成/恩旸完善这部分文字

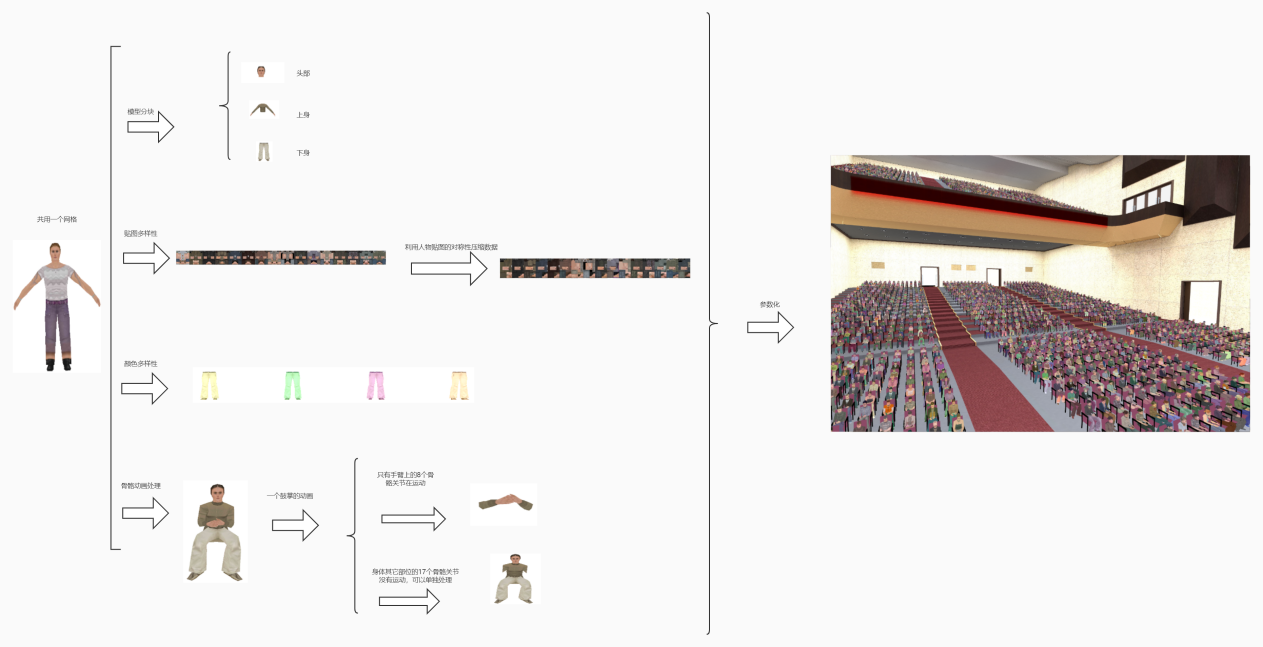


图 5 大规模人群轻量化处理技术的运用

在项目中我们只需要实现听众鼓掌的动作，这个动作只涉及到手臂的8个骨骼，其它的17个骨骼的状态没有发生变化，所以首先可以将骨骼数据分成两部分，手臂处骨骼和固定动作的骨骼。

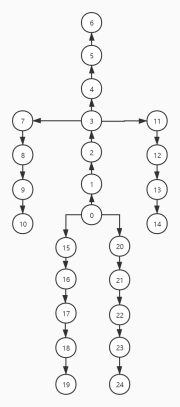


图 6 人物模型骨骼结构的分析（未完成）

每个骨骼的变换矩阵因为最后一行是固定值，所以一个变换矩阵只需要存储12个float数据。有17个骨骼在动画播放的过程中的状态没有发生变化，所以这可以提前计算好这17\*12=204个数据。涉及到手臂的8个骨骼，由于我们在项目中只需要实现鼓掌这个简单的动作，经过测试使用16帧就可以获得比较好的动画效果，这是16帧的鼓掌动画中后8帧可以看作前8帧的倒放，所以我们只需要8个骨骼8帧分别的变换矩阵，这8个骨骼需要的数据量为8\*8\*12=768。将这些数据在预处理阶段直接计算好可以减少在客户端的计算量。

**增加对象间的差异性**

**需求**

为了使得人群渲染的多样性效果更加明显，我们希望使用相同头部贴图的模型距离越远越好。

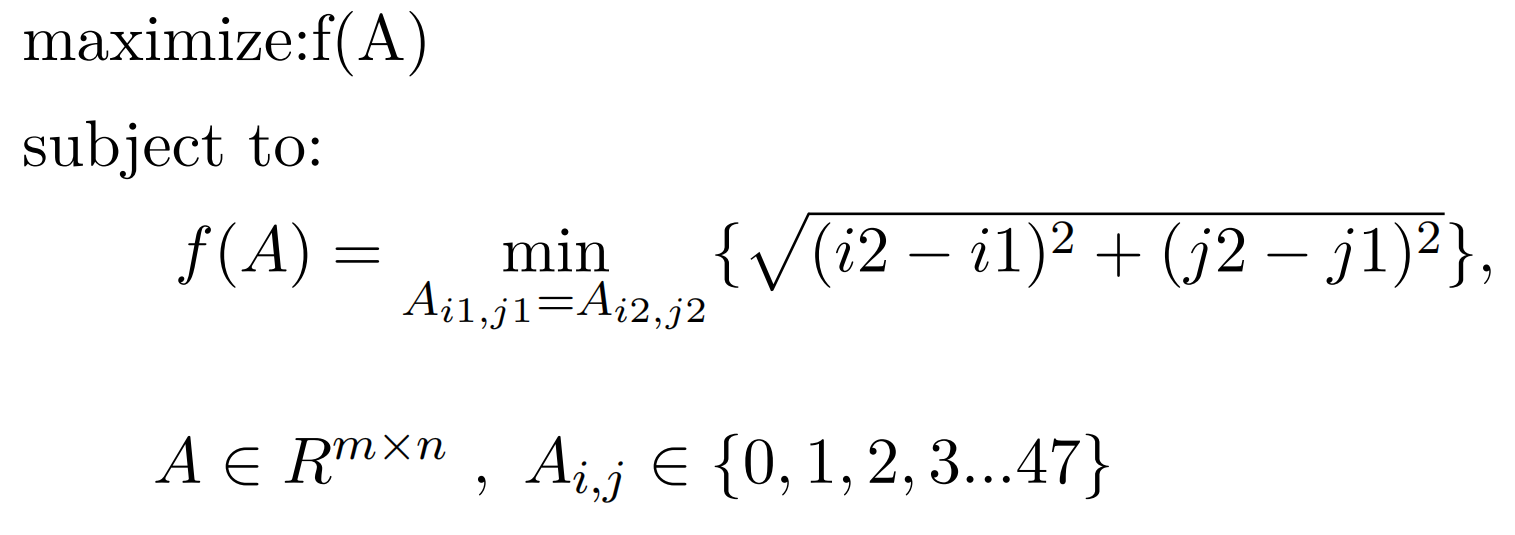
**问题分析**

可以将这一需求抽象为以下数学问题：

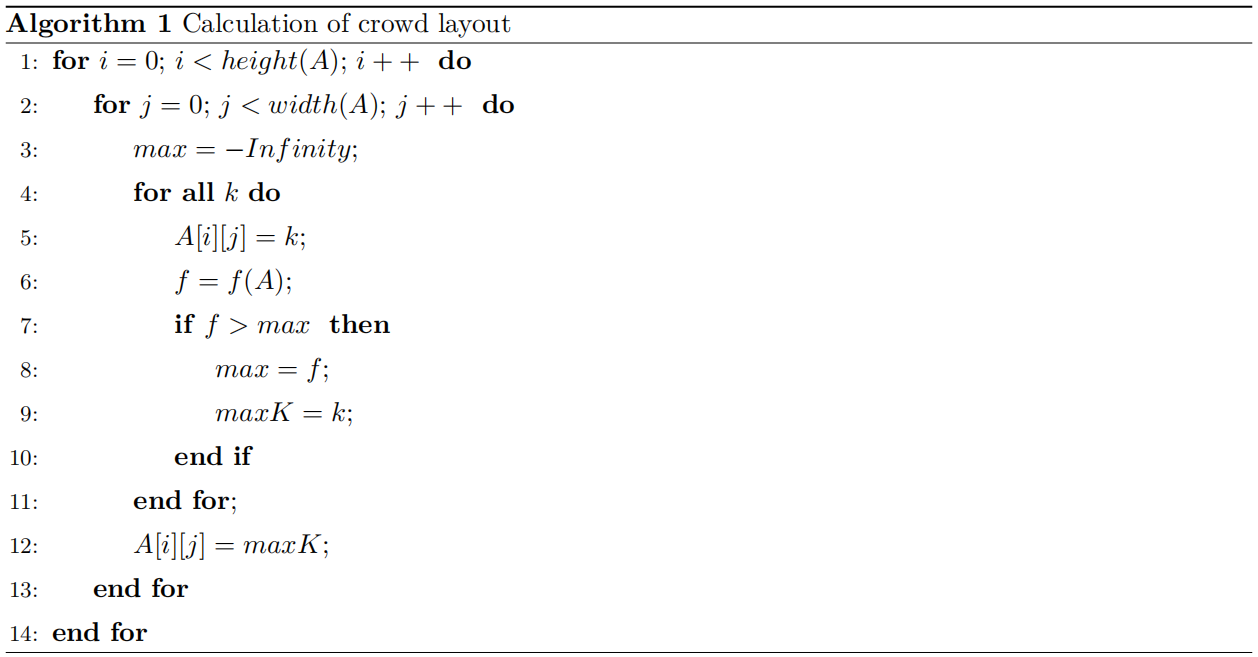
人和椅子位置相同，椅子的摆放为规则的矩阵，可以用一个长宽固定的矩阵A记录信息。头部贴图共用48种，可以用0-47这48给整数表示矩阵A中每个元素。

我们定义一个叫**分散度**概念，用来描述相似对象的分散程度。分散度的值为相同类型相同的元素间距离的最小值。我们要求一个最佳的人群设置方式A，使得分散度最大。

可以将这个问题表示为下面的公式形式，其中矩阵A为人群设置方式，f表示分散度，i和j表示矩阵的行号和列号：



**解决问题**



**结果分析**

我们进行了1000次不经过处理的测试，分散度f的值始终为1，既存在相邻的两个人物对象贴图相同。经过我们的处理后分散度约为5.831，所以在我们的场景中任意使用相同贴图的人物对象之间的最小距离是5.831，他们之间相隔4人以上。

1. **关键技术之二：细粒度化渐进式传输调度（参考李柯论文相关内容）**

5.1轻量化缓存管理

**志成这一块**

将场景中需要的数据细化拆分，相同部件只传输一次。

分级传输，提高初始加载速度。人物贴图进行了多细节层次处理，先传输低像素的纹理贴图，再传输高像素的贴图。

5.2基于兴趣度的细粒度化在线打包

**志成写这一块**

5.3 带宽自适应的渐进式传输调度

志成写这一块

为了确保复用的人物对象资源不被重复存储，需要将场景中的人群统一管理，我们用JS设计了以一个对象来管理整个的人群。

为了提高渲染效率，所有的人物在渲染时作为一个整体，一次性将人群所需的全部数据输入着色器。这就需要我们管理好每个人的所有参数信息（位置，大小，贴图类型，等等），渲染时一次性将这些信息传入shader。

**6 轻量级大规模会议场景在线渲染**

志成完成这一块

骨骼动画数据的输入格式是32的float浮点数，经过测试发现骨骼动画数据实际上并不需要这么高的精度，所以我们设计了一种16位的浮点数格式，这样就使得传入着色器的骨骼数据量减少了一半。

项目中人物的贴图左右对称，所以可以只传入着色器左半部分贴图，这样传入着色器的贴图数据量就减少了一半。

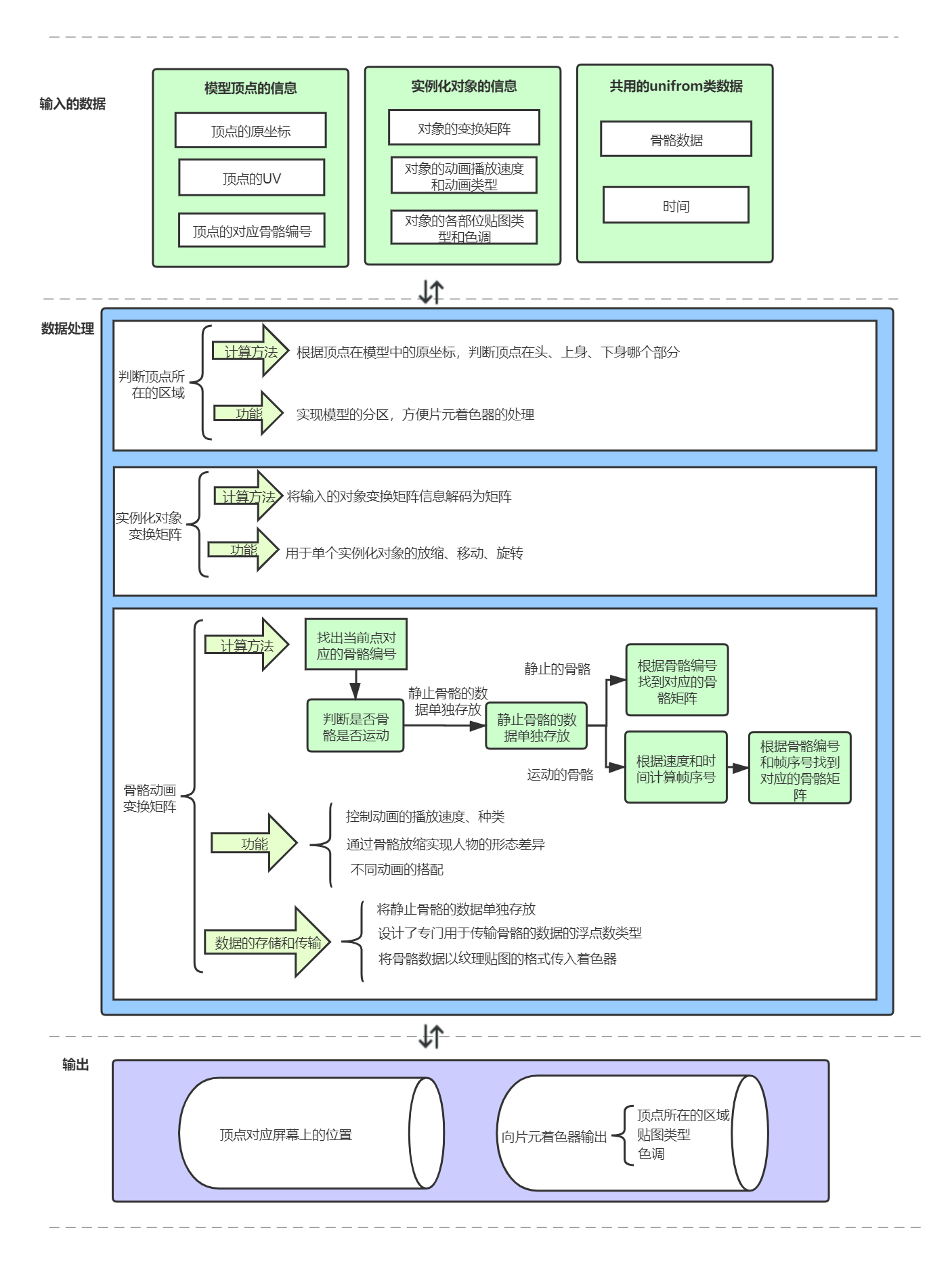


图 7 顶点着色器的优化处理

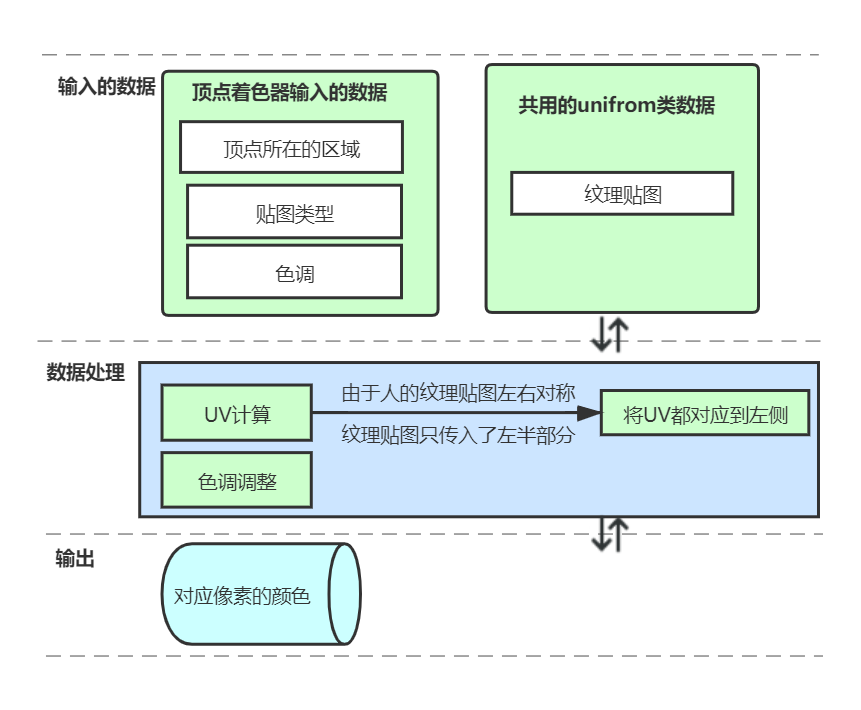


图 8 片元着色器的优化处理

5.实验结果与性能分析

志成/恩旸搞这一块

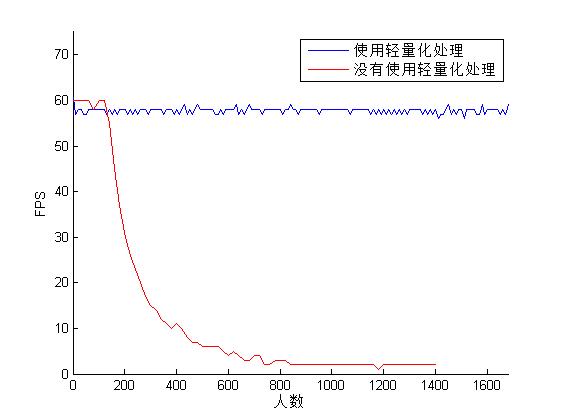


图 9 人群的轻量化处理技术在项目中的效果



图 10 大规模人群的轻量化处理的效果

1. 结论

志成/恩旸搞这一块

1. Reference
2. Reference

[1] VSWORK虚拟空间[EB/OL] <http://www.vswork.com/>

[2] ENGAGE Virtual Reality Education & Corporate Training. VR Education Holdings PLC [EB/OL] https://engagevr.io/

[3] Mozilla Hubs [EB/OL] <https://hubs.mozilla.com/>

[4]Xiaojun Liu, Ning Xie, Kai Tang, Jinyuan Jia. “Lightweight for Web3D Visualization of Large-scale BIM Scenes in Real-time”. Graphical Models, vol. 88, pp. 40-56, November，2016.

[5] G. G. X.-G. H. Jing Sun, Yu-Shen Liu, IFCCompressor: A

content-based compression algorithm for optimizing Industry Foundation

Classes files, Automation in Construction 50 (2015) 1–15.

doi:10.1016/j.autcon.2014.10.015.

[6] J. L. G. Arthaud, Automatic semantic comparison of step product models,

in: Innovations in Design & Decision Support Systems in Architecture

and Urban Planning, Springer Netherlands, 2006, pp. 447–463.

doi:10.1007/978-1-4020-5060-2 29.

[7] S. H. Y. S. Ghang Lee, Jongsung Won, Metrics for quantifying the

similarities and differences between ifc files, Journal of Computing in

Civil Engineering 25 (2) (2011) 172–181. doi:10.1061/(ASCE)CP.1943-

5487.0000077.

[8] R. R. A. I. Le Zhang, Development of ifc-based construction industry

ontology for information retrieval from ifc models, in: Proceedings of

the 2011 EG-ICE Workshop, University of Twente, The Netherlands,

2011, pp. 6–8.

[9] M. W. M. G. J.-H. Y. Ge Gao, Yu-Shen Liu, A query expansion

method for retrieving online bim resources based on industry

foundation classes, Automation in Construction 56 (2015) 14–25.

doi:10.1016/j.autcon.2015.04.006.

[10] M. S. Hamid Laga, Michela Mortara, Geometry and context for semantic

correspondences and functionality recognition in man-made 3d shapes, ACM Transactions on Graphics (TOG) 32 (5) (2013) 150.

doi:10.1145/2516971.2516975.

[11] M. A. N. J. M. Youyi Zheng, Daniel Cohen-Or, Recurring part arrangements

in shape collections, Computer Graphics Forum 33 (2) (2014)

115–124. doi:10.1111/cgf.12309.

[12] Laixiang Wen, Ning. Xie, and Jinyuan. Jia. “Fast accessing Web3D contents using lightweight progressive meshes”. Computer Animation and Virtual Worlds. 27(5): 466-483, May, 2016.

[13]Oğuzcan Oğuz, Ateş Akaydın, Türker Yılmaz, and Uğur Güdükbay. 2010. Emergency

crowd simulation for outdoor environments. Computers & Graphics 34, 2 (2010),

136–144.

[14]Alessandro Pluchino, Cesare Garofalo, Giuseppe Inturri, Andrea Rapisarda, and Matteo

Ignaccolo. 2013. Agent-based simulation of pedestrian behaviour in closed spaces:

a museum case study. arXiv preprint arXiv:1302.7153 (2013).

[15]Fengting Yan, Jinyuan Jia, Yonghao Hu, Qinghua Guo, and Hehua Zhu. 2019. Smart

fire evacuation service based on Internet of Things computing for Web3D. Journal

of Internet Technology 20, 2 (2019), 521–532.

[16]Daniel P Savoy, Marcio C Cabral, and Marcelo K Zuffo. 2015. Crowd simulation

rendering for web. In Proceedings of the 20th International Conference on 3D Web

Technology. 159–160.

[17]Ke Li, Qian Zhang, Hantao Zhao and Jinyuan Jia. User Interests Driven Collaborative Cloud-Edge-Browser Architecture for WebBIM Visualization. ACM Web3D, Nov. 13-15, 2020.

[18]M Satyanarayanan. 2017. The Emergence of Edge Computing. IEEE Computer 50, 1(2017), 30–39.

[19]Tianchu Zhao, Sheng Zhou, Xueying Guo, Yun Zhao, and Zhisheng Niu. 2015. ACooperative Scheduling Scheme of Local Cloud and Internet Cloud for Delay-Aware Mobile Cloud Computing. (2015), 1–6.

[20]Weisong Shi, Jie Cao, Quan Zhang, Youhuizi Li, and Lanyu Xu. 2016. Edge Computing:Vision and Challenges. IEEE Internet of Things Journal 3, 5 (2016), 637–646.

[21]Xueshi Hou, Yao Lu, and Sujit Dey. 2017. Wireless VR/AR with Edge/Cloud Computing.(2017), 1–8.