

# Cave VR 直播系统的手势控制系统

张训雷, 帖云, 齐林

(郑州大学, 郑州 450001)

**摘要:** 在 5G 不断商用背景的推动下, 虚拟现实 (Virtual Reality, VR) 技术将在生活中的诸多场景中具有广阔的前景, 并且目前基于 VR 技术已有诸多应用。然而, 若需要在虚拟场景中实现近乎真实的体验, 用户指令的输入和系统的实时反馈显得尤为重要。目前在 VR 场景中的主流指令输入方式通常由 VR 头戴设备相搭配的手柄等硬件实现, 由于用户需要利用手动的方式操作手柄以实现对相关动作的实现, 从而降低了用户在 VR 场景中的沉浸式体验。为了使得 VR 场景适用于更加自主化和现实化的需求, 本文在直播平台的支持下, 开发了基于 Leap Motion 的手势控制系统, 并且实验结果表明了所设计的手势控制系统满足了在直播场景下的基本需求。

**关键词:** Cave 系统; VR 直播; Leap Motion; 手势控制系统

**中图分类号:** TP37 **文献标识码:** A

## Gesture Control System for Cave VR Live Streaming System

Zhang Xunlei, Tie Yun, Qi Lin

(Zhengzhou University, Zhengzhou, 450001, China)

**Abstract:** Driven by the background of continuous commercialization of 5G, virtual reality (VR) technology will have a broad prospect in many scenes of life, and there are already many applications based on VR technology. However, the input of user commands and real-time feedback from the system are particularly important if a near-real experience is needed in virtual scenes. At present, the mainstream command input in VR scenes is usually achieved by the VR headset with the handle and other hardware, which reduces the immersive experience of the user in VR scenes because the user needs to operate the handle manually to achieve the relevant action. In order to make the VR scene suitable for more autonomous and realistic needs, this paper develops a Leap Motion-based gesture control system with the support of a live streaming platform, and the experimental results show that the designed gesture control system meets the basic needs in the live streaming scene.

**Key words:** Cave System; VR Live Streaming; Leap Motion; Gesture Control System

### 1 引言<sup>1</sup>

随着移动宽带传输技术的不断成熟, 促进了宽带流媒体业务的极大发展。尤其是VR技术成为当前

---

作者简介: 张训雷 (1994 - ), 男 (汉族), 安徽滁州人,

郑州大学硕士研究生, 1602512235@qq.com。

产业化的热点方向，为了进一步规范其发展，工业和信息化部主办了世界VR产业大会，并制定了《虚拟现实产业发展白皮书(2019年)》，根据最新的统计数据表明在未来几年内，VR产业将直接或间接产生上万亿的市场效益。由于VR技术可以实现足不出户就能体验到近乎真实的各种生活场景，如购物、旅游等。此外，VR技术还为虚拟网络资源加入了沉浸式的体验，使得用户可以享有身临其境的体验，从而极大提升用户使用体验。

VR技术的概念和理论从20世纪70年代开始形成，首先，由Dan Sandin等研制出数据手套<sup>[1]</sup>实现了在虚拟环境中对用户的手部动作信息的实时捕捉和跟踪；其次，美国国家航天局为了满足对太空研究的需求开发了VIEW系统<sup>[2]</sup>，该系统实现了在虚拟环境中对火星的探索任务，并记录了相关的实验数据从而促进了后期的研究工作；最后，由VPL Research公司以及任天堂游戏公司发布了基于VR技术的相关软硬件，从而推动了VR技术的商用化进程。但是，由于VR技术对硬体的处理性能和网络的传输性能要求较高，因此VR技术陷入了发展的瓶颈。

近年来，随着计算机硬件和软件的高速发展，尤其是数据传输技术的不断发展，从而实现了网络内容近乎无损的传输，从而促进了VR技术进入新的发展时期。并且相关厂商为推动VR商业化的进程，相继推出了面向消费者的VR产品及配套设备，如HTC Vive、Facebook Oculus等，此外相关厂商也为其VR设备建立了相关的生态，使得其内容和体验的不断提升。

## 2 VR 控制系统研究现状及挑战

VR技术的开发与应用通常应具备沉浸式、交互性以及自主化的特点。相较于传统的技术发展，由于VR始终基于虚拟场景和模型以实现逼近真实的体验感，所以如何让用户与所处的虚拟场景融为一体就显得尤为重要。但是目前VR技术的应用主要存在VR控制方式非自然化、媒体内容解析度受限、刷新率偏低以及观看视场角范围较小的挑战。直播技术作为一种新兴的异地场景远程云端共享的形式，目前主流的方式有：1) 依托于网页Web技术的桌面级别的产品以及2) 基于VR技术的特定场景应用。前者依托于Web技术虽然实现了流媒体内容的异地共享和观看，但由于其始终在桌面环境以视频的形式为用户呈现远端的环境，从而极大地影响了

直播的效果，并最终降低了用户的使用体验；然而，VR技术通过模型虚拟场景搭建技术和全景视频渲染技术，通过借助于相关硬件和软件平台的基础上不仅为用户带来了云端实景体验，而且在一定程度上实现了流媒体的高解析度接收，目前主要的VR开发与应用平台有HTC Vive以及Oculus。

目前在VR视频的传输和接收领域已经有相关研究人员做了大量的研究工作，并也取得了显著的效果。然而，如何对VR的内容进行沉浸化的控制和操作仍需要做更为深入的研究工作。一般的都是通过与相关Vive或Oculus等设备相配套的控制手柄实现对虚拟场景的控制<sup>[3-5]</sup>，通过使用配套的手柄虽然极大的简化了开发的流程，但由于手柄已对相关动作的触发进行了指定，使得用户并不能任意的完成操作指令，从而限制了用户在使用过程中的操作自由度。因此，相关研究者通过研究表明<sup>[6]</sup>，借助于手势对虚拟场景进行控制将极大推动VR控制技术向着更为拟人化的方向发展。

由于通过相关VR设备相配套的手柄硬件，一方面对用户的控制操作进行了限制，一方面由于其依托于VR头戴显示器或追踪设备对其进行实时的位置定位和用户动作指令反馈，因此存在追踪域的盲区，从而导致其并不能很好的对用户的动作做出相应的反馈。所以，研究人员通过在二维平面安装以重量驱动的装置，以实现控制设备空间形状的捕捉<sup>[7]</sup>。此外，为了对人体手掌的动作进行真实的模拟，通过搭建以手掌为基准的触觉反馈系统，该系统通过内部集成的小型马达动态控制和呈现手部的力量和动作变化<sup>[8]</sup>。通过设计特定的硬件以实现对相关虚拟场景输入的检测和控制，一方面实现了更为精确的追踪和反馈，一方面对任务的需求紧密相关，从而并不具备普适性的特定，并且设计的相关硬件设备体积较大，并不能很好的商用。因此，在基于数据驱动的引导下，通过数据回归建模的方式实现了对相关的用户输入的反馈，并且实现了对目标输入动作指令的感知<sup>[9]</sup>。

研究人员通过自主设计相关硬件设置的方法和通过数据学习的方法对虚拟的控制输入进行检测和反馈，均在特定的场景中实现了较好的效果表现。但是，由于手势的非刚体特性（即手势的几何形状和运动状态受到用户所表达的特定语义内涵影响，从而可以表现出多种状态）导致直接通过检测其纹理特征识别具体的动作将对计算资源产生较大消耗，并且相对于肤色等纹理信息的骨骼结构通过直

接对手部的骨骼进行建模，因此降低了特征的建模过程，并在一定程度上降低了误检的概率。因此，2013年Leap公司面向PC和Mac的消费市场发布了Leap Motion体感控制器，该控制器通过对手势的关键节点进行捕获和追踪，从而实现将人体手部的骨骼节点信息在虚拟环境中加以呈现。受到Leap Motion设备的启发，人机交互领域的研究人员在其基础上开展了大量的研究工作<sup>[10-15]</sup>。因此，本文在结合之前研究工作的基础上，基于Leap Motion体感控制器为Cave VR直播系统开发手势控制系统。

### 3 Cave VR 直播系统

Cave系统通过应用高解析度立体成像技术和高保真声场还原技术，结合激光定位技术与动作捕捉等模块构建了具有高度沉浸体验的虚拟现实系统。通过对各种现实环境和人体感觉的反馈和模拟，为用户提供了真实的远程三维场景体验，从而推动了VR技术的进一步完善，同时也拓宽了虚拟现实技术的应用场景。相对于传统的VR环境，Cave极大的

激发了VR领域的潜力，并在商用化的进程中不断进化。目前基于Cave系统的商业化应用有：面向医疗的远程诊断、面向教育的智慧课堂、面向旅游的远程景点游览以及面向工业界的智能装配等。

#### 3.1 Cave 系统组成及优势

##### (1) 系统组成

Cave系统由图形工作站、仿真视觉模块、投影反射模块、定位追踪模块和多种信息的仿真模块构成<sup>[16]</sup>，整体结构如图1所示。Cave系统通过高性能图形工作站提供的算力支持，实现了对相关VR流媒体场景的呈现以及对复杂三维模型的精确计算和效果渲染，从而实现对接收的流媒体以高解析度显示以及对模型的逼真渲染和还原。此外，通过高流明的投影设备与具备高刷新率的多块显示模块相搭配，使用视频同步分流技术将获取到的实时媒体内容即时的分发到与之相匹配的多块显示器之中，从而实现对用户生活和工作的真实空间环境的高度模拟。

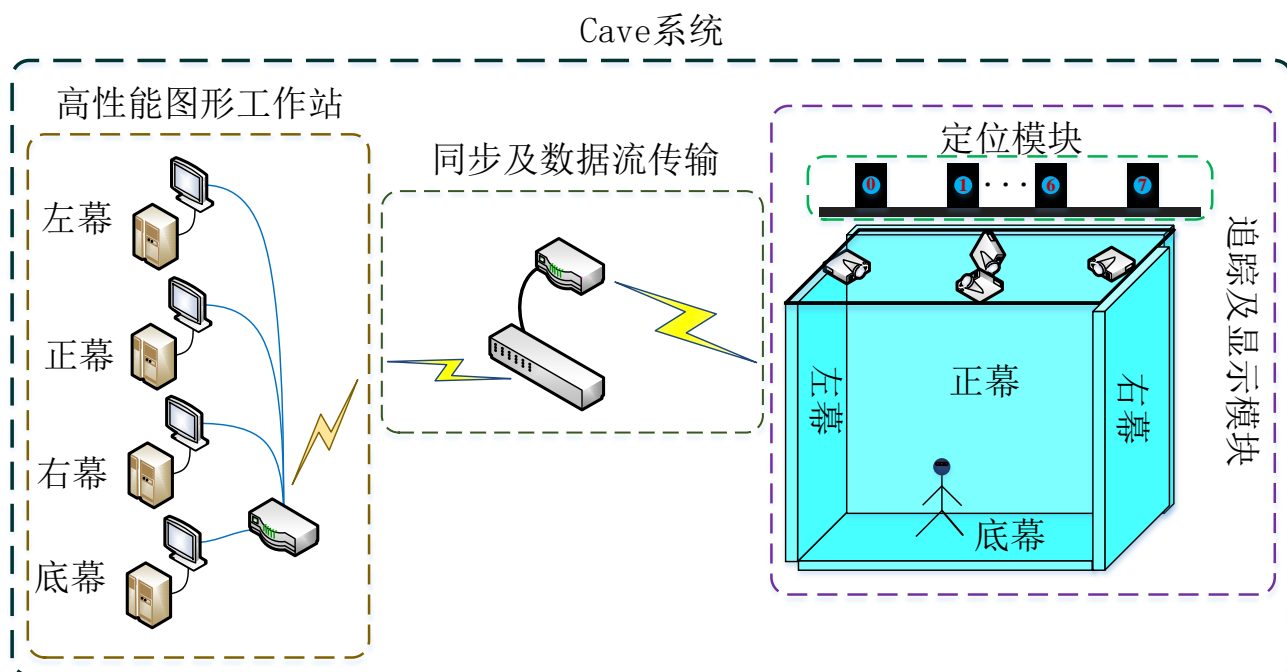


图1 Cave系统整体模型

##### (2) 系统优势

相对于VR头戴显示设备和一体机设备，Cave系统只需要借助于主动式立体眼镜就可以完成对虚拟场景的接收，从而可以明显降低用户在使用过程中所产生的疲劳感，并且可以通过追踪定位装置对

用户的实时视角和位置进行计算，从而计算出匹配于当前用户即时视角的显示效果，通过对多块显示区域因此可以尽可能的降低由于用户空间位置变化引起的显示畸变的问题。此外，由于不需要借助于全封闭式的头戴显示装置，也明显的减轻了由于虚拟场景过度沉浸而导致的用户眩晕等问题。并且通

过使用无接触式的交互装置替代手柄完成基本的交互任务，实现了用户在虚拟环境中的非约束式体验。直播作为一种强调场景渲染和用户体验的技术，如何在保证用户体验的基础上增加对相关实际任务的完成就显得尤为重要。相对于头戴显示设备，Cave 系统只需要用户佩戴质量较轻的立体眼镜即可，并且相关的直播场景可以将用户实际的包围在其中，使得用户可以切实的体验到参与实地直播的乐趣。并且通过配合特定的手势控制系统，使得用户不仅可以对直播的场景进行全范围的观看，而且可以实现实时的对相关直播物件的虚拟触碰和抓取操作，从而实现了双向实时互动式的直播体验。

因此，进一步推动直播技术向着更为沉浸式和自然化的方向发展。

### 3.2 直播系统搭建

VR 直播任务对采集环境、系统的整体处理能力以及网络传输性能要求较高，为了达到较好的直播体验，这里通过对采集场景的实时拼接技术以提升接收端的显示效果，同时使用 H.265 编码技术以压缩流媒体传输对网络带宽的需求。直播系统主要由视频源采集、源视频云端分发传输以及流媒体接收端构成，其总体结构如图 2 所示。

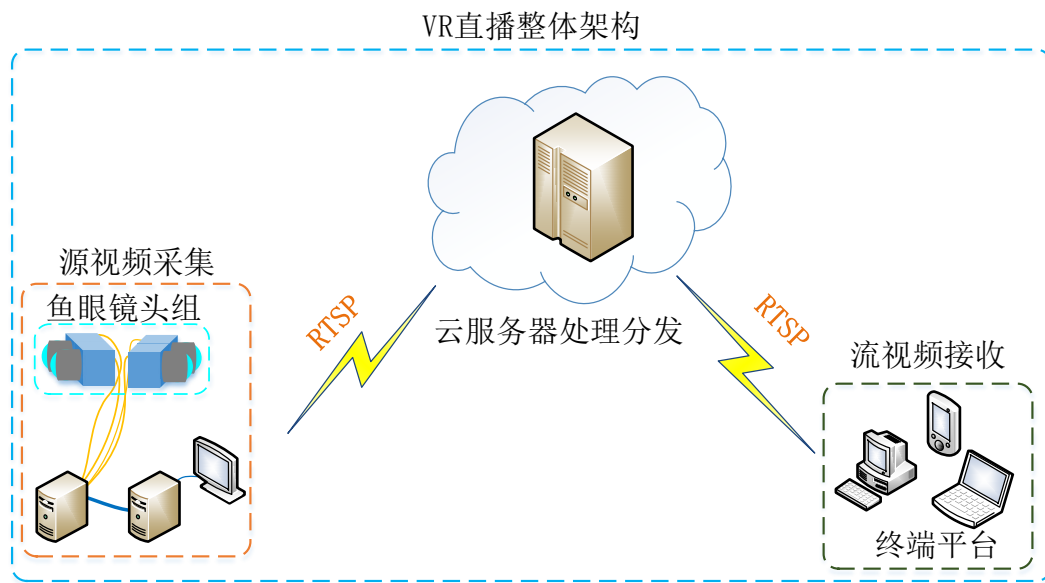


图 2 VR 直播系统模型

#### (1) 直播源采集

为了实现对直播场景的全景采集以及尽可能的模拟人体观察外界环境的模式，通过使用鱼眼相机配合相关的相机支架以实现对人体左右双目结构的模拟，从而为用户在接收端实现真实直播场景的高度还原。由于需要相关镜头的组合才能共同完成对同一方向的环境进行 360°的采集和还原的任务，从而导致了单独的相机所采集到的视频流存在部分区域的重叠，正是由于重叠区域的存在实现了对人体双目成像原理的模拟，所以通过对相关的重合区域存在的局部特征点进行标定和匹配操作，实现了将两路视频流拼接成一路完整的立体视频流的目的，从而实现了直播场景的全景和立体化采集。

#### (2) 视频流传输

由于直播过程中需要传输大量的数据信息，由于 HTTP/HTTPS 传输协议需要保证传输的有效性，从而在传输宽带数据时将存在较大的传输时延，从而使得用户在观看直播时存在明显的卡顿、延迟以及其他问题，因此极大降低用户的使用体验。此外，相对于目前使用较多的使用本地客户端既实现对场景采集，又需要完成对采集视频的上传业务的模式相比。若将相同的模式应用到 VR 的直播应用中，由于本地服务器在完成采集的同时，还需要对数据的传输进行编码任务，从而将极大的增加采集端的计算负担，从而导致直播过程出现视频的掉帧卡顿、采集视频延时传输等异常情况，并且拼接和编码任务对硬件的性能要求较高以及目前网络传输带宽尤其是移动网络带宽较为受限。为了在克服上述挑战的同时实现数据的高速传输，选取专为流媒体传输业务设计的 RTMP/RTSP 协议作为传输协议，并通

过架设专有的服务器对采集到的视频数据流进行处理和分发，从而既实现了采集和传输业务的分离操作，又实现了对 VR 直播场景的流畅处理和传输分发。

### (3) 流媒体接收

因为在视频的采集端使用的是可以对场景进行 360° 采集的鱼镜头，所以在视频的接收端需要搭建适应于接收鱼眼全景流媒体的环境，为了提高对接收环境的开发效率，使用了主流的场景建构和设计软件 Unity 3D 2018 个人版，并在 Unity 3D 平台的支持下实现了基本接收环境的搭建，并通过相关实验验证了所设计场景在显示 360° 视频中的可使用性。此外，由于在传输过程中使用了 H.265 协议实现了在保证传输质量的前提下降低传输码率的需

求，并且使用 RTMP/RTSP 协议实现了对流媒体的传输。因此，需要在接收端设计特定的解码方案和接收协议以实现实时获取和解码显示，从而最终实现对 VR 直播视频的接收任务，根据对所接收的直播实际效果的分析，其呈现状态和效果表明了整个 VR 直播系统设计的有效性和可用性。

## 4 Leap Motion 手势控制系统

在 VR 场景中，若仍借助键鼠等输入设备对用户的指令进行输入将明显降低用户的沉浸式体验。为了提升用户在虚拟场景中的使用体验，通过 Leap Motion 开发基于手势的控制系统，以完成用户在直播过程中的需求输入，其整体手势识别和控制结构如图 3 所示。

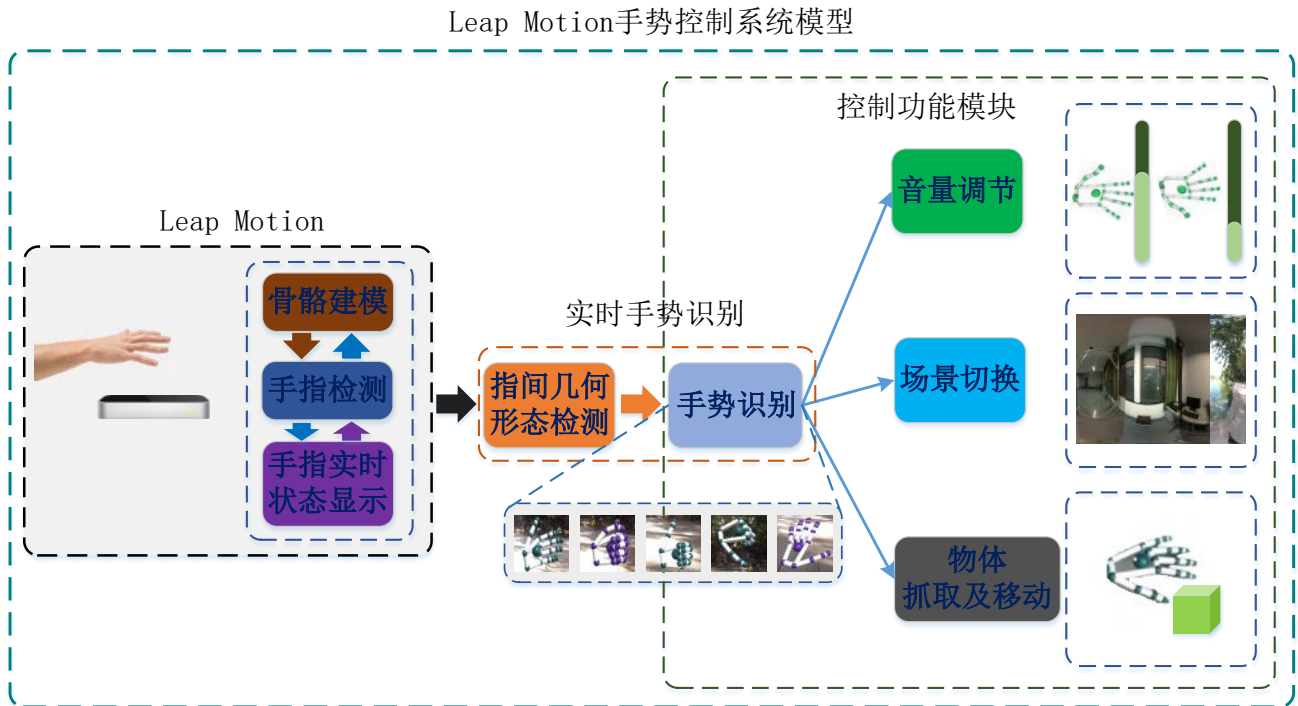


图 3 手势控制系统模型

### 4.1 开发环境简介

#### (1) Unity 3D

作为三维虚拟场景搭建和开发的平台，Unity 3D 通过模型编程的思想实现了丰富的虚拟场景应用，并成为 VR 开发的主流平台之一。Unity 3D 通过可

视化互动式的开发流程，支持跨平台开发以及多脚本语言的开发模式，从而降低了开发门槛并极大的减少了开发周期和明显降低了开发难度。其主要分为各类模型的导入接口、场景搭建显示界面和程序运行窗口显示等，通过在开发视角下对其中的模型位置和状态设置进行微调，使得调试过程变得更为简洁。此外，Unity 3D 提供了丰富的在线资源市场，

使得用户可以轻易的获取到当下热门的开发素材和教学类资源，并且为了实现对真实世界进行模拟，其通过结合目前主流的现实仿真引擎实现了对真实环境的模拟，如重力以及动力学仿真等。正是 Unity 3D 在开发过程中的诸多优势，因此在直播接收平台的搭建过程中选取其作为主要的开发平台，通过使用 Unity 3D 提供的幕布功能实现了对全景直播的 360° 接收和显示，并且借助于常规的 3D 模型设计出了基本的控制界面。

## (2) Leap Motion 控制器

为了推动 VR 技术的进一步发展，体感控制公司 Leap 在 2013 年发布了基于手势骨骼模型的控制类硬件设备 Leap Motion。用户只需要将手置于其上方适当距离的位置，Leap Motion 即可对手势的骨骼进行自动的识别和建模操作。Leap Motion 通过红外成像原理对手势的 3D 模型进行构建，并且设计了较大的可视角范围使得用户的手势不用局限于空间中某一位置和特定状态，从而提升了用户在使用过程中自由度。此外，Leap Motion 提供了较为丰富的应用开发接口，对复杂的手势定义和识别过程进行简化，降低了开发中的难度和设计冗杂的开发文档。

其通过对采集到的手势掌心位置、移动速度、掌心方向以及虚拟掌心球的位置和半径等进行计算和传输，实现了对相关手势的显示。并且为了进一步提升 Leap Motion 在实际应用中的适用性，设备内部也完成了对指尖位置和速度的检测，从而可以实现对虚拟空间中的物体的操作。并且 Leap Motion 提供了基于 Unity 3D 的开发插件，通过相关插件可以在 Unity 3D 中对其捕获到的手势进行直接的显示和计算，并且可以在可视化手势的基础上，对相关特定的手势进行显示。然而，在实际开发 VR 直播系统的手势控制系统过程中，由于需要依据不同的手势状态触发不同动作，并且其只能提供手势的骨骼信息，所以需要对手指和掌心之间的骨骼进行判断和定义，以实现特定手势的定义，这里主要通过手势骨骼的几何学特征进行判断。

## 4.2 手势控制系统及功能实现

因为在直播过程中，用户可能需要对接收到的视频音量进行调整、对某些特定直播场景中的物体进行交互以及实现不同直播场景之间的切换操作，所以所设计的基于 Leap Motion 的 VR 直播控制系统主要由手势识别、音量调节、物体抓取与释放以及场景切换组成。其主要功能界面如图 4 所示。



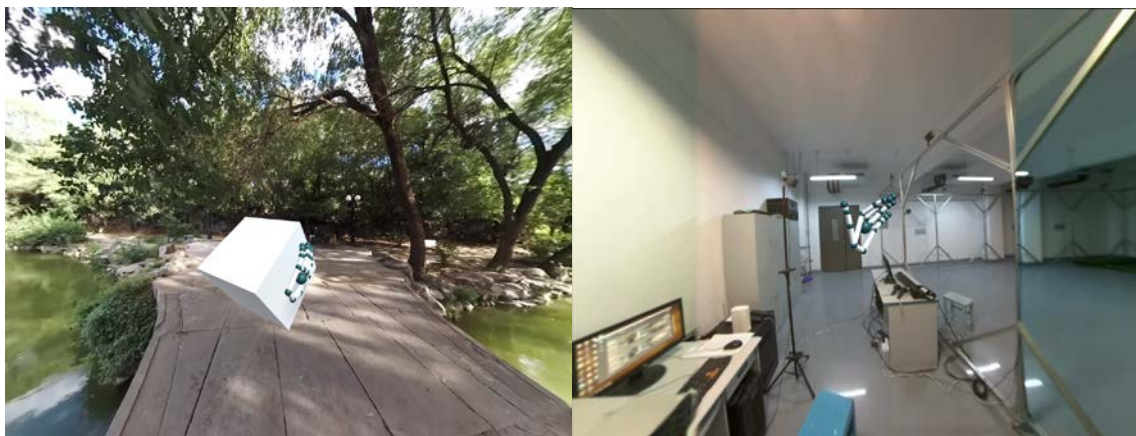


图 4 Cave VR 控制系统

### (1) 手势识别

Leap Motion 虽然实现了对用户手势的骨骼模型显示，但是其并不能对用提供的不同的手势进行识别。由于 Unity 3D 对深度学习模型的支持并不良好，并且为了充分利用 Leap Motion 控制器提供的骨骼信息，这里主要通过手势骨骼的几何学形态信息对手势进行判别，通过计算手指骨骼间以及手指骨骼与掌心骨骼之间的距离关系，以实现对用户不同手势之间的识别，并且实现了对握拳、手掌伸张、大拇指伸展等相关常用手势的检测和识别任务。

### (2) 音量调节

首先通过 Unity 3D 提供的 3D 常规模型，搭建音量的可视化模块以实现其高低定量显示，通过可视化模块用户可以直观的观察直播视频的即时音量高低并进行适当的调节，以实现合适的音量播放和观看。由于音量的高低设置和控制与手势的向上或向下状态相类似，所以通过将大拇指的向上或向下过程与之进行类比，所以通过将大拇指的向上与向下的动作与其相关联以实现直播过程中音量的控制。

### (3) 场景切换

此外，不同的用户可能在不同的时间对直播的内容和平台有不同的需求，所以用户可能同一时间需要对不同的直播进行观看，所以根据上述相关需求，开发了应对不同直播场景的切换功能。因为挥手的动作类似于对不同状态的切换，所以将不同场景的切换任务类比于用户的挥手动作，以实现用户的挥手动作与场景的切换功能相绑定，从而完成对不同直播节目切换的需要。为了使得 Leap Motion 实现对运动信息进行建模和识别，通过调用其应用开发接口中的速度检测模块结合对任意手势的识别

过程，最终实现了对不同场景的切换。并且通过将不同直播的流媒体地址集成到 Unity 3D 环境中，实现了地址与直播场景的绑定，并且当用户使用手指指定某一场景达到一定的时间长度时实现对直播场景的选择操作。

### (4) 物体抓取与移动

因为直播目前较为热门的方向之一即为“带货业务”，所以为了进一步贴合当下的应用方向，在直播场景中添加了 Unity 3D 提供的常规虚拟三维模型立方体（此处的三维模型可以根据用户需求进行更换和添加）作为抓取和释放对象，通过对 Leap Motion 的抓取和松手动作进行检测和识别，当用户的手触摸到虚拟物体之后，当用户进行抓取之后则触发用户抓取物件动作，使得用户可以对相关物体进行随意的移动以及将相关物体移动到最适宜的观察视角内；并且当用户张开手之后，Leap Motion 手势的检测过程结合 Unity 3D 提供的重力和动力学模拟引擎，实现了相关物体的自由状态释放。

## 5 结论

本文在 Cave VR 直播的环境下开发了替代传统人机交互的手势控制系统，通过将 Leap Motion 手势采集设备和 Unity 3D 虚拟仿真平台相结合的方式搭建实验场景及环境。由于 Unity 3D 软件的跨平台性和 Leap Motion 设备对手势骨骼模型的敏感性，所以本文所设计的系统可以轻易的移植到当下主流直播平台之中。并且在具体手势的识别过程中，由于 Leap Motion 并不能直接对手势进行识别，所以，通过设计基于手势骨骼几何学形态的算法以实现实时手势输入的状态进行检测和识别。此外，

本文通过将不同手势与不同动作相挂载的方式，实现了对直播场景的简单控制。根据实验结果表明，本文所设计系统在 VR 直播中的应用具有明显的优势。为了应对更为复杂的场景需求以及对手势检测精度的要求，后续准备基于本文工作的基础上对 VR 场景的控制系统进一步进行更为深入研发并结合最新的手势检测识别算法以推动 VR 手势控制系统在更为普遍环境中的应用。

## 参考文献

- [1] DeFanti, T., and D. J. Sandin. Sayre Glove Final Project Report. US NEA R60-34-163 Final Project Report, 1977.
- [2] Fisher S S, McGreevy M, Humphries J, et al. The virtual environment display system. In: Proceedings of the Workshop on Interactive 3D Graphics[C], 1987, 77-87
- [3] Lucio Tommaso De Paolis, Valerio De Luca. The impact of the input interface in a virtual environment: the Vive controller and the Myo armband[J]. Virtual Reality, 2019, 1-20.
- [4] P. Dempsey. The Teardown: HTC Vive virtual reality headset[J]. Engineering & Technology, 2016, 11(7): 80-81.
- [5] Jost Tyler A, Nelson Bradley, Rylander Jonathan. Quantitative analysis of the Oculus Rift S in controlled movement[J]. Disability and rehabilitation. Assistive technology, 2019, 1-5.
- [6] Potter, Leigh Ellen, Jake Araullo, and Lewis Carter. The leap motion controller: a view on sign language. Proceedings of the 25th Australian computer-human interaction conference: augmentation, application, innovation, collaboration[C], 2013, 175-178.
- [7] Shigeyama, Jotaro, et al. Transcalibur: weight moving VR controller for dynamic rendering of 2D shape using haptic shape illusion[J]. ACM SIGGRAPH 2018 Emerging Technologies, 2018, 1-2.
- [8] Sinclair, Mike, et al. Capstancrunch: A haptic vr controller with user-supplied force feedback[C]. Proceedings of the 32nd annual ACM symposium on user interface software and technology, 2019, 815-829.
- [9] Fujinawa, Eisuke, et al. Computational design of hand-held VR controllers using haptic shape illusion[C]. Proceedings of the 23rd Acm symposium on virtual reality software and technology, 2017, 1-10.
- [10] Alimanova, Madina, et al. Gamification of hand rehabilitation process using virtual reality tools: Using leap motion for hand rehabilitation[C]. 2017 First IEEE International Conference on Robotic Computing, 2017, 336-339.
- [11] Park, Gangrae, et al. Virtual figure model crafting with VR HMD and Leap Motion[J]. The Imaging Science Journal, 2017, 65(6): 358-370.
- [12] Vaitkevicius, Aurelijus, et al. Recognition of American sign language gestures in a virtual reality using leap motion[J]. Applied Sciences, 2019, 9(3): 445.
- [13] Caggianese, Giuseppe, Luigi Gallo, and Pietro Neroni. The vive controllers vs. leap motion for interactions in virtual environments: a comparative evaluation[C]. International Conference on Intelligent Interactive Multimedia Systems and Services. Springer, Cham, 2018, 24-33.
- [14] Bachmann, Daniel, Frank Weichert, and Gerhard Rinkenauer. Evaluation of the leap motion controller as a new contact-free pointing device[J]. Sensors, 2015, 15(1): 214-233.
- [15] Pirker, Johanna, et al. Gesture-based interactions in video games with the leap motion controller[C]. International Conference on Human-Computer Interaction. Springer, Cham, 2017, 620-630.
- [16] Cruz-Neira, Carolina, Daniel J. Sandin, and Thomas A. DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the CAVE[C]. Proceedings of the 20th annual conference on Computer graphics and interactive techniques, 1993, 135-142.