**北京航空航天大学计算机学院**

**学士学位论文开题报告**

**论文题目**：虚拟现实头眼协同对象操纵方法设计与实现

**专 业**：计算机科学与技术

**学 生**：刘兆薰

**学 号**：19373345

**指导教师**：王莉莉

**北京航空航天大学计算机学院**

2023年02月24日

目 录

[1. 论文选题背景与意义 2](#_Toc97078850)

[2. 国内外研究现状 4](#_Toc97078851)

[3. 论文的研究内容和拟采取的技术方案 8](#_Toc97078852)

[4. 论文研究计划 9](#_Toc97078853)

[5. 主要参考文献 10](#_Toc97078854)

**虚拟现实头眼协同对象操纵方法设计与实现**

# 论文选题背景与意义

虚拟现实（Virtual Reality）技术，简称VR，是一种利用计算机技术来模拟生活环境或创造虚拟现实的新型多媒体技术，是扩展现实（Extended Reality）技术的一个分支。目前主流的VR设备可利用头戴式显示器，按键式手柄或压力反馈手套，定位器和特定软件支持，建立起一个完全虚拟的三维空间。使用者在这个虚拟的环境里进行位置移动和动作改变时，计算机可以立即进行高度实时的、复杂的运算，将精确的三维世界影像传回产生临场感，让使用者身处完全的视觉环境中，并且可以进行形象的操作和交互。该技术整合了计算机图形学、仿真模拟、人工智能、传感器、网络以及并行计算等技术的最新发展成果，是一种高技术模拟系统。

VR已经在影视娱乐、教研教学、设计辅助等领域颇有建树，然而学界和工业界依旧留存着许多非常关键的问题亟待解决，比如如何高效准确并且无压力地进行虚拟环境中的物体控制和交互，如何提高虚拟环境中的图形渲染效率等，拥有非常大的研究价值。

在影视制作行业，VR主要通过构建出可与影视场景交互的虚幻三维空间场景，结合对观众的头、眼、手等部位动作捕捉，及时调整影像呈现内容，继而形成人景互动的独特体验。在网络直播行业，观众往往不能全方位了解传统的直播对象周围环境状况，无法切身感受现场氛围，而 VR 直播将活动现场还原到虚拟空间中，让观众可以身临其境地自由选择观看位置和角度，有极高的互动性。在数字展馆应用中，虚拟现实技术与展馆展示相结合，不仅体现了其开放、共享、多媒体呈现的特点，数字化呈现实体展的全部内容，还突破实体展的时空局限性，利用图文、视频、三维模型等深度资料，对重点展品进行延展和补充，加强了可视化的网络互动体验，使得展览内容更丰富和多样。在课堂教学中，虚拟现实技术可通过自然的交互方式，将抽象的学习内容可视化、形象化，为学生提供传统教材无法实现的沉浸式学习体验，提升学生获取知识主动性，实现更高的知识保留度。目前，教育已成为虚拟现实应用行业中发展最快也是最先落地的领域，随着政策的鼓励和市场的驱动，预计虚拟现实教育市场还将持续增长。在产品设计领域，以工业互联网或物联网平台为基础，虚拟现实已经成为实现数字孪生（Digital Twins）的核心技术之一。依托特定工具软件可以在虚拟空间中构建出与物理世界完全对等的数字镜像，成为将产品研发、生产制造、商业推广三个维度的数据全部汇集的基础，实现了数据信息与真实物理环境间的互动，为进行阶段性数据验证、业务流程参考的提供了重要支撑。

目前，虚拟现实已是预计增长潜力最大的高新技术之一。根据IDC研究公司在2018年发布的预测，在未来四年内，VR相关的投资将增长约20倍，具体表现为在2022年达到155亿欧元。此外，这项技术将成为企业数字化转型计划的关键。到2019年，企业在这一领域的支出将超过消费领域的支出；换言之，预计到2020年，超过一半的大型公司将有一个以VR为核心的战略。截至目前，至少有230家公司在开发VR相关产品。亚马逊（Amazon）、苹果（Apple）、Meta、谷歌（Google）、微软（Microsoft）、索尼（Sony）等互联网产品巨头都有专门的VR研发团队，为VR的未来提供了非常广阔的可能性。在未来，随着相关技术的继续发展和制造成本的逐步降低，VR势必会带来更强大的应用和市场潜力。

然而，就目前而言，限制VR普及和发展的较为直接的阻碍，除开较高的市场售价，则来自于其仍旧较低的易用性，即虚拟环境中对物体的操控和交互的准确度依旧不容乐观，或是操作指令和交互动作过于复杂繁琐。这个缺陷直接降低了用户对VR技术的接受度和使用期望。因此，VR中的对象操纵方法的优越性是提高其使用体验和普及度的基本问题之一。许多研究者已经进行了大量的研究，但仍有较大的提升空间。常规的对象操纵动作包括点击、按压、抓取、移动、释放等，其对应对象的直接具体表现主要为创建、销毁、位移、形变、旋转和缩放。对象操纵的速度、准确性、学习成本、使用压力和多样性将直接影响应用程序的效果，而在虚拟环境中实现高效且易用的对象操纵具有一定的挑战性。因此，本研究拟针对虚拟现实中对象操纵的关键问题进行研究，旨在提出相较于目前国际一流水准方法更加高效易用的基于头眼协同的虚拟现实对象操纵方法。

# 国内外研究现状

对象操纵是VR技术包含的基本且必要的交互行为之一。在过去的二十年里，国内外许多相关领域学者致力于研究基于扩展现实技术的对象操纵。扩展现实是一个总括性术语，包含了虚拟现实、增强现实、混合现实，以及三者之间的一切。针对对象操纵这一特定的研究主题，扩展现实的所有门类，甚至传统计算机图形界面的研究成果皆可启发虚拟现实的更新发展。根据已有的研究内容，扩展现实中的对象操纵方法主要由以下九大思想构成：（1）基于手部（含手柄）动作的追踪；（2）基于语音的交互；（3）基于眼动的追踪；（4）基于头部动作的追踪；（5）基于脑电波（EEG）的分析；（6）基于面部动作的追踪；（7）基于足部运动追踪；（8）基于体态的追踪；（9）基于手臂动作的追踪。

其中，又以基于手部（含手柄）追踪、基于语音交互和基于眼动追踪的方法居多，占据了所有对象操纵相关研究的80%以上[1][2]。

在基于手部（含手柄）追踪的方法研究早期（20世纪末），学界主要的两大思路为在虚拟环境中延长用户手臂和射线广播（ray-casting）,见图 1。虚拟延长手臂思路的代表研究来自于Ivan Poupyrev团队在1996年发表的Go-Go沉浸式交互方法[3]。Go-Go使用交互式增长用户手臂的元函数和非线性映射来指定和操纵远处的物体。与同时期的其他技术不同的是，Go-Go允许对附近的和远处的物体进行无缝直接操纵。然而，Go-Go技术提出的物体选择和操控模式并不能完全作为一个完整的交互方法来供人们使用； Go-Go应该被视为以同时期技术为基础的一个补充，而不能完全取而代之。射线广播的思路和虚拟延长手臂类似；射线广播的思路是将射线束从使用者的手中延伸出来，从而指定操纵物体。然而，射线广播思路存在比较明显的弊端。由于在射线广播的对象操纵中物体是被连接到射线末端的，所以除了以射线本身为轴可以完成的动作，许多操纵都是难以简单实现的，因为只有一个自由度（围绕射线轴的旋转）可以用射线广播的方式独立控制。比如，若用户希望以与射线方向垂直的轴向旋转一个物体，单纯以射线广播是无法完成的。此外，射线广播还缺乏一种控制物体与用户之间距离的方法，导致用户无法准确地将物体拉近或推远，而这也是对象操纵的基础功能之一。

当把Go-Go方法和射线广播以及其他方法（例如通过将虚拟手臂延伸到无限远来改进Go-Go的Stretch Go-Go）进行比较时，并没有明显的赢家[5]。用户评估结果显示，所有技术都有明显的缺点。在这次评估中，HOMER方法被提出来了；这是一种以手为中心的基于射线广播的对象操作技术[5]。HOMER使用射线来选择物体，在选择物体后，它将虚拟的手移动到物体上；用户的身体和手之间的当前距离被映射为到虚拟物体的距离。因此，HOMER方法操纵对象的方式与Go-Go技术类似，但缩放系数是针对每个选定的物体独立计算的。

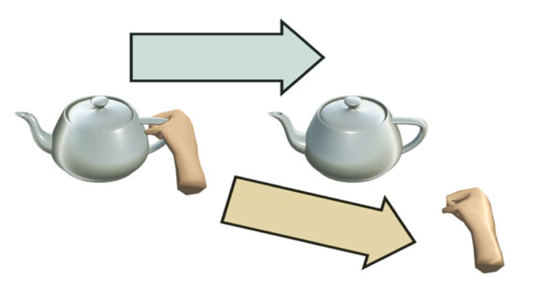
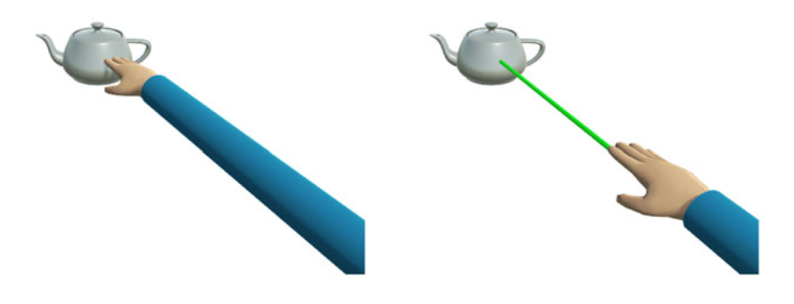


图 1 研究早期三种主要的基于手部动作追踪的对象操纵方法

(a) PRISM

(b) Go-Go (c) 射线广播

到了21世纪，学界将研究重心放在虚拟环境对象操纵的精确性上。一个具有代表性的方法是Scott Frees团队在2005年提出的PRISM[4]。PRISM是一种通过缩放操作进行精确和快速交互的方法，它在同时期是一种非常新颖且具有开创性的交互技术。PRISM主动根据用户在虚拟环境中的行为特征来确定他们所想的操纵目标是明确还是不明确的。当操纵目标明确时，PRISM动态地调整“控制/显示”比例来提高对象操纵的精确度。该比例决定了物理手部运动和受控虚拟物体运动之间的关系，降低了传感器对手部运动监测不必要的灵敏度，从而减少操作误差。使用PRISM，用户始终完全控制着被操纵物体的位置。与像Go-Go这样的技术相比，PRISM在能力范围上也有很大提升，最明显的进步在于PRISM扩大了手部运动的规模以允许“远距离”操纵，同时在特定情形中可以主动缩小手部运动的幅度以提高精确度。

在这之后，Curtis Wilkes等人将PRISM与HOMER相结合，在2008年提出了融合了两种方法精华的Scaled HOMER[6]。Scaled HOMER使用基于速度的缩放，允许用户在近距离和远距离进行更为精确的操作。它比原始的HOMER在各种任务条件下，尤其是有关需要高度精确、远距离放置物体或大运动距离的任务中的性能都有所提高。2015年，在Go-Go和PRISM研究之后，Chris Auteri等人将这两种技术结合起来，以提高延伸的三维操作的精确性[7]。该方法首先将PRISM直接应用于用户的手（基础光标）的运动，从而基于运动速度计算出一个新的光标位置（PRISM光标）。然后，PRISM光标移动的距离被基于Go-Go距离的启发式方法所放大。与 PRISM和HOMER的结合一样，Go-Go和PRISM的结合提供了一些改进，尤其是在任务完成的成功率和精细度上。

基于双手的操纵在2008年被Noritaka Osawa团队提出[8]。该团队提出了一种用于在沉浸式虚拟环境中精确定位3D虚拟物体的单手和双手控制技术。这个方法提出了一种位置调整策略，包括一个类似于PRISM的用于减缓手部运动的比例系数以及一个被动的视角调整。该交互系统会自动将视角接近抓取点，使被操纵的物体看起来更大，从而更易于操控。为了有效控制这些调整，该团队提出了两种技术。第一种是基于单手操纵的；因为当用户想精确地操纵一个物体时，他们的手会慢慢移动，所以通过对单手的速度监测，系统可以判断当前对象是否需要精确操纵。另一种是基于两手间距离的；当用户两手之间的距离很小时，调整就会被激活。通过用户评估，位置和视点的调整比禁用这种调整有更好的操纵效率和用户体验。此外，该团队的测试结果还显示，双手控制比单手表现更好。承接双手直接操纵的方法，Martinet团队提出了两种移动3D对象的技术[12]。第一种扩展了许多CAD（Computer-aided Design，计算机辅助设计）应用程序中的视窗概念；它引入了四个视窗，每个视窗显示3D对象的不同视图。在其中一个视窗中触摸并拖动物体，可以在与该视窗平行的平面上平移物体。第二种方法被称为Z技术；Z技术只使用场景的一个视图（见图2）。在这种技术中，第一次触摸触发在平行于视图的平面上移动物体，第二次触摸触发垂直于视图平面的前后运动。Martinet的初步评估表明，用户更喜欢Z技术。Martinet等人在Z技术的基础上进行了改进，推出了DS3，一种基于DOF分离的三维对象操纵技术[13]。与Z技术类似，一次直接触摸可以在屏幕平面上移动物体，间接触摸可以操纵物体深度，两次直接触摸可以实现旋转。Martinet将DS3与之前的类似方法，比如Hancock团队提出的Sticky Tools方法（见图3）和Reisman团队提出的Screen-Space方法进行了比较，结果显示DOF分离导致了更好的结果[14][15]。

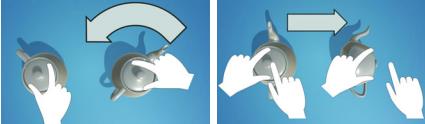


图 2 基于双手直接操纵的Sticky Tools方法

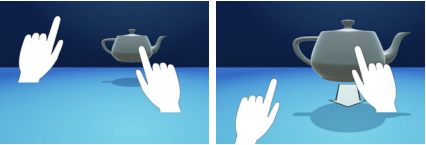
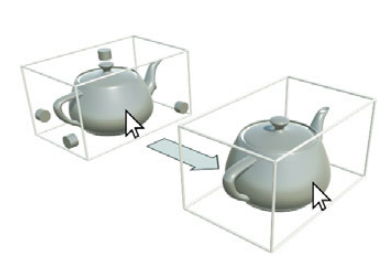
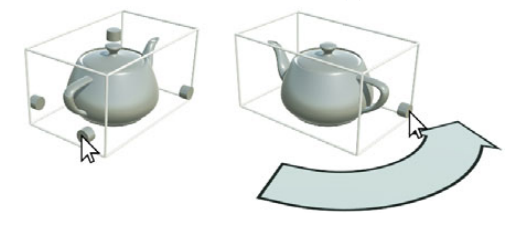
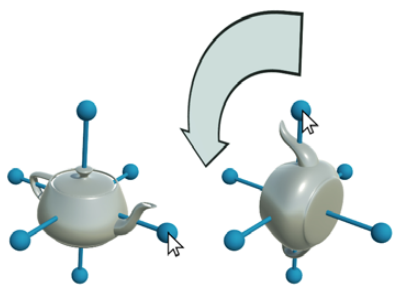
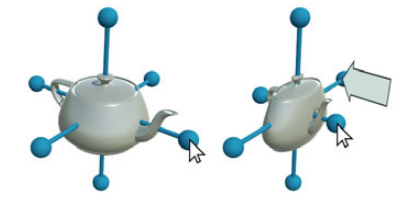


图 3 基于双手直接操纵的Z技术

另外一种值得一提的基于手部动作追踪的方法是外加虚拟控制柄（见图4）；虽然这类方法暂未被应用到虚拟环境中，但是它们对于对象操纵的研究是颇具启发的。Houde团队在1992年开发了一种基于“操纵盒”的方法；这种方法由一个围绕着物体的边界长方体框组成，拖动长方体即拖动物体，另外还有三个旋转柄用于围绕其中心轴旋转物体[9]。Conner团队也采用了设置虚拟控制柄来进行对象操纵；他们的方法允许完整的9-DOF（Degree of Freedom，自由度）控制（平移、旋转和缩放）甚至诸如扭曲的其他变形[10]。该方法的虚拟控制柄的两端有一个小球体，它们将几何变换约束在一个平面或轴上；用户拖动其中一个球体可以平移、旋转或缩放物体。秉承着这两个方法的思想，Mendes团队在2016年提出了相似的一个基于外加虚拟控制柄的方法。他们从实验结果中提出了一套发展准则：（1）直接操作很适合粗略的变换；（2）位移和旋转操作应尽可能分离，以防止不需要的变换；（3）单一的DOF分离对于精确的变换是非常理想的，通常用于细粒度的调整。

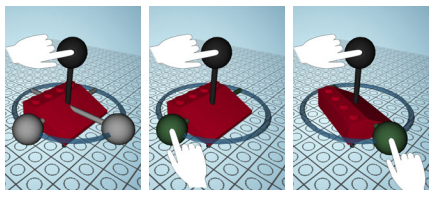


(a) Houde团队的方法

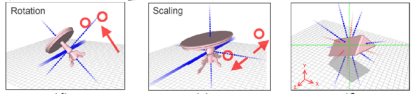
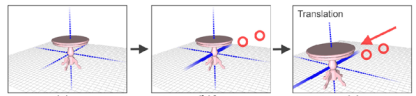
(b) Conner团队的方法

图 4 外加虚拟控制柄的两种代表方法

2010年以后，基于手部动作（含手柄）追踪的非直接对象操纵方法开始出现。其中较有代表性的两个方法为Mendes团队在2011年提出的LTouchIt和Kin-Chung Au团队在2012年提出的多点触摸方法（见图5）。LTouchIt虽然使用了直接操纵平移的方法，但在DOF分离之后，它可以控制物体在不超过两个维度上的位置，并使用旋转手柄一次围绕一个轴进行旋转；用户可以选择一个手柄来定义一个旋转轴，并通过另一只手的操作来指定旋转角度[16]。Au团队利用多点触摸表面的高输入带宽，将标准变换部件的操作能力委托给多点触摸手势。这使得使用单一的多点触控动作就能对约束和变换操作进行无缝控制。用户可以用两个触摸点选择一个候选轴，通过按住并移动两个手指来进行物体的变换[17]。



(a) LTouchIt



(b) 多点触控方法

图 5 两种基于手部动作（含手柄）追踪的非直接对象操纵方法

# 论文的研究内容和拟采取的技术方案

本研究的主要内容和目标是：（1）探索比较多种头眼协同交互方法的优劣性；（2）确定一套高效易用的基于头眼协同的虚拟现实对象操纵方法；（3）确定针对操纵效率和易用性的评估指标；（4）根据评估指标对比目前国际一流方法（baseline）并作结果分析。

# 论文研究计划

* 2021.12 – 2022.2：阅读文献，分析国内外研究现状。
* 2022.2 – 2022.3：**Object-target视觉度量**
* 2022.2 – 2022.3：**操纵类型分配函数的构建和滤波器的构建**
* 2022.3 – 2022.4：确定方法的pipeline并做实验
* 2022.4 – 2022.5：论文撰写

# 主要参考文献

1. Mendes, Daniel & Caputo, Ariel & Giachetti, Andrea & Ferreira, Alfredo & Jorge, Joaquim. (2018). A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments: Survey on 3D Virtual Object Manipulation. Computer Graphics Forum. 38. 10.1111/cgf.13390.
2. Radianti, Jaziar & Majchrzak, Tim A. & Fromm, Jennifer & Wohlgenannt, Isabell. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. Computers & Education. 147. 10.1016/j.compedu.2019.103778.
3. Poupyrev, Ivan & Billinghurst, Mark & Weghorst, Suzanne & Ichikawa, Tadao. (1998). The Go-Go Interaction Technique: Non-linear Mapping for Direct Manipulation in VR. Proc. of UIST'96.
4. Frees, Scott & Kessler, G.D.. (2005). Precise and rapid interaction through scaled manipulation in immersive virtual environments. Proceedings - IEEE Virtual Reality. 99-106. 10.1109/VR.2005.1492759.
5. Bowman, Doug & Hodges, Larry. (1999). An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. Symposium on Interactive 3D Graphics. 182. 10.1145/253284.253301.
6. Wilkes, Curtis & Bowman, Doug. (2008). Advantages of velocity-based scaling for distant 3D manipulation. 23-29. 10.1145/1450579.1450585.
7. Auteri, Chris & Guerra, Mark & Frees, Scott. (2015). Increasing Precision for Extended Reach 3D Manipulation. International Journal of Virtual Reality. 12. 10.20870/IJVR.2013.12.1.2859.
8. Osawa, Noritaka. (2008). Two-Handed and One-Handed Techniques for Precise and Efficient Manipulation in Immersive Virtual Environments. International Symposium on Visual Computing.
9. Stephanie Houde. (1992). Iterative design of an interface for easy 3-D direct manipulation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92). Association for Computing Machinery, New York, NY, USA, 135–142. <https://doi.org/10.1145/142750.142772>
10. Conner, D. & Snibbe, Scott & Herndon, Kenneth & Robbins, Daniel & Zeleznik, Robert & van Dam, Andries. (1992). Three-Dimensional Widgets. Computer Graphics (1992 Symposium on Interactive 3D Graphics). 25. 10.1145/147156.147199.
11. Mendes, Daniel & Relvas, Filipe & Ferreira, Alfredo & Jorge, Joaquim. (2016). The Benefits of DOF Separation in Mid-air 3D Object Manipulation. 10.1145/2993369.2993396.
12. Martinet, Anthony & Casiez, Géry & Grisoni, Laurent. (2010). The Design and Evaluation of 3D Positioning Techniques for Multi-touch Displays. 3DUI 2010 - IEEE Symposium on 3D User Interfaces 2010, Proceedings. 115-118. 10.1109/3DUI.2010.5444709.
13. Martinet, Anthony & Casiez, Géry & Grisoni, Laurent. (2010). The Effect of DOF Separation in 3D Manipulation Tasks with Multi-touch Displays. Proceedings of VRST'10, the 17th ACM Symposium on Virtual Reality Software and Technology. 10.1145/1889863.1889888.
14. Hancock, Mark & Cate, Thomas & Carpendale, Sheelagh. (2009). Sticky tools: Full 6DOF force-based interaction for multi-touch tables. 133-140. 10.1145/1731903.1731930.
15. Reisman, Jason & Davidson, Philip & Han, Jefferson. (2009). A screen-space formulation for 2D and 3D direct manipulation. UIST 2009 - Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology. 69-78. 10.1145/1622176.1622190.
16. Mendes, Daniel & Lopes, Pedro & Ferreira, Alfredo. (2011). Hands-on interactive tabletop LEGO application. 19. 10.1145/2071423.2071447.
17. Au, Oscar & Tai, Chiew-Lan & Fu, Hongbo. (2012). Multitouch Gestures for Constrained Transformation of 3D Objects. Computer Graphics Forum. 31. 651-660. 10.1111/j.1467-8659.2012.03044.x.