

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
2. 国内外研究现状.....	3
3. 论文的研究内容和拟采取的技术方案.....	12
4. 论文研究计划.....	14
5. 主要参考文献.....	15

# 虚拟现实头眼协同对象操纵方法设计与实现

## 1. 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 and 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。

## 2. 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<sup>[1][2]</sup>。

在基于手部（含手柄）追踪的方法研究早期，学界主要的三大思路为在虚拟环境中单手直接操纵、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<sup>[3]</sup>。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<sup>[5]</sup>。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<sup>[5]</sup>。HOMER 使用射线来选择物体，在选择物体后，它

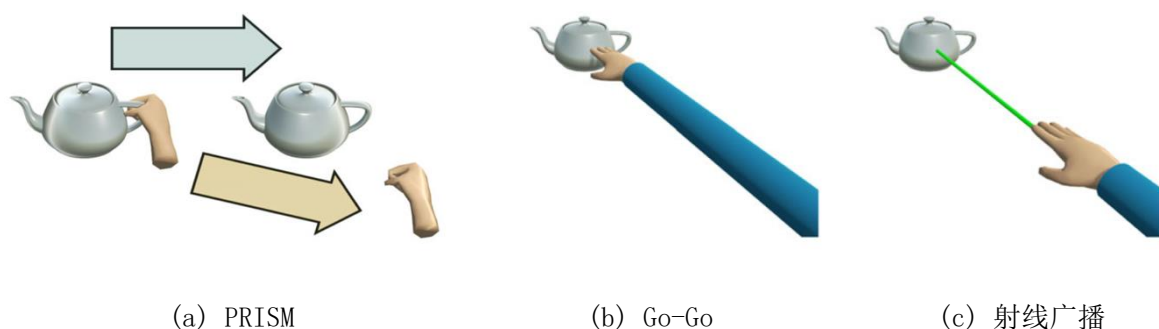


图 1 研究早期三种主要的基于手部动作追踪的对象操纵方法

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。

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<sup>[4]</sup>，（见图 1(a)）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<sup>[6]</sup>。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<sup>[7]</sup>。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<sup>[8]</sup>。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



图 2 基于双手直接操纵的 Z 技术

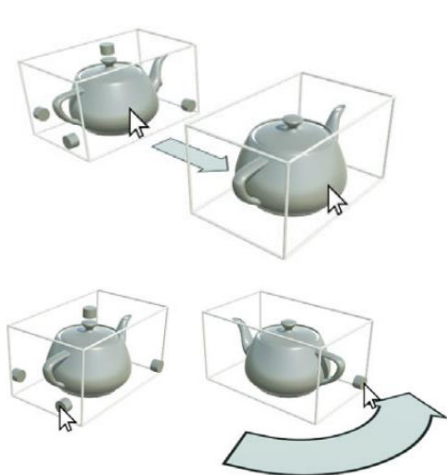


图 3 基于双手直接操纵的 Sticky Tools 方法

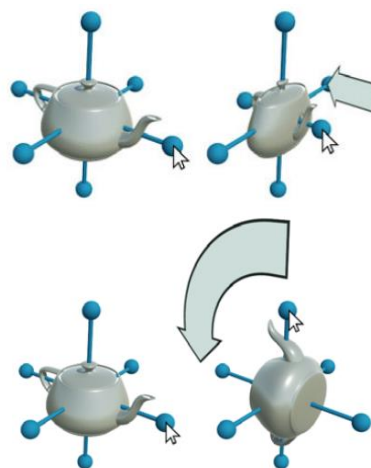
这种调整有更好的操纵效率和用户体验。此外，该团队的测试结果还显示，双手控制比单手表现更好。承接双手直接操纵的方法，Martinet 团队提出了两种移动 3D 对象的技术<sup>[12]</sup>。第一种扩展了许多 CAD（Computer-aided Design，计算机辅助设计）应用程序中的视窗概念；它引入了四个视窗，每个视窗显示 3D 对象的不同视图。在其中一个视窗中触摸并拖动物体，可以在与该视窗平行的平面上平移物体。第二种方法被称为 Z 技术；Z 技术只使用场景的一个视图（见图）。在这种技术中，第一次触摸触发在平行于视图的平面上移动物体，第二次触摸触发垂直于视图平面的前后运动。Martinet 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<sup>[13]</sup>。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）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<sup>[14][15]</sup>。

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）；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<sup>[9]</sup>。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<sup>[10]</sup>。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





(a) Houde 团队的方法

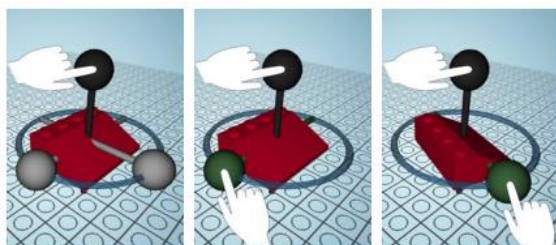


(b) Conner 团队的方法

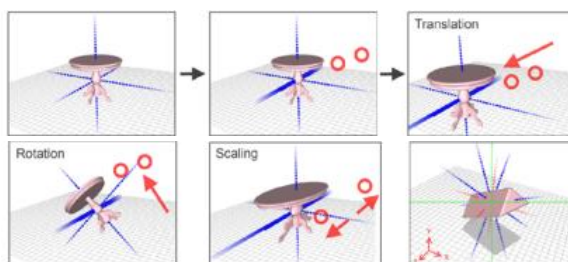
图 4 外加虚拟控制柄的两种代表方法

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。

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<sup>[16]</sup>。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<sup>[17]</sup>。



(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



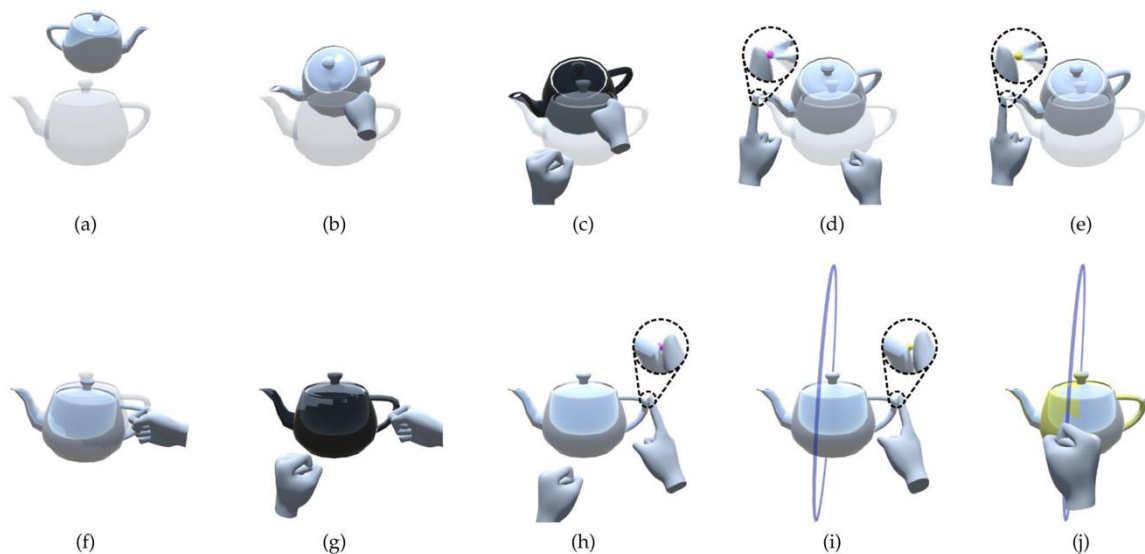


图 6 PinNPivot 完成物体转移的过程：(a)初始位置和目标位置；(b)6DOF 操纵迅速将物体翻转过来；(c)按比例 3DOF 平移准确地放置壶口的尖端；(d)一个销钉被创建；(e)销钉被锁定；(f)物体在 3DOF 中快速旋转；(g)物体在按比例 2DOF 中被精确旋转；(h)创建第二个销钉；(i)第二个销钉被锁定，出现一个环；(j)该物体在按比例 1DOF 中旋转。当目标变成黄色时，表示与目标位置吻合良好。

目前，基于手部动作（含手柄）的追踪的对象操纵方法的 SOTA (state-of-the-art, 最先进方法) 为 Gloumeau 团队在 2020 年提出的 PinNPivot 方法<sup>[39]</sup>。这个方法使用“销钉 (pin)”来约束 1DOF/2DOF/3DOF 旋转；PinNPivot 还支持 6DOF 操纵和 3DOF 平移，具体的一个操作流程可参考图 6。在该团队与以往技术的比较表明，PinNPivot 拥有更准确和更快的操纵效率。

同样在 2010 年以后，基于语音交互的对象操纵方法出现<sup>[18]</sup>。该类方法代表是 Carruth 团队在 2019 年提出的一种使用 HMD 对小型无人驾驶地面车辆的语音操作进行沉浸式培训方法<sup>[19]</sup>、Mirzaei 团队在 2014 年提出的一种使用自然语言和用户手势在真实规模的数字模拟中进行导航和互动的的方法和 Ferracani 团队在 2017 年提出的一种通过语音指令游览虚拟博物馆的方法。然而，基于语音的交互模式存在许多不理想的因素。首先，语音交互对用户的要求相对较高，用户必须以标准的发音、强逻辑性和强简洁性陈述自己的语音指令来与系统交互，这无可避免地提高了用户使用语音交互系统时的负担；其次，语音交互的干扰（如噪音）很难消除，错误率较高；再者，语音交互的局限性较大，因为在某些需要安静和隐私的特殊环境（如办公室，学生宿舍等）语音交互系统是不会被考虑的；最后，语音交互难以实现撤回指令，用户无法以简易的方式更正操作失误。

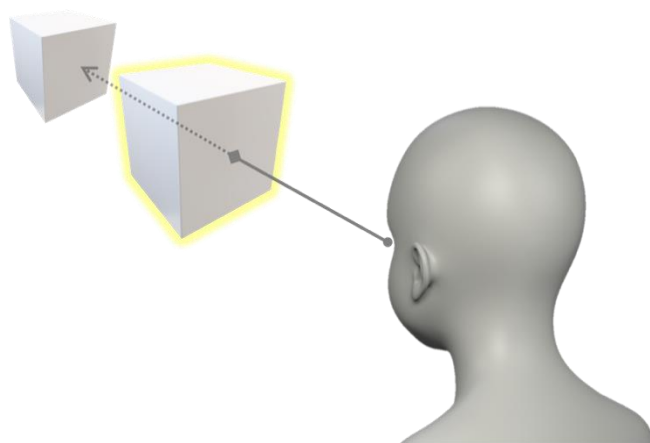


图 7 基于眼睛注视的指向选择

在 2015 年以后，眼球追踪在头戴式虚拟现实显示器中的应用越来越多，各种集成了眼球追踪器的头盔已经在市场上销售。根据 Adhanom 团队在 2023 年发表的研究，眼动跟踪在虚拟现实中的应用是高度多样化的，并且跨越了多个学科<sup>[21]</sup>。因此，近年来基于眼动追踪的对象操纵方法应运兴起。

在 VR 中实现基于眼睛注视的指向的最常见方法是使用眼动仪提供的 3D 注视方向向量，并观察场景中的哪些对象与方向向量相交<sup>[22]</sup>。通常，射线是基于方向向量投射的，并且射线相交的第一个物体被认为是被指向的项目（见图 7）。这与射线广播的基本思想是一致的。各种研究表明，基于凝视的指向比基于手的指向更快，因为我们能够比我们的手更快地将目光移向目标<sup>[23]</sup>。然而，由于眼球运动的固有生理特性和眼动追踪的技术限制，与其他常见的指点界面相比，基于眼睛注视的准确度还是稍显逊色的<sup>[24][25]</sup>。基于眼睛注视的指向界面中的不准确性主要有两种形式，一是由眼睛跟踪数据中的自然噪声引起的，二是由眼睛跟踪数据质量不稳定引起的。

基于眼动追踪的操纵方法的一大难点是目标对象选择。仅通过眼睛注视进行选择是一项相对具有挑战性的任务，需要实施更为精密的机制以在虚拟环境中使用基于眼睛的交互。根据以往的方法，我们可以实施其他选择确认技术来辅助眼动交互。这样做的一个额外好处是可以解决人机交互领域经典的“点石成金”问题，即“无论你看哪里，都有东西被激活；你不能在没有发出命令的情况下看任何地方”<sup>[26][27]</sup>。

目前已有许多学者使用各种技术来实现虚拟环境中基于注视的交互选择确认。Hansen 团队在 2018 年提出了一种基于眼睛注视的停留进行选择确认的技术<sup>[24]</sup>。Sidenmark 和 Gellersen 在 2019 年实施了两种头部辅助的眼动交互技术，第一种是 Eye

& Head Dwell 第二种是 Eye & Head Convergence<sup>[22]</sup>。Eye & Head Dwell 是一种停留以确认的技术，其中停留计时器仅由头部支持的凝视转移触发，但可以通过仅眼睛凝视暂停和恢复；Eye & Head Convergence 是一种用于快速目标确认的替代技术，它允许用户通过将眼睛指示器和头部指示器对准目标来确认选择。Kumar 和 Sharma 团队在 2016 年提出了一种使用眨眼进行选择确认的技术<sup>[28]</sup>。Pfeuffer 团队在 2017 年提出了一种手眼协统的选择确认方法；这个方法允许用户用眼睛注视物体并同时加以一种“捏合”手势来辅助确认选择<sup>[29]</sup>。Pai 团队在 2019 年提出了另外一种协同辅助操纵技术；用户可以用目光指向目标，并使用肌电图检测到的手臂肌肉收缩来触发操纵动作<sup>[30]</sup>。Qian 和 Teather 团队在 2017 年提出了一种通过键盘按钮按下进行选择确认辅助，并使用眼睛注视进行指向选择的方法<sup>[31]</sup>。最近，Sidenmark 团队在 2020 年提出了 Outline Pursuits 方法，它利用平滑追踪来允许用户在虚拟环境中选择被遮挡的对象<sup>[32]</sup>。

伴随选择技术而来的另外一个值得关注的问题是反馈技术。一个完整的虚拟环境中的对象操纵技术应该向用户提供反馈，让用户能够清楚地了解系统的状态<sup>[33]</sup>。由于眼睛对视野中的视觉变化很敏感，它们会本能地尝试将注意力转移到这些视觉变化上。因此，在向用户提供反馈时应该格外小心，因为视觉上突出的反馈机制可能会产生意想不到的后果，即转移用户的视线以产生不期望的交互动作。Boyer 团队在 2017 年提出的一种非视觉反馈方法，他们使用听觉反馈来避免不必要的视线转移<sup>[34]</sup>。然而，学界依然有很多基于视觉的反馈方法：Blattgerste 团队在 2018 年提出了一种突出显示所选对象的反馈方法；Mohan 团队也在 2018 年提出了一种在所选对象周围显示确认标志的方法；Sidenmark 团队在 2020 年提出了一种在所选对象周围显示轮廓的反馈方法<sup>[35][36][32]</sup>。

根据目前基于眼动追踪的对象操纵方法的研究数量，眼动追踪将很快成为 HMD 系统

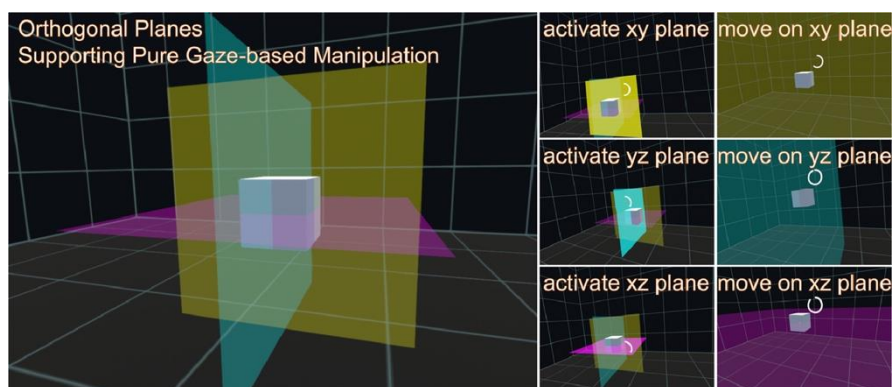


图 8 OrthoGaze 的正交平面

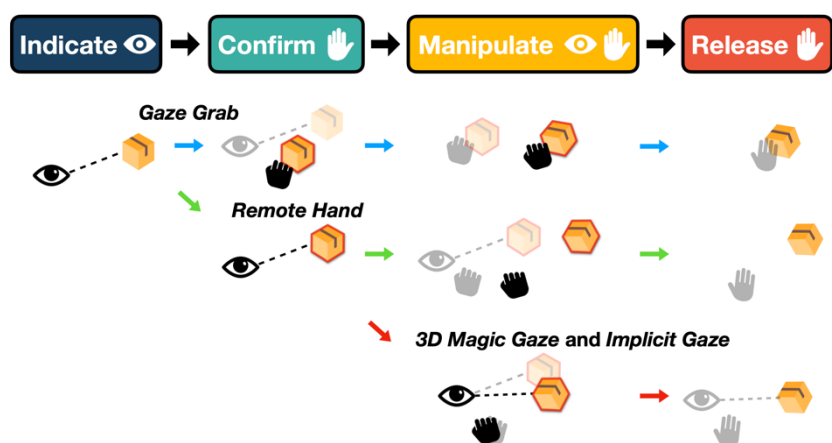


图 9 Yu 团队提出的基于凝视的三维物体操纵方法操作流程

不可或缺的一部分。因此，我们预计围绕 HMD 眼动追踪的研究和开发将在未来几年加速和扩展。然而，目前大部分眼动追踪方法依旧存在许多值得优化的问题。除了硬件限制，交互动作所带来的生理性不适也需要得到改善。大多数基于眼动追踪的操纵方法都包含眨眼命令（包括眨眼、双眼眨眼和眨眼眼球运动）<sup>[28]</sup>。要求用户改变他们的自然眨眼频率可能会导致用户眼睛疲劳、眼睛干涩和眼睛疲劳<sup>[37]</sup>。Kumar 和 Sharma 团队在 2016 年的研究结果也表明，频繁眨眼和眨眼会导致用户眼睛疲劳<sup>[28]</sup>。而且，基于眨眼的界面往往不准确，因为下意识的眨眼很难与自然眨眼区分开来，所以系统往往需要用户做出完全下意识的眨眼动作，如快速多次眨眼。然而，长时间眨眼有明显的缺点，例如减慢交互流程并在长时间眨眼期间阻挡用户的视线。因此，基于眼睛注视的系统控制需要在交互动作上做出合理的优化。

一个值得学习的基于凝视的对象移动方法是 Liu 团队在 2020 年提出的 OrthoGaze，它允许用户只需用眼睛凝视就能直观地操纵虚拟物体的三维位置<sup>[41]</sup>（见图 8）。该方法利用了三个可选择正交平面，其中每个平面不仅有助于引导用户在任意虚拟空间中的注视，而且还允许对物体位置进行 2DOF 操作，是一种完全不依赖手部的对象操纵方法。

目前基于眼动追踪的对象操纵方法的 SOTA 为 Yu 团队在 2021 年提出的一种基于凝视的三维物体操纵方法<sup>[40]</sup>。该方法的整个操纵任务可以分解为四个阶段：指示、确认、操作和释放（见图 9）。该研究表明，当所有目标都位于用户前方且在手臂可触及的距离内时，基于凝视的交互对于物体操作没有明显的性能优势；但对于有远处物体的较大空间，凝视输入可以减轻手臂的疲劳问题。眼动和其他模态的不同整合、协调和过渡策略可以为构建更高效的物体操纵技术提供优势。

### 3. 论文的研究内容和拟采取的技术方案

本研究的主要探究问题是如何确定一套快速、准确、易用的虚拟现实头眼协同对象操纵方法。这个问题的主要难点包含：（1）通过眼动可获取的信号有限，因为眼球本身无法做出多样化的动作；（2）由于本能性的动作干扰（如无意识的眨眼），眼动信号解析难度大；（3）已有的基于眼动追踪的对象操纵方法使用负担非常大，主要是因为操作流程复杂并且需要视线高度集中，容易产生眼球生理性疲劳。需要考虑以尽可能少的头眼动作组合和尽可能小的专注度要求完成复杂的对象操纵步骤。

基于此考虑，本研究的主要内容是：（1）头眼协同的场景与目标浏览；（2）头眼协同的目标选择；（3）头眼协同的对象操纵。大致研究步骤是：（1）确定针对操纵速度、准确度和使用负担的评估指标。这可以在接下来的文献查阅中获取相应的指引；（2）探索比较多种头眼协同交互方法；（3）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，支持操纵对象位移、旋转和缩放；（4）根据评估指标对比目前国际一流方法（baseline）并作结果分析。

针对探索比较多种头眼协同交互方法，我们首先需要探究尽可能多的头眼协同方式可能性。要想获得尽可能多的可能性，就需要先整理眼部动作和头部动作的所有可支配动作。由于头部运动，或者严格意义上来说，颈部运动的种类相对较少且显而易见，我们在此仅整理眼部动作。

按照动作的适用性程度，我们可以将眼部运动整理为七种：眨眼（blink）、凝视（gaze/fixation）、跳视与微跳视（saccade & microsaccade）、平滑追瞄（smooth pursuit）、视动震颤（nystagmus）、辐辏运动（vergence）和前庭眼反射（vestibular ocular reflex, VOR）<sup>[38]</sup>。从技术层面来讲，眨眼不是眼睛本身的动作，但是这项行为在眼动研究中十分重要。一个例子是眨眼可以中断凝视行为，从而可以让眼动追踪系统做出相应的交互动作。凝视通常用来描述双眼处理视觉信息时停留相对稳定的一段时间。一般而言，凝视通常会 是 200 到 300 毫秒之间；但取决于具体的任务，也可能是更长或更短的时间。比如在文字阅读时，时间会较短；在观看影像资料时，时间会较长。跳视是一种下意识的行为，指双眼非常快速且同步地在两个距离较大的凝视点之间移动；而微跳视多是一种无意识的本能行为，用来快速切换小距离的凝视点。跳视之所以在多数情况下是一种下意识的行为，是因为人体在涉及较大距离的凝视点切换时，多以微跳视辅以头部转动来代替跳视。平滑追瞄是由一系列微跳视以及凝视组成的，多用以持续



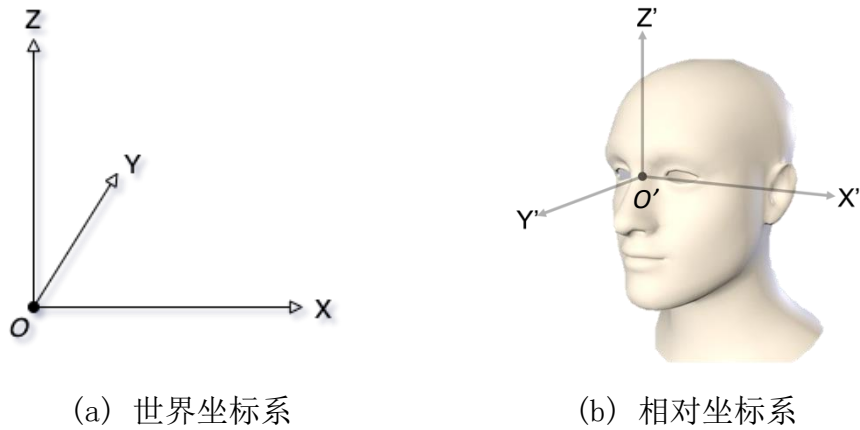


图 10 相对于头部的坐标系

追踪移动物体。视动震颤是指眼球快速左右移动，可以视为大幅加快速度的平滑追瞄。幅奏运动是指当我们将眼睛聚焦在很靠近面部的目标上时，我们的眼睛凝视会同时向内聚集，俗称“对眼”；这是一种非正常的眼动，因为健康的眼球运动应该是共轭的，即同方向的，所以交互系统应该尽量避免依赖这种眼动方式。前庭眼反射（VOR）是指当我们凝视着一个目标时同时转动头部，VOR 系统会确保眼睛调整到适当的位置以持续凝视目标；具体的一个常见表现为当我们直视前方时将头部向右旋转，我们的眼睛会自动向左旋转，以确保凝视点持续在前方。在本次研究中，我们希望能够尽可能地尝试不同的头眼协同组合，并且记录、评估、对比每组组合的速度、准确度与使用负担。

关于具体交互系统的开发实现，我们在系统中引入一个恒定的三维直角坐标系作为世界坐标系，和一个相对于头部静止的三维直角坐标系（见图 10）；这两个坐标系是所有对象操纵的基础。我们在初始化世界坐标系时规定世界坐标系与初始状态下的相对坐标系有如下关系：

$$\overrightarrow{OX} \cdot \overrightarrow{O'X'} = 0$$

$$\overrightarrow{OY} \cdot \overrightarrow{O'Y'} = |\overrightarrow{OY}| |\overrightarrow{O'Y'}|$$

$$O = O'$$

交互系统的运作遵循有限状态机，见图 11。状态机具体的状态内行为和状态转移条件需要在后期研究中进一步的筛选和确认，在此我们仅讨论目前期望值最高的几种构思。

在进入 OBJECT\_MOVING 状态时和在该状态下，目前主要构思的一个交互模式为快速两次眨眼以进入，进入后物体根据头部转动的轴向产生相应方向的位移。一个普适的规则是，头部顺时针转动控制对象延其世界坐标轴正方向运动，逆时针转动则控制对象延

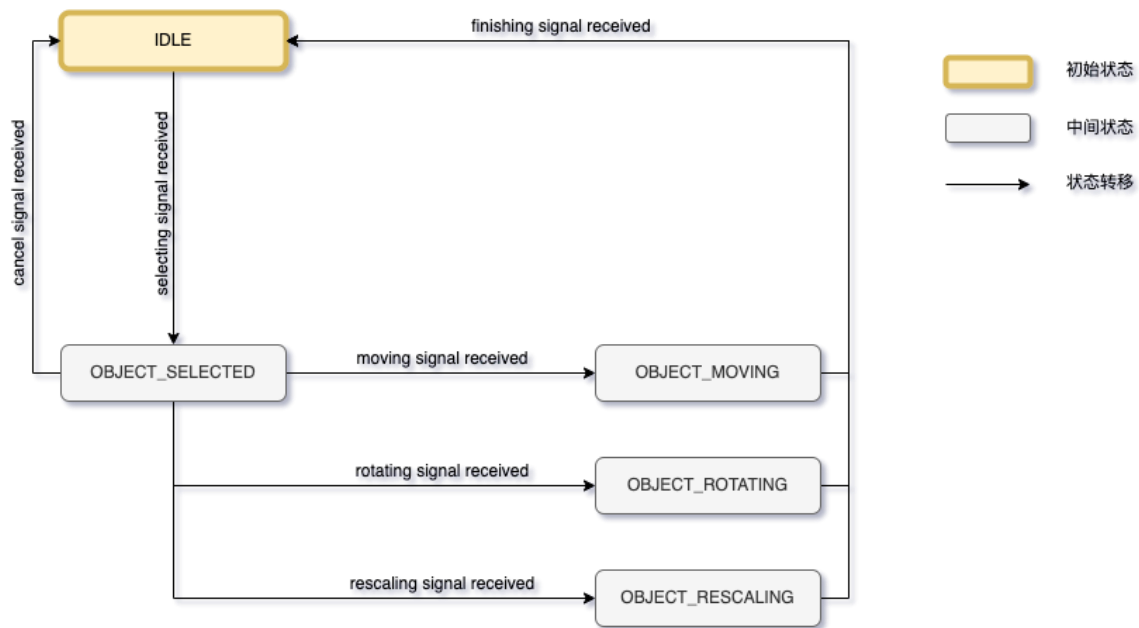


图 11 交互系统的有限状态机

其世界坐标轴负方向运动。具体的规则是，头部以相对坐标轴 Y 轴为轴转动控制物体在世界坐标轴的 Y 轴上运动，以相对坐标轴 Z 轴为轴转动控制物体在世界坐标轴的 X 轴上运动，以相对坐标轴 X 轴为轴转动控制物体在世界坐标轴的 Z 轴上运动。

在进入 OBJECT\_ROTATING 状态时和在该状态下，目前主要构思的一个交互模式为半闭双眼以进入，进入后仿照 Conner 团队外加一个虚拟的控制柄以辅助旋转<sup>[10]</sup>。

在进入 OBJECT\_RESCALING 状态时和在该状态下，目前主要构思的一个交互模式为快速完成一次眨眼紧接一次半闭双眼以进入，进入后同样外加一个虚拟的控制柄以辅助缩放。

## 4. 论文研究计划

- 2022.12 - 2023.2: 阅读文献，分析国内外研究现状，找准评估标准和可作比較的 baseline;
- 2023.2 - 2023.3: 对象操纵——位移：多种方法构思、实现和比较;
- 2023.2 - 2023.3: 对象操纵——旋转和缩放：多种方法构思、实现和比较;
- 2023.3 - 2023.4: 设计实验，开展 Pilot Study 和 User Study;
- 2023.4 - 2023.5: 分析实验结果，撰写论文。



## 5. 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.
- [18]P. Monteiro, G. Gonçalves, H. Coelho, M. Melo and M. Bessa. (2021). "Hands-free interaction in immersive virtual reality: A systematic review," in IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 5, pp. 2702-2713, doi: 10.1109/TVCG.2021.3067687.
- [19]Carruth, Daniel & Hudson, Christopher & Bethel, Cindy & Pleva, Matus & Ondáš,

- Stanislav & Juhár, Jozef. (2019). Using HMD for Immersive Training of Voice-Based Operation of Small Unmanned Ground Vehicles. 10.1007/978-3-030-21565-1\_3.
- [20]MIRZAEI, Mohammad & Chardonnet, Jean-Remy & Merienne, Frédéric & Genty, Ariane. (2014). Navigation and interaction in a real-scale digital mock-up using natural language and user gesture. ACM International Conference Proceeding Series. 2014. 10.1145/2617841.2620716.
- [21]Adhanom, Isayas & MacNeilage, Paul & folmer, eelke. (2023). Eye Tracking in Virtual Reality: a Broad Review of Applications and Challenges. Virtual Reality. 1-24. 10.1007/s10055-022-00738-z.
- [22]Sidenmark, Ludwig & Gellersen, Hans. (2019). Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection. 1161-1174. 10.1145/3332165.3347921.
- [23]Tanriverdi, Vildan & Jacob, Robert. (2001). Interacting With Eye Movements In Virtual Environments. Conference on Human Factors in Computing Systems - Proceedings. 10.1145/332040.332443.
- [24]Hansen, John & Rajanna, Vijay & MacKenzie, I. & Baekgaard, Per. (2018). A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. 1-5. 10.1145/3206343.3206344.
- [25]Luro, Francisco & Sundstedt, Veronica. (2019). A comparative study of eye tracking and hand controller for aiming tasks in virtual reality. 1-9. 10.1145/3314111.3318153.
- [26]Robert J. K. Jacob. (1990). What you look at is what you get: eye movement-based interaction techniques. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90). Association for Computing Machinery, New York, NY, USA, 11–18. <https://doi.org/10.1145/97243.97246>
- [27]Jacob, Rob & Stellmach, Sophie. (2016). What you look at is what you get: gaze-based user interfaces. interactions. 23. 62-65. 10.1145/2978577.
- [28]Kumar, Devender & Sharma, Amit. (2016). Electrooculogram-based virtual reality game control using blink detection and gaze calibration. 2358-2362. 10.1109/ICACCI.2016.7732407.
- [29]Pfeuffer, Ken & Mayer, Benedikt & Mardanbegi, Diako & Gellersen, Hans. (2017). Gaze + pinch interaction in virtual reality. 99-108. 10.1145/3131277.3132180.

- [30]Pai, Yun Suen & Dingler, Tilman & Kunze, Kai. (2019). Assessing hands-free interactions for VR using eye gaze and electromyography. *Virtual Reality*. 23. 10.1007/s10055-018-0371-2.
- [31]Qian, Yuan Yuan & Teather, Robert. (2017). The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality. 91-98. 10.1145/3131277.3132182.
- [32]Sidenmark, Ludwig & Clarke, Christopher & Zhang, Xuesong & Phu, Jenny & Gellersen, Hans. (2020). Outline Pursuits: Gaze-assisted Selection of Occluded Objects in Virtual Reality. 10.1145/3313831.3376438.
- [33]Majaranta, Päivi & Bulling, Andreas. (2014). Eye tracking eye-based human-computer interaction. *Advances in Physiological Computing*. 39-65. 10.1007/978-1-4471-6392-3\_3.
- [34]Boyer, Eric & Lorenceau, Jean & Portron, Arthur & Bevilacqua, Frédéric. (2017). Continuous Auditory Feedback of Eye Movements: An Exploratory Study toward Improving Oculomotor Control. *Frontiers in Neuroscience*. 11. 10.3389/fnins.2017.00197.
- [35]Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. (2018). Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the Workshop on Communication by Gaze Interaction (COGAIN '18)*. Association for Computing Machinery, New York, NY, USA, Article 1, 1–9. <https://doi.org/10.1145/3206343.3206349>
- [36]Mohan, Pallavi & Goh, Boon & Fu, Chi-Wing & Yeung, Sai-Kit. (2018). DualGaze: Addressing the Midas Touch Problem in Gaze Mediated VR Interaction. 79-84. 10.1109/ISMAR-Adjunct.2018.00039.
- [37]Hirzle, Teresa & Cordts, Maurice & Rukzio, Enrico & Bulling, Andreas. (2020). A Survey of Digital Eye Strain in Gaze-Based Interactive Systems. 1-12. 10.1145/3379155.3391313.
- [38]Purves D, Augustine GJ, Fitzpatrick D, et al., editors. (2001). *Types of Eye Movements and Their Functions*. Neuroscience. 2nd edition. Sunderland (MA): Sinauer Associates
- [39]P. C. Gloumeau, W. Stuerzlinger and J. Han. (2021). "PinNPivot: Object Manipulation Using Pins in Immersive Virtual Environments," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 4, pp. 2488-2494, doi: 10.1109/TVCG.2020.2987834.
- [40]Yu, Difeng & Lu, Xueshi & Shi, Rongkai & Liang, Hai-Ning & Dingler, Tilman & Velloso,

Eduardo & Goncalves, Jorge. (2021). Gaze-Supported 3D Object Manipulation in Virtual Reality. 1-13. 10.1145/3411764.3445343.

[41]Liu, Chang & Plopski, Alexander & Orlosky, Jason. (2020). OrthoGaze: Gaze-based Three-dimensional Object Manipulation using Orthogonal Planes. Computers & Graphics. 89. 10.1016/j.cag.2020.04.005.