

参赛队员姓名: 梅毅轩

中学: 上海中学

省份: 上海市

国家/地区: 中国

指导教师姓名: 姜忠鼎 吴奕明

论文题目: 虚拟现实头盔中面向眼球追踪
的复合式物体检测算法

虚拟现实头盔中面向眼球追踪的复合式物体检测算法

上海中学 梅毅轩

摘要: 虚拟现实和眼球追踪技术相结合, 极大的提升了虚拟现实的用户体验, 也推动了眼球追踪技术在各领域的应用。但系统精度的不足, 影响了虚拟现实眼球追踪技术进一步的推广和应用。针对虚拟现实眼球追踪系统精度不足, 软件优化差, 导致人机交互过程中误触发率高, 三维空间复杂场景下虚拟物体无法选中的问题, 设计了复合式物体检测算法。在算法中, 数据预处理过程中通过聚焦有效性检测, 降低了人机交互过程中误触发率; 通过对待检测物体 AOI 概率, 可见性概率、预测概率的计算, 得到待检测物体被选中的综合概率, 给出了一种三维空间复杂场景的物体检测算法, 改善了三维空间复杂场景下虚拟物体无法选中的问题。设计了多种实验场景和数据可视化分析软件, 对复合式物体检测算法的准确度进行了测试、分析和优化。复合式物体检测算法不仅可以应用于虚拟现实眼球追踪设备中, 也可广泛应用于集成眼球追踪技术的其他设备中。

关键词: 虚拟现实, 眼球追踪, 物体检测

Abstract: The combination of virtual reality and eye tracking technology greatly improves the user experience of virtual reality and promotes the application of eye tracking technology in various fields. However, the lack of system precision affects the further promotion and application of eye tracking in virtual reality. The cause of the issue is that eye trackers currently available in virtual reality is not precise enough and that the software optimization is also poor, which leads to a high false triggering rate in human-computer interaction as well as selection failure in a complex 3D scene. Aimed at the problem, a compound object detection algorithm is designed. In the algorithm, the false triggering rate in human-computer interaction is reduced by focusing validity testing during data preprocessing. To select objects, compound object detection algorithm computes comprehensive probability based on each object's AOI probability, visibility probability and prediction probability. We demonstrate the effectiveness of compound object detection algorithm by testing it in various complex 3D test scenes. Moreover, we designed a data visualization program to analyze the data and optimize the algorithm. Compound object detection algorithm can be used not only in virtual reality eye tracking devices, but also in other devices integrated with eye tracking technology.

Keywords: Virtual Reality, Eye Tracking, Object Detection

目 录

1. 引言	1
1.1 研究背景和前景.....	1
1.2 研究现状.....	2
1.3 现存问题.....	2
2. 系统工作原理.....	3
2.1 虚拟现实眼球追踪头盔原理.....	3
2.2 现有物体检测算法工作过程.....	4
2.3 复合式物体检测算法工作过程.....	5
3. 复合式物体检测算法.....	5
3.1 数据预处理.....	5
3.2 复合式物体检测.....	6
3.2.1 物体检测的原理.....	6
3.2.2 三维空间中AOI概率.....	8
3.2.3 三维空间中可见性概率.....	9
3.2.4 三维空间中预测概率.....	10
3.2.5 综合概率.....	11
4. 实验结果与分析.....	11
4.1 数据预处理的测试与分析.....	11
4.1.1 实验设计.....	11
4.1.2 比较及分析.....	12
4.2 算法精确度的验证.....	13
4.2.1 实验场景设定.....	13
4.2.2 实验流程.....	14
4.2.3 数据可视化分析工具.....	15
4.2.4 近距离场景的验证.....	16
4.2.5 远距离场景的验证.....	17
4.2.6 极端情况场景的验证.....	18
4.3 示例程序对算法的验证.....	18
5. 结论.....	19
参考文献	20
致谢.....	21

1 引言

1.1 研究背景和前景

眼球追踪技术, 通过对用户眼部动作的采集和分析, 获取用户关注度信息, 在心理研究, 用户体验研究, 教育辅助, 消费习惯研究, 医疗辅助等领域有着广泛的应用前景。在眼球追踪领域不乏有 Tobii、SMI、七鑫易为等厂商早已布局[12][13][14]。但大部分产品, 主要用于医疗眼控辅具, 例如 Tobii 公司的眼动仪, 七鑫易为生产的眼控系列产品。这些设备价格昂贵, 且需要专业硬件平台才能正常工作, 平台兼容性差, 不适合大规模的推广和应用。

随着虚拟现实技术发展, 交互性、清晰度、真实感将会成为用户关注的重点。眼球追踪技术与虚拟现实的结合将极大的提升用户的体验。现实生活中, 人们习惯以头动控制视野, 眼控锁定目标的方式进行交互, 因此眼球追踪和虚拟现实相结合, 符合用户的习惯, 比其他虚拟现实交互方式效率更高, 体验更自然。通过眼球追踪得到用户的关注点, 可以在用户关注区域以更丰富的细节进行渲染, 周边区域逐步降低渲染分辨率, 从而有效减少计算量, 降低虚拟现实设备的延迟和眩晕感, 并大幅提升渲染画质, 增强虚拟现实场景清晰度和真实感。

虚拟现实头戴式显示器性价比高、便携性好, 结合眼球追踪系统, 采集用户关注度信息将变得更加容易, 极大的推动了眼球追踪技术各个领域中的应用。例如: 根据虚拟现实眼球追踪设备上记录的关注点轨迹, 为眼动测试提供最真实的数据基础, 这些数据可应用于用户研究, UI 设计, 广告设计等领域; 虚拟现实中的眼动广告可根据人眼的关注点来完成画面的转换, 关注点信息的放大等, 并记录广告被关注的次数及轨迹, 用于广告分析; 眼控交互和虹膜识别的结合, 可以完成商品的选择和用户身份确认, 用户在虚拟现实环境中购物流程更加自然和流畅; 虚拟现实中, 眼球追踪可以在被测试者无察觉的情况下记录真实的关注点, 应用于各种心理疾病的研究, 或者应用于测谎、心理教学等领域。

鉴于虚拟现实和眼球追踪结合的种种优势, 各大厂商相继推出自己的产品, 2017 年年初, Fove 公司首先推出了带眼球追踪仪的虚拟现实头戴式显示器(简称虚拟现实头盔)Fove0[15], 随后 SMI, 七鑫易为等公司先后推出了眼球追踪开发者套件[13][16], 可以单独或者与现有普通头戴式显示器配合使用, 极大的推动了眼球追踪技术的推广和应用。

基于虚拟现实平台所研究的眼球追踪各种技术和算法, 不仅可以用于虚拟现实各个场景中, 也可以移植到其他眼控产品中, 用于判断人眼状态和注视点, 具有广泛的应用前景。例如: 在疲劳驾驶或视线偏移状态时进行预警, 并配合 HUD 技术提供更好的车载娱乐交互体验; 通过眼球动作完成密码的输入, 既防止自己的输入动作被偷窥, 又不会在键盘上留下指纹; 模拟真实的人眼对焦机制, 解决 AR 穿戴式智能眼镜的交互问题, 通过眼球动作实现对焦、拍照等指令下达; 通过眼球动作, 完成对电脑的操作, 为有沟通障碍的患者的治疗和日常护理提供方便。随着眼球追踪技术不断的发展完善, 它将广泛应用于智慧医疗、VR/AR、智能手机、广告传媒、智能汽车、机器人、航空航天等各行各业中。

1.2 研究现状

虚拟现实眼球追踪相关的应用开发和用户体验, 建立在高效而准确的获取用户关注点信息的基础上, 这就对眼球追踪系统的精度提出了极高的要求。眼球追踪系统的精度受设备精度的影响最大, 设备精度主要取决于眼球追踪仪内置相机的拍摄精度和内置瞳孔检测算法精度。人机交互系统设计也是影响眼球追踪系统精度的重要因素, 并且它的优化设计可以弥补设备精度的不足对眼球追踪系统精度的影响。

目前虚拟现实眼球追踪系统的研究, 主要集中于眼球追踪设备硬件设计以及瞳孔检测算法优化。在硬件设计方面, Fove 公司第一时间推出了内置眼球追踪仪的 Fove0 虚拟现实头戴式显示器, 推出时在同类设备中技术最先进, 效果最好。七鑫易维推出了可用于 HTC Vive 头戴式显示器中的眼球追踪模组, 同样具有很好的效果。在瞳孔位置检测方面, Thiago Santini 等人提出了 PuRe 算法[8], 对瞳孔进行重建, 提高了瞳孔检测的准确度。Wolfgang Fuhl 等人构建了卷积神经网络 Pupil Net v2.0[9], 使用了大量人工标定的眼部图片对其进行训练, 相较传统的图像处理算法, 该方法具有更高的准确度。Fove 公司的 Fove0 头戴式显示器中也集成了瞳孔检测算法, 其优点在于无需大量数据进行训练, 简单校准之后即可使用。

在人机交互系统的软件和算法的研究方面, 国际上研究相对较少, 较差的软件优化严重影响了虚拟现实眼球追踪的用户体验。目前 FOVE 公司设计了基于直接选中的物体检测算法, 在 2017 年推出时, 是当时最准确的虚拟现实头戴式显示器眼球追踪物体检测算法, 但在复杂三维场景中使用效果并不好。Thies Pfeiffer 等人研究了部分遮挡情况下的识别问题[10], 但对于距离较远以及物体密集等复杂场景并没有系统的研究。EyeFluence 公司在景深渲染和交互系统的设计上进行了一些研究, SMI 公司在景深渲染方面也做了一些研究, 截至目前, 两家公司都没有发布任何设计细节。

鉴于硬件设备和瞳孔检测这两方面的研究相对成熟, 人机交互系统方面的优化设计较为欠缺, 本文直接采用 Fove 公司的 Fove0 内置眼球追踪仪的虚拟现实头戴式显示器做为硬件平台, 研究人机交互系统的相关技术和算法, 提升虚拟现实眼球追踪系统的精度, 改善虚拟现实眼球追踪的用户体验。

1.3 现存问题

目前虚拟现实眼球追踪系统的研究中, 人机交互技术与算法的研究较为欠缺, 导致眼球追踪系统精度的恶化, 主要表现在人机交互过程中物体检测、选中和触发的准确率差(此处的物体定义包括虚拟现实中的菜单控件, 物体, 虚拟人物等)。当用户扫视一个有许多控件(例如按钮, 输入框)的界面时, 会经常错误地触发界面上的控件。而在距离较远, 物体密集, 有遮挡等三维空间复杂场景下, 几乎无法实现用户关注物体的检测和选中, 用户有很大概率无法选中三维空间中 50 米外的人体大小的物体。致使虚拟现实眼球追踪的人机交互体验较差, 这些都大大的影响了眼球追踪技术在虚拟现实环境中的推广和应用。

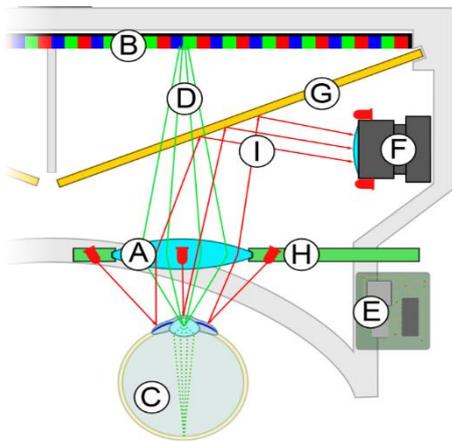
针对这些问题, 本文提出了复合式物体检测算法, 通过对眼球追踪输入数据的预处理,

改善误触发的问题。通过计算待检测物体被选中的综合概率，决定物体的选中，改善物体无法检测和选中的问题。为了验证算法的效果，本文设计多种实验场景和数据可视化分析软件，对复合式物体检测算法进行测试、分析和优化。本文所设计的复合式物体检测算法，相较 Fove0 自带的直接选中物体检测算法，有效的改善眼球追踪人机交互中误触发和物体无法检测和选中的问题，极大的提升了虚拟现实眼球追踪系统精度，改善人机交互的体验。

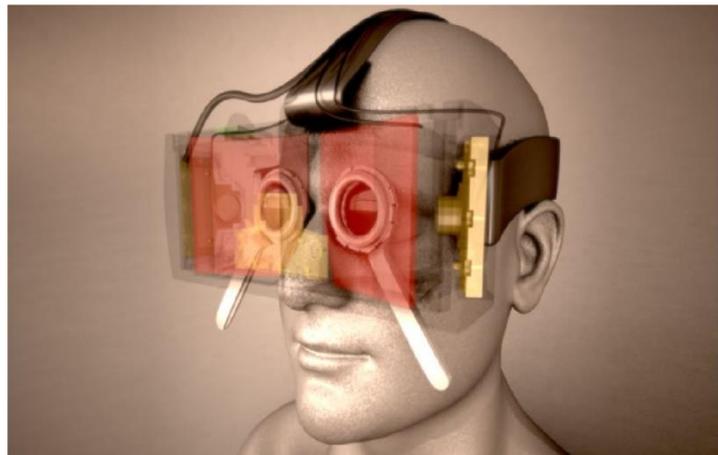
2. 系统工作原理

2.1 虚拟现实眼球追踪头盔原理

如图 2.1 (a) 所示，为了借助眼球追踪实现人机交互，眼球追踪设备会通过左右两侧的红外相机 (F) 来记录用户的眼部数据 (图 2.1 (a) 仅给出了设备右侧设计图，设备左右两侧是对称的)。红外相机捕获的图像经过眼球追踪头盔自带的芯片 (E) 中的瞳孔检测算法的处理，可以确定瞳孔的位置，并依靠检测到的瞳孔位置估计注视的方向。当眼球追踪仪记录到双眼的眼部数据时，人机交互系统就可以根据这一数据，计算用户三维空间中的焦点的位置和深度信息，得出用户正在关注的物体，从而响应用户操作，实现人机交互。



(a) 眼球追踪头盔设计图



(b) 眼球追踪头盔佩戴图

图 2.1 眼球追踪仪设备原理图： 图 (a)：菲涅尔透镜 (A)，显示屏 (B)，眼球 (C)，可见光光线 (D)，数据处理芯片 (E)，眼球追踪相机 (F)，分光镜 (G)，红外光源 (H)，红外光光线 (I)。 图 (b)：用户佩戴虚拟现实眼球追踪头盔的示意图

2.2 现有物体检测算法工作过程

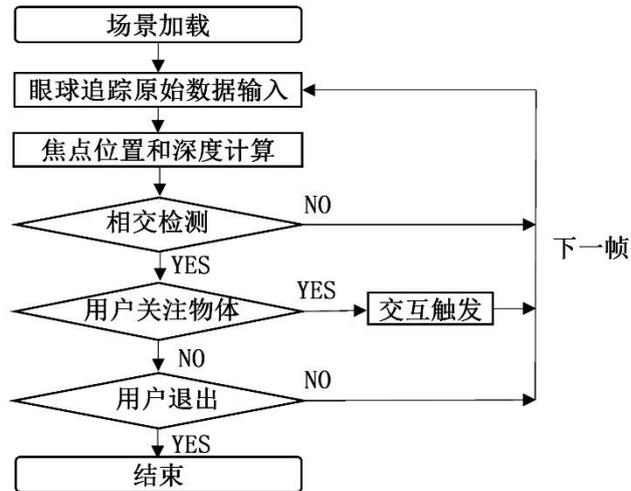


图 2.2 直接选中物体检测算法流程图

如图 2.2 所示，在每一帧中，交互系统通过眼球追踪仪获取用户眼球原始数据，其中包含代表用户两眼注视方向的两条射线和用户两眼在场景中位置。交互系统对这些数据进行处理，计算出用户的焦点位置和深度等信息。物体检测算法根据这些数据进行相交检测，相交检测主要分为两个部分，首先是位置测试，检测用户焦点的屏幕坐标是否位于某一个物体上，如果确认，则会进行深度测试，判断用户的焦点深度是否在物体上，如果位置和深度测试都通过，则判定用户这一帧的焦点位于通过测试的物体上，此物体是用户关注的物体。这一过程是目前眼球追踪系统通用的物体选中方法，在本文被定义为直接选中物体检测算法。直接选中物体检测算法有其不足之处，主要体现在以下两个方面：

1. 误触发率较高，当用户扫视界面时，焦点会在界面上连续移动，如果某一个控件面积较大，用户焦点从进入控件到离开控件所需的时间大于触发阈值，就会造成误触发。改善此问题的一种方式是提高触发阈值，用户焦点停留在控件上的时间更长，才能完成触发。但由于眼球追踪设备精度不足，存在用户注视控件时焦点抖动现象，致使屏幕边缘处或面积较小的控件，难以触发，触发阈值的提高，将导致这类控件触发难度进一步加大。同时，对于较为熟练的用户，过长的触发阈值会降低操作的流畅度，破坏用户体验。

2. 直接选中物体检测算法在距离较远，物体密集，有遮挡等三维空间复杂场景下，由于设备中瞳孔检测算法的误差，导致计算出的用户焦点位置偏离用户实际关注的物体位置，无法完成相交检测，也就无法正确的返回被选中的物体。这种情况下，就算用户盯着某一个物体看了很久，直接选中物体检测算法也无法检测到用户正在关注这一物体。

2.3 复合式物体检测算法工作过程

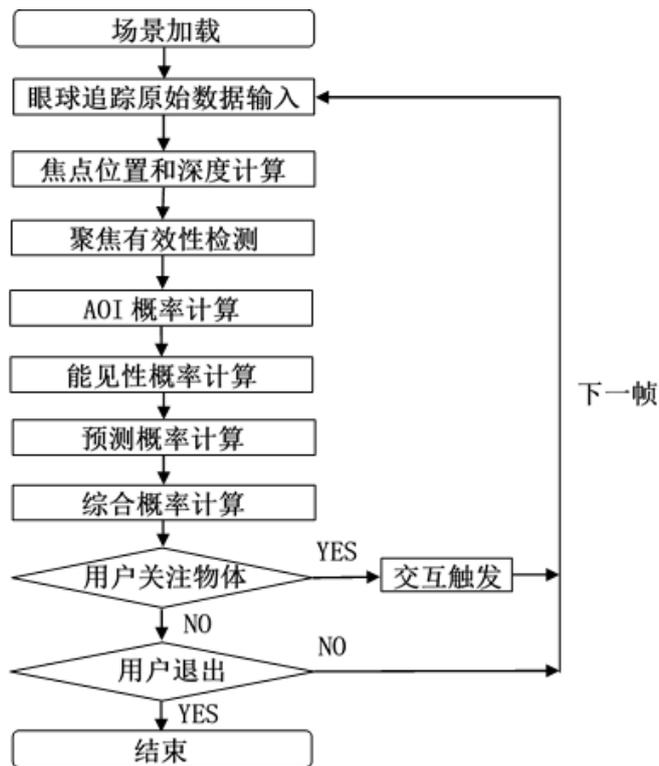


图 2.3 复合式物体检测算法流程图

为了改善 2.2 节中所述问题，本文设计了复合式物体检测算法，如图 2.3 所示，为了减少误触发的发生，增加了聚焦有效性检测，对眼球追踪输入数据进行预处理，来判定用户的焦点属于扫视界面时的焦点，还是已经聚焦物体时的焦点，只有聚焦有效性检测成功，才有可能进行交互触发。聚焦有效性检测，可有效降低误触发率，其设计细节将在 3.1 节中讨论，测试结果将在 4.1 节展示。

为了改善直接选中物体检测算法在距离较远，物体密集，有遮挡等三维空间复杂场景下无法选中物体的问题，复合式物体检测算法对被检测物体的 AOI 概率，可见性概率，预测概率进行计算，得到待检测物体被选中的综合概率，基于综合概率决定物体是否被选中，有效提高了物体检测和选中的准确性。AOI 概率，可见性概率，预测概率以及综合概率的定义及计算方法，在 3.2 节中将会详细讨论，复杂场景下复合式物体检测算法准确性的测试结果将在 4.2 节展示。

3. 复合式物体检测算法

3.1 数据预处理

复合式物体检测算法通过聚焦有效性检测，对眼球追踪输入数据进行预处理，降低误触发率，提高虚拟现实眼球追踪系统精度。聚焦有效性检测原理是：用户扫视界面时用户的焦

点信息和聚焦物体时的焦点信息最大的差别在于焦点移动速度, 根据大量的实验数据得出如下经验值, 聚焦物体时最大焦点移动速度为扫视界面时焦点移动速度的三分之一。如果在一段时间内焦点的移动速度小于聚焦物体时最大焦点移动速度, 可以判断用户进入聚焦状态。结合待检测物体被选中综合概率的计算结果, 决定物体选中是否有效, 可以有效降低物体的误触发率。



图 3.1 记录用户扫视速度的实验场景图

由于每个人的反应时间、扫视界面时的焦点移动速度都不一样, 必须要进行用户适配, 聚焦有效性检测才能达到更好的效果。用户适配的方法是通过扫视界面速度测试, 获取用户浏览界面时的正常速度。如图 3.1 所示, 在测试中, 测试程序让用户数出界面上的文字中的字母 A 的数量, 通过这一方法, 可以获得用户在不主观干预的情况下, 浏览界面的平均速度。这一段文字中包含了稀疏和密集的情况, 尽可能的模拟了真实使用场景。通过这一实验可以获得用户扫视界面的焦点移动速度。

在复合式物体检测算法中, 只有当用户的焦点移动速度小于聚焦物体时最大焦点移动速度, 用户的焦点停留才会被判定有效, 这个焦点所关注物体才可能被触发。聚焦有效性检测可以有效降低误触发率, 并使得触发阈值可以安全的设定为一个较小的值, 提高交互系统的流畅度。

3.2 复合式物体检测

复合式物体检测通过计算待检测物体被选中的综合概率, 决定待检测物体的选中, 改善了三维空间复杂场景下物体无法检测和选中的问题。具体算法的实现, 本节以下内容将会详细论述。

3.2.1 物体检测的原理

为了改善距离较远或目标物较小时, 相交检测无法实现物体选中问题, 基于 AOI (Area Of

Interest) 的物体检测算法被提出[2]。AOI 模型来源于 Stellmach 等人提出的 OOI (Object of Interest) 模型[11]。OOI 是一个用户注意力分配模型, Stellmach 等人将其用于三维空间中用户的注意力分析中。而 AOI 则是一个简化后的二维注意力分配模型, 这个模型认为: 当用户关注一个物体时, 所关注的物体位于焦点周围的一个范围之内, 这个范围被定义为注意力半径。根据此模型, Alam 等人提出了公式 3.1 中所示的 AOI 概率计算方法, 并设计了二维平面上的物体检测算法[2]。

$$Gs_i = 1 - \min\left(1, \frac{d}{r}\right) \quad (3.1)$$

公式 3.1 中, Gs_i 代表物体 i 被选中的概率。d 为物体与焦点间的距离, r 为注意力半径。公式 3.1 可以很方便的计算注意力半径内每一个物体被选中的概率。

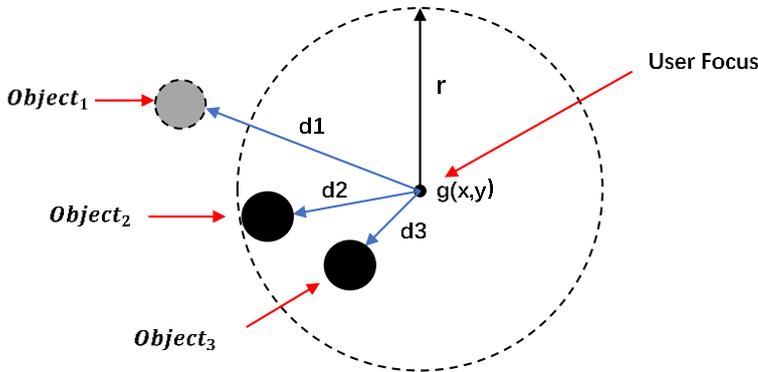


图 3.2 基于 AOI 的物体选中示意图

图 3.2 展示了基于 AOI 的物体检测算法的选中方式, 左侧三个圆形物体是待选中的物体, 半径为 r 的虚线圆为用户的注意力半径覆盖区域, $g(x, y)$ 为这一帧中用户的焦点位置, d 为焦点到可被选中物体的距离, r 为用户的注意力半径, 这一数值需要预先确定。由于距离 d_1 大于注意力半径 r, 所以 $Object_1$ 的 AOI 概率等于 0, $Object_1$ 不可能被选中。 d_2, d_3 小于注意力半径 r, 所以 $Object_2, Object_3$ 的 AOI 概率大于 0, 由于 d_3 小于 d_2 , 根据公式 3.1 计算出 $Object_3$ 的 AOI 概率大于 $Object_2$ 的 AOI 概率, 因此 $Object_3$ 被选中。这种检测算法的本质是选中用户注意力半径覆盖区域中离焦点最近的物体。它很好的弥补了眼球追踪仪设备精度不足的问题, 使得选中变得更加容易。如果注意力半径这一参数选择合理, 这个算法在二维场景下有很好的效果。

在三维空间, 该算法有其不足。首先, 它没有考虑到用户不能透视这一限制, 如果距离焦点最近的物体被其他物体部分或完全遮挡, 则无法判断用户在关注哪个物体。其次, 此算法的注意力半径 r 不能自动调整, 在不同距离下选择同样大小的注意力半径, 会导致近距离下的误选或远距离时的无法选中。综上所述, 这一算法并不适用于三维空间中的物体检测。

为了提高二维空间中的物体选中成功率, Alam 等人又提出了一种预测式的物体检测算法[2], 在二维空间中有着很好的效果。然而, 这种预测式算法需要对物体的转移概率预先设定,

仅适用于静态场景。并且此算法没有考虑物体到用户距离对物体检测的影响, 无法应用在三维空间中。

针对二维空间物体检测算法应用于三维空间场景中的不足, 在二维空间基于 AOI 的概率基础上, 本文设计了三维空间基于 AOI 的概率, 在二维空间预测式物体检测算法的基础上, 本文设计了三维空间的预测概率, 结合计算物体能见度的能见性概率, 最终得到物体被选中的综合概率, 每一帧中综合概率最大的物体会被选中。上述概率的具体计算方法在 3.2.2-3.2.5 节将会详细介绍。

3.2.2 三维空间中 AOI 概率

在 3.2.1 节中提到, Alam 等人提出的基于 AOI 的物体检测算法不适用于三维空间一个原因, 在于它不能根据焦点与用户间距离自动调整注意力半径 r 的大小, 进而导致注意力范围过大或过小。为解决这一问题, 本文提出了三维空间中基于 AOI 的概率。

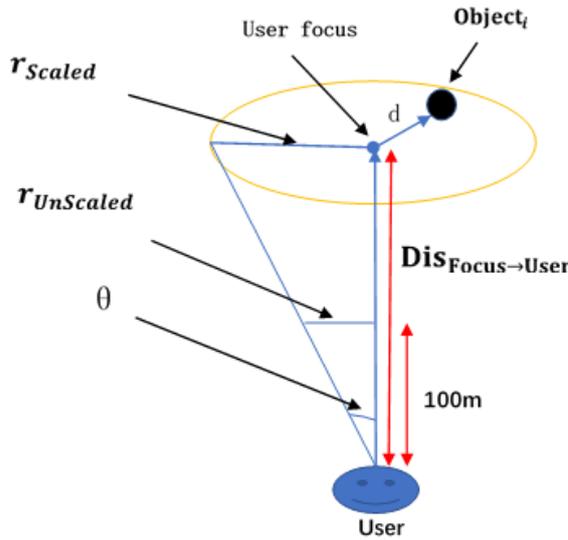


图 3.3 三维空间中注意力半径计算示意图

$$r_{Scaled} = \frac{Dis_{Focus \rightarrow User}}{100} * r_{Unscaled} \quad (3.2)$$

公式 3.2 为当前焦点与用户距离下, 用户的注意力半径 r_{Scaled} 的计算公式, 如图 3.3 所示, $r_{Unscaled}$ 为焦点距离用户 100 米时用户的注意力半径, θ 为用户的注意力覆盖角, 大量实验表明 θ 约为 15 度左右, 所以 $r_{Unscaled}$ 被设定为 30 米。 $Dis_{Focus \rightarrow User}$ 为当前焦点与用户之间的距离。

$$Gs_i = 1 - \min\left(1, \frac{d}{r_{Scaled}}\right) \quad (3.3)$$

公式 3.3 为三维空间中基于 AOI 的概率计算公式, 其中 Gs_i 代表物体 i 在这一帧中被选中的概率, d 为物体 i 在这一帧中与焦点间的距离, r_{Scaled} 为当前焦点与用户距离下, 用户的注

注意力半径，由公式 3.2 计算得到。

三维空间的 AOI 概率，根据物体距离用户的远近，自动调节注意力半径 r_{Scaled} 的大小，使复合式物体检测算法在三维空间物体远近距离变化的情况下，都具有较高准确率。

3.2.3 三维空间中可见性概率

在 3.2.1 节中提到，Alam 等人提出的基于 AOI 的物体检测算法不适用于三维空间另一个原因，没有考虑物体被遮挡的情况。物体被部分或全部遮挡情况下，会导致距离焦点很近但被障碍物挡住的物体被错误的选中。针对物体被遮挡的情况，本文提出了基于可见性测试的可见性概率。通过可见性测试，可以体现物体可见性对物体选中的影响。

$$VS_i = \frac{T_{visible}+1}{T_{all}+1} \quad (3.4)$$

公式 3.4 为可见性概率公式。 VS_i 为物体 i 的可见性概率， $T_{visible}$ 为物体可见的测试点数量， T_{all} 为物体所有的测试点数量。为了防止出现物体部分可见，但由于其他物体挡住了所有测试点造成可见性判断出现偏差的情况，分子和分母都加上了 1，确保这一项不会为 0。测试点的选取数量越多，可见性测试的精确度就越高，上述情况出现的概率也越小，但相应的对所有测试点进行遍历的时间开销也更大。一个折中的做法是以模型的部件作为测试点，对每个部件进行可见性测试，如图 3.4 为示例模型所有的测试点分布，橙色的线条为模型各部件的边缘，如图 3.5 为示例模型对应的部件列表，测试时对列表中的每一个部件进行可见性测试。对于每一个测试点，先测试测试点是否位于用户视角视椎体之内，如果测试点通过了这一测试，说明测试点在用户视角的渲染范围内。然后，对测试点进行深度测试，如果通过了深度测试，说明测试点没有被遮挡，那么这个测试点就是可见的。



图 3.4 示例模型测试点分布

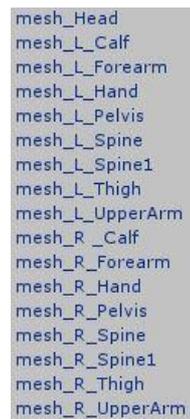


图 3.5 示例模型部件列表

通过对可见性概率的计算，得到物体在遮挡情况下被选中的概率，结合 AOI 概率的结果，可以进一步提升复合式物体检测算法的准确率。

3.2.4 三维空间中预测概率

预测概率体现了用户在前一帧中对各物体的注意力分布, 对后一帧物体选中的影响。在 3.2.1 节中提到, 二维空间预测式的物体检测算法需要对物体的转移概率预先设定, 仅适用于静态场景, 并且没有考虑物体到用户距离对物体检测的影响, 无法应用在三维空间中。本文参考 Alam 等人提出的二维预测式模型设计[2], 提出了三维空间中预测概率, 考虑物体间距离、物体到用户距离对转移概率的影响, 自动计算转移概率, 适用于三维的动态场景。

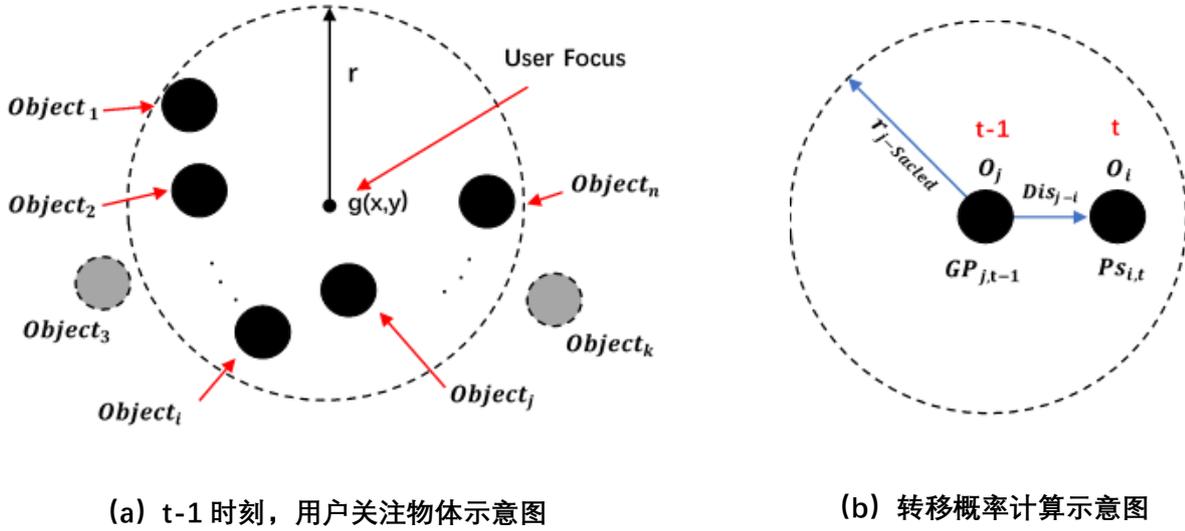


图 3.7 预测概率计算示意图

由于用户在场景中的焦点的移动是连续的, 所以用户在关注过一个物体之后, 再关注与之最近的相邻物体的可能性更大。如图 3.7a 所示, 如果物体在上一帧 t-1 时刻关注物体 j 后, 在当前帧 t 时刻, 关注物体 i 可能性最大。如果可以确定 t-1 时刻物体 j 是唯一被关注的物体, 如图 3.7b 所示, 上一帧 t-1 时刻用户真实焦点则位于物体 j, 考虑焦点位于物体 j 位置时的用户注意力半径, 物体 j 和物体 i 之间的距离, 可以得到当前帧 t 时刻焦点从物体 j 转移到物体 i 的转移概率。

$$Tr_{j \rightarrow i} = 1 - \min\left(1, \frac{Dis_{j \rightarrow i}}{r_{j-Scaled}}\right) \quad (3.5)$$

公式 3.5 中 $Tr_{j \rightarrow i}$ 为转移概率, 计算了焦点从物体 j 转移到物体 i 的可能性, 其中 $Dis_{j \rightarrow i}$ 为物体 j 和物体 i 之间距离, $r_{j-Scaled}$ 为焦点位于物体 j 时的用户注意力半径, 计算方法见 3.2.2 节。如果 $Dis_{j \rightarrow i}$ 小于 $r_{j-Scaled}$, 物体之间距离小于用户注意力半径, 那么 $Tr_{j \rightarrow i}$ 的取值范围为 (0, 1], $Tr_{j \rightarrow i}$ 的取值随物体之间距离的增加而减小; 如果 $Dis_{j \rightarrow i}$ 大于 $r_{j-Scaled}$, 物体之间距离超过用户注意力半径, 那么 $Tr_{j \rightarrow i}$ 的取值为 0, 用户的注意力不可能从物体 j 转移到物体 i。

如图 3.7b 所示, 如果确定物体 j 为 t-1 时刻唯一被关注选中的物体, 那么物体 i 在当前帧 t 时刻的预测概率 $PS_{i,t}$ 就是 $Tr_{j \rightarrow i}$ 。但根据 AOI 概率定义, 如图 3.7a 所示, 在 t-1 时刻, 并不能完全确定用户关注的物体, 注意力半径内的任何一个物体都存在被用户关注的可能性, 因此计算 t 时刻物体 i 的预测概率, 要考虑到注意力半径内所有物体在 t-1 时刻被关注选中

的综合概率(基于 AOI 概率, 可见性概率, 预测概率之积, 计算方法见 3.2.5 节), 以及这些物体到物体 i 的转移概率。因此可以对所有上一帧 $t-1$ 时刻综合概率 $Gp_{j,t-1} > 0$ 的物体 j ($1 \leq j \leq n, j \neq i$) 到物体 i 的转移概率 $Tr_{j \rightarrow i}$, 做加权平均, 加权值采用综合概率 $Gp_{j,t-1}$ 。公式 3.6 给出了计算当前帧 t 时刻, 物体 i 的预测概率的公式。

$$Ps_{i,t} = \frac{\sum_j Gp_{j,t-1} Tr_{j \rightarrow i}}{\sum_j Gp_{j,t-1}} \quad 1 \leq i \leq n, 1 \leq j \leq n, i \neq j \quad (3.6)$$

公式 3.6 计算了物体 i 的预测概率。如上文所述, $Tr_{j \rightarrow i}$ 代表用户关注的对象从物体 j 转移到物体 i 的可能性, $Gp_{j,t-1}$ 代表在上一帧 $t-1$ 时刻物体 j 被关注的综合概率。公式 3.6 考虑了上一帧 $t-1$ 时刻中所有可能被用户关注的物体 j , 对当前帧 t 时刻物体 i 被关注选中的影响。如图 3.7a 所示, 物体 3 和物体 k 这类物体, 在上一帧 $t-1$ 时刻处于用户注意力范围之外, 被关注选中综合概率为 0, 因此对于当前帧 t 时刻物体 i 的预测概率贡献为 0。

预测概率考虑到用户焦点移动的连续性, 通过上一帧物体被关注选中的综合概率以及转移概率, 来预测当前帧的物体被关注选中的概率, 确保连续几帧中被选中的物体的变化不会太剧烈, 提高复合式物体检测算法的准确率。

3.2.5 综合概率

$$Gp_i = Gs_i * Vs_i * Ps_i \quad (3.7)$$

公式 3.7 计算了物体 i 在这一帧中的被选中的综合概率 Gp_i 。综合概率由三维空间中 AOI 概率 Gs_i , 可见性概率 Vs_i , 预测概率 Ps_i 三者相乘得出。AOI 概率确保最终被选中的物体在用户的注意力范围之内; 可见性概率确保被选中的物体未被大面积遮挡; 预测概率确保两帧间被选中的物体的关联性。每一帧中综合概率最大的物体才可能被选中, 综合概率充分考虑了各种因素对物体选中的影响, 可以极大的提高复合式物体检测算法在三维空间中各种复杂场景下的物体选中准确率。

4. 实验结果与分析

4.1 数据预处理的测试与分析

4.1.1 实验设计

为了评估数据预处理中聚焦有效性检测的效果, 本文设计了如下实验。在实验中, 被测者根据一系列连续的指令, 看屏幕上不同区域的按钮。在扫视过程中误触发的所有控件都会被记录, 最终的实验结果以一维时间序列的热度图形式展现。实验场景如图 4.1 和图 4.2 所示, 两幅图均截自本文设计的示例程序, 仅为部分实验场景。实验场景中包含按钮, 滑动条, 选项, 文本框等大小不一的控件。测试中两套物体检测算法同时运行, 使用了聚焦有效性检测的物体检测算法记录触发数据并负责控件触发; 没有使用聚焦有效性检测的物体检测算法仅记录触发数据。在实验场景中, 需要连续触发的控件并不都是相邻放置, 在需要连续触发

的控件间的焦点移动路径上有可能存在其他控件，因此可以有效证明聚焦有效性检测对于减少焦点移动时的误触发的作用。



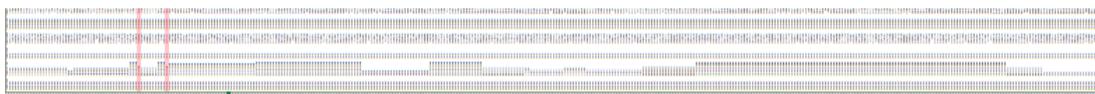
图 4.1 实验场景 1



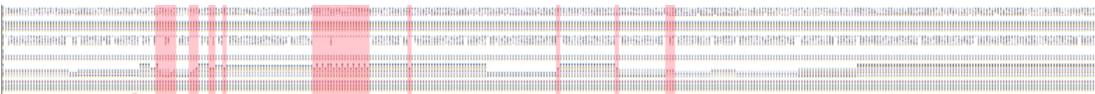
图 4.2 实验场景 2

4.1.2 比较及分析

图 4.3 展示了聚焦有效性检测的测试结果。图 4.3(a)为使用了聚焦有效性检测的物体检测算法返回数据做成的热度图，图 4.3(b)为未使用聚焦有效性检测的物体检测算法返回数据做成的热度图。图中每一列都代表一帧中物体检测算法返回的用户正在看的控件，红色的部分为根据误触发记录标定的误触发控件，根据误触发控件的数量和分布，可以分析得出算法的准确度。



(a) 使用聚焦有效性检测的数据热度图



(b) 未使用聚焦有效性检测的数据热度图

图 4.3 聚焦有效性检测的测试结果

由图 4.3 的实验结果可以看出，使用聚焦有效性检测对眼球追踪输入数据进行预处理后，误触发的数量会很大程度的减少。仅在极少帧中由于用户视线停留时间过长，而发生了误触发。由此可见，聚焦有效性检测对于误触发现象的减少是有效的。

4.2 算法精确度的验证

4.2.1 实验场景设定

本节将会测试各种场景对于复合式物体检测算法准确率的影响。实验场景依靠三个参数来控制：物体距离，物体密集程度，物体遮挡程度。物体距离分为 20-50 米和 50-100 米两种情况，分别模拟近距离场景和远距离场景。

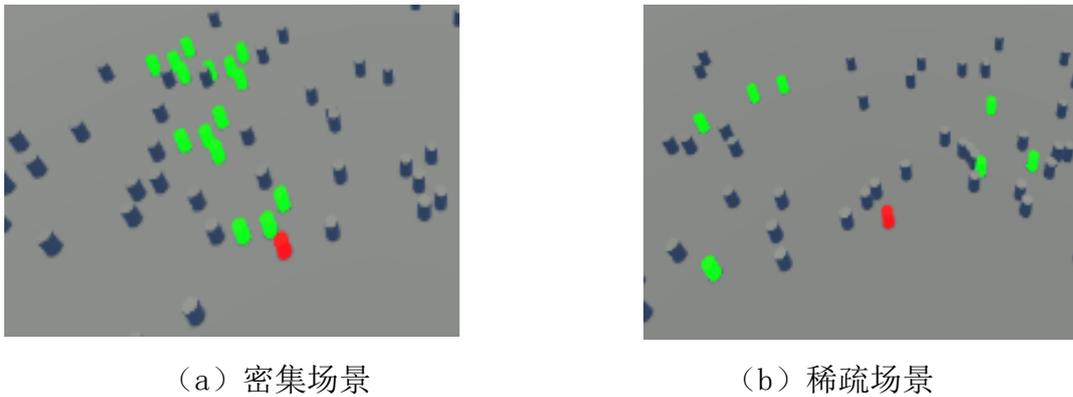


图 4.4 物体密集程度示意图

物体密集程度分为密集和稀疏两种情况，在物体密集的情况下，可被选中物体以四个一簇的形式分布，如图 4.4 (a) 所示，图中红色的物体为用户需要选中的物体，周围三个绿色的可被选中物体做为障碍物，完成干扰选中的目的，蓝色物体是不可选中的障碍物，实现遮挡的作用。而在物体稀疏的情况下，如图 4.4 (b) 所示，红色和绿色标注的可被选中物体分散分布。

物体遮挡程度分为三种情况：障碍物少、障碍物一般、障碍物多，以物体的密度来区分，物体距离 20-50 米场景下，对应遮挡程度的三种情况，在 20-50 米的圆环区域中分别放置 30, 100, 200 个物体。物体距离 50-100 米的场景下，对应遮挡程度的三种情况，在 50-100 米的圆环区域中分别放置 108, 362, 763 个物体。保证不同距离场景下，三种遮挡程度的物体密度基本相同。

图 4.5 (a) 为距离 20-50 米，物体总数 200 个，可被选中物体稀疏分布的场景图，如图 4.5 (a) 所示，用户位于环形区域的中心，200 个物体成环形围绕用户分布，红色物体为用户当前需要选中的物体，绿色物体是可被选中的物体，蓝色物体是完成遮挡作用的不可选中的障碍物。图 4.5 (b) 为通过位于环形区域中心的用户视角看到的场景分布。在所有场景中，可被选中物体始终保持在 30 个，其中当前需要选中的红色物体 1 个，可被选中的绿色物体 29 个，场景中蓝色物体不可被选中。蓝色物体和绿色物体一起做为障碍物，去遮挡当前需要选中的红色物体。将距离远近，密集程度，遮挡程度三种参数全排列，总共有 12 组实验，每组实验中物体在满足实验场景参数限制的情况下随机分布。

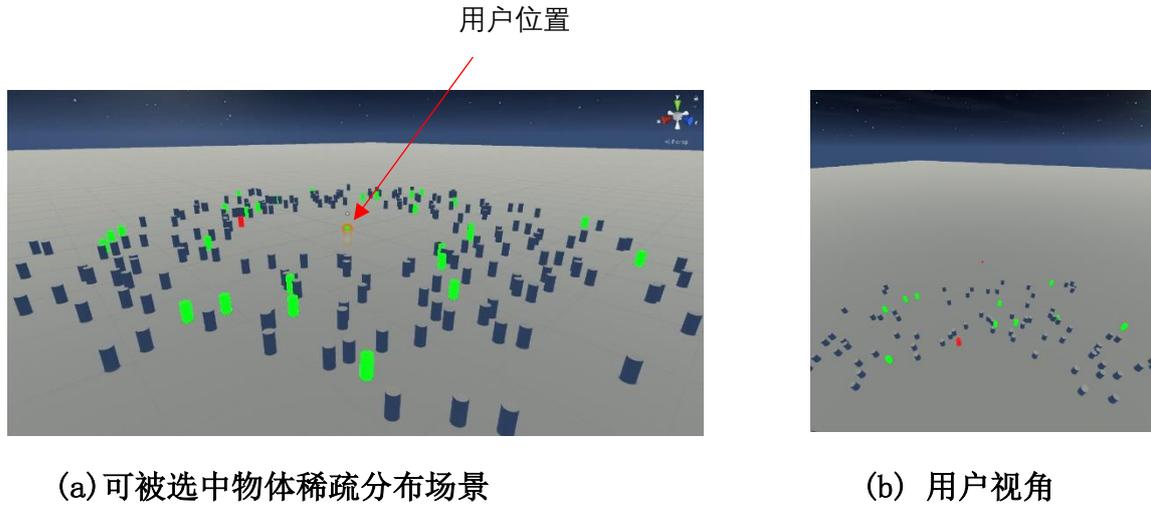


图 4.5 实验场景图：距离 20-50 米，物体总数 200 个

最后，还设置一组极端情况对算法进行测试，在这组测试中，物体距离为 100 米到 200 米，可被选中物体密集程度为密集（四个一簇），物体的遮挡程度为障碍物多，场景中总共有 3000 个障碍物。当然，这一场景在实际应用中极少出现，实验结果仅供参考。在更远或更密集的情况下，用户无法看清单个物体，因此没有分析价值。

4.2.2 实验流程

表 4.1 物体选中的测试数据

测试次数	选中物体编号		
	直接选中算法选中物体	用户真实选中物体	复合式算法选中物体
1	6	6	6
2	3	3	3
3	0	0	0
4	-1	19	19
5	-1	8	8
6	-1	33	33
7	28	28	28
8	-1	7	7
9	-1	5	5
10	22	22	10
11	-1	17	17
12	-1	21	21
13	-1	11	11
14	35	35	35
15	26	26	26
16	13	13	-1
17	-1	20	-1

在实验中，场景设定后，系统会随机选择一个可被选中的物体，要求用户选中，之后每次要求用户观测距离前一个被选中物体最近的可被选中物体，通过这一方法，可以有效的模拟连续的用户视线。系统要求用户选中的物体会变成红色，用户看到这一物体后按下空格键，确认正在关注这一物体。空格键按下的同时，复合式物体检测算法和直接选中物体检测算法

同时运行，系统同时记录下两套物体检测算法所返回的物体编号和用户按空格键确认的物体编号（红色的物体）。

表 4.1 为一组测试结果的部分测试数据，左侧为眼球追踪设备自带的直接选中算法的选中数据，中间为用户真实选中的结果，右侧为复合式算法返回的选中数据。测试场景中的每一个物体都有唯一的编号，表 4.1 中选中物体编号代表上述三种情况下，每次测试中被选中物体的编号，如果编号为-1，代表算法无返回值，未选中任何物体。

两种算法返回结果与用户真实选中结果进行对比，如果选中物体的编号不同，则判定为误选，设置为红色底色；如果算法返回物体编号为-1，表示算法未选中任何物体，无返回值，则判定为选中失败，设置为黄色底色。比较算法返回的选中结果与当前用户真实选中结果的差异，计算误选和选中失败结果的占比，可以得出两种算法的准确率。

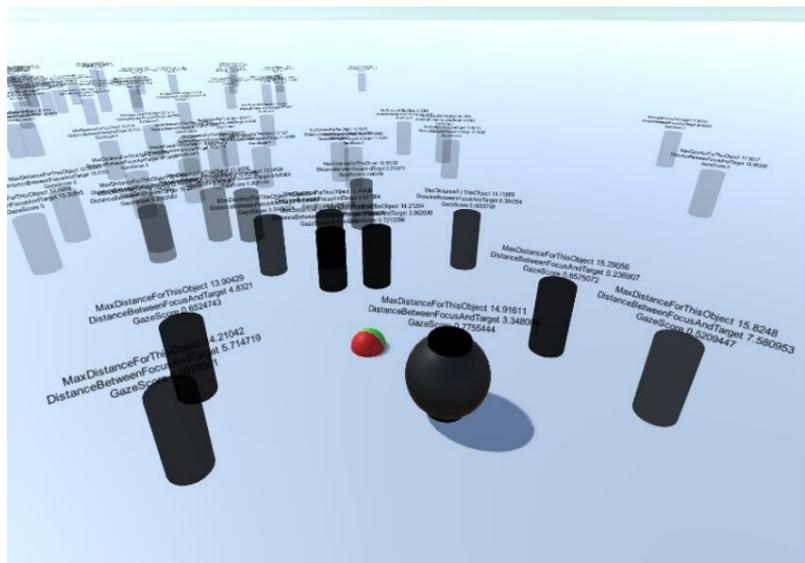
为了验证实验的一致性，对 20-50 米和 50-100 米两个场景的 12 组实验，进行了多次重复实验，发现实验的一致性非常好，本文给出了其中两次实验的结果。

4.2.3 数据可视化分析工具

为了方便实验数据的分析，设计了数据可视化分析软件。如图 4.6 所示，数据可视化分析软件基于 Unity 环境搭建，其中包含了 AOI 概率，能见度概率，预测概率，综合概率的热度图，焦点轨迹，选中物体轨迹可视化，以及实验场景逐帧复现的功能。通过数据可视化分析软件，可以很好地分析出复合式物体检测算法各部分在每一帧中对物体选中的贡献，并分析错误选中的原因，以便优化程序和算法。



(a) 功能列表



(b) 概率热度图

图 4.6: 数据可视化分析软件

4.2.4 近距离场景的验证

在实验数据的整理中, 由于距离对测试结果的影响最大, 所以按照距离对数据进行分析。本节分析 20-50 米的情况下复合式物体检测算法的准确率。



(a) 可被选中物体密集分布场景

(b) 用户视角

图 4.7 实验场景图: 距离 20-50 米, 物体总数 200 个

图 4.5, 图 4.7 为距离 20-50 米的条件下, 物体总数 200 个的实验场景及用户视角。物体总数为 30 和 100 的场景图, 由于论文篇幅的原因, 不在此展示。

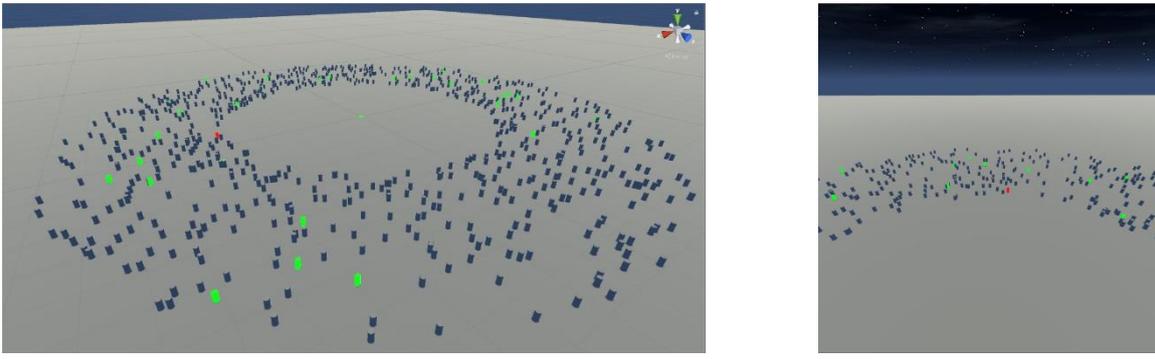
表 4.2 距离 20-50 米情况下物体检测算法的准确率。

实验场景参数			第一次实验		第二次实验	
场景编号	物体数量	可被选中物体密集程度	准确率 (%)		准确率 (%)	
			复合式	直接选中	复合式	直接选中
1	30	密集	93.3	26.7	86.7	23.3
2		稀疏	100.0	46.7	90.0	43.3
3	100	密集	76.7	23.3	76.7	26.7
4		稀疏	90.0	43.3	86.7	43.3
5	200	密集	66.7	23.3	66.7	26.7
6		稀疏	96.7	43.3	96.7	40.0

表 4.2 是采用 4.2.2 节实验方法得到的近距离场景下两种物体检测算法的准确率结果, 从实验结果可以看出, 在距离 20-50 米、可被选中物体稀疏分布的情况下, 无论障碍物数量多少, 复合式物体检测算法的准确率都能维持在 90%左右。但是, 在可被选中物体密集分布时, 复合式算法的准确率会有一定程度的下降。通过数据可视化分析软件分析发现, 在距离 20-50 米的场景下, 由于设备自身的精度问题, 会导致计算出的焦点位置在一个范围内抖动, 使得 AOI 概率和转移概率随焦点位置变化而变化, 当物体密集分布时, 需要选中的物体周围存在可被选中的障碍物 (绿色物体), 导致误选情况出现, 算法的准确率下降。

在表 4.2 实验结果中, 对比直接选中物体检测算法的准确率, 复合式物体检测算法明显有很大优势, 不管物体的数量多少, 疏密程度如何, 复合式物体检测算法的准确率都大大高于直接选中物体检测算法。

4.2.5 远距离场景的验证



(a) 可被选中物体稀疏分布场景

(b) 用户视角

图 4.8 实验场景图：距离 50-100 米，物体总数 763 个

图 4.8 (a) (b) 为距离在 50-100 米时，物体总数为 763 个时，可被选中物体稀疏分布的实验场景图，此场景设定下，可被选中物体密集分布的场景图在场景截图中看不出明显差别，故不在此展示。但是从头戴式显示器的屏幕上，可以看出场景中的疏密差别。物体总数为 108 和 362 时的实验场景图，由于论文篇幅的原因，不在此展示。

表 4.3 距离 50-100 米情况下物体检测算法的准确率

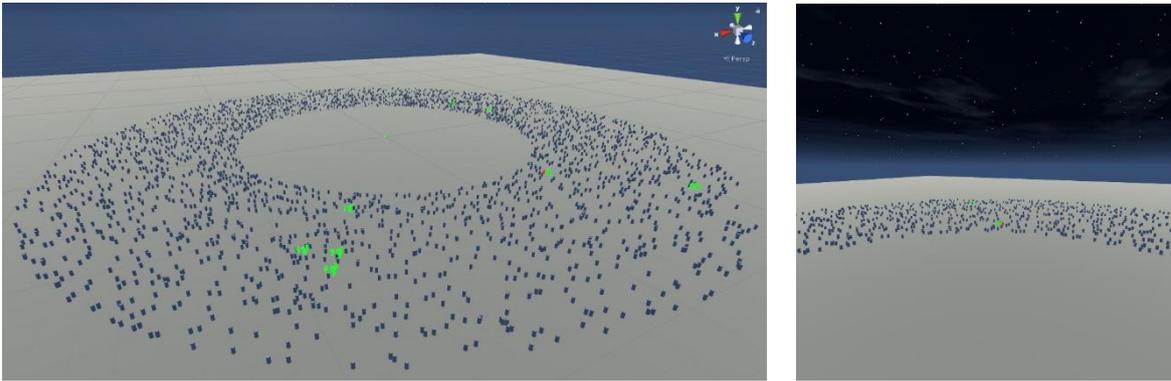
实验场景参数			第一次实验		第二次实验	
场景编号	物体数量	可被选中 物体密集程度	准确率 (%)		准确率 (%)	
			复合式	直接选中	复合式	直接选中
1	108	密集	73.3	13.3	70.0	16.7
2		稀疏	96.7	23.3	93.3	23.3
3	362	密集	70.0	13.3	66.7	13.3
4		稀疏	96.7	26.7	83.3	26.7
5	763	密集	76.7	10.0	70.0	10.0
6		稀疏	96.7	26.7	90.0	23.3

表 4.3 是采用 4.2.2 节实验方法得到的远距离场景下两种物体检测算法的准确率结果，从实验结果可以看出，复合式物体检测算法在远距离场景下的准确率，相较近距离场景下的准确率有 5%-10% 的下降。通过数据可视化分析软件分析测试数据发现，在距离较远时，用户的视线与地面的夹角较小，导致左右眼视线与地面的两个交点，在三维空间中距离相差较大，以致计算出的焦点在三维空间中的位置与用户实际关注点的位置有所偏移。这种偏移在 4.2.6 节中距离用户大于 100 米极端场景更为明显。通过数据可视化分析软件分析测试数据还发现，距离较远时设备误差被进一步放大，导致距离远且物体密集分布的场景中，需要选中的物体，被周围可被选中的障碍物（绿色物体）干扰更加严重，算法的准确率进一步下降至 70% 左右。但是，不可选中的障碍物（蓝色物体）的数量对选中准确率的影响几乎可以忽略不计，复合

式物体检测算法中可见性测试, 对被遮挡的无关物体进行了有效的剔除。

从表 4.3 实验结果中, 可以看出, 在 50-100 米的距离上, 直接选中物体检测算法准确率进一步降低, 几乎无法选中物体。复合式物体检测算法, 准确率随着距离的增加, 有所下降, 但是其总体准确率基本高于 70%。

4.2.6 极端情况场景的验证



(a) 可被选中物体密集分布场景

(b) 用户视角

图 4.9 实验场景图: 距离 100-200 米, 物体总数 3000 个

在极端情况场景的测试中, 复合式物体检测算法的准确率为 50%。通过数据可视化分析软件分析测试数据发现, 造成性能进一步下降的主要原因在于设备误差在 100-200 米距离上被进一步放大, 使得计算出的焦点位置进一步偏离用户的实际关注点, 从而更加严重影响复合式物体检测算法的准确度。在此场景中, 依赖于相交检测的直接选中物体检测算法准确率接近零, 完全无法选中物体。

4.3 示例程序对算法的验证

示例程序为一个第一人称射击游戏, 如图 4.10 所示, 该程序在 Unity 环境下搭建, 使用 Fove0 头戴式显示器作为 HMD 硬件设备, 游戏的场景为一张随机生成的无限地图, 用户仅通过眼球的运动, 便可以控制游戏中的 UI 控件、角色, 实现游戏中各种菜单的选中, 游戏人物的移动, 敌人的选中, 射击等各种动作。本文中许多测试都是在示例程序的场景中完成。

示例程序中的关键技术物体检测、选中和触发 (此处的物体定义包括虚拟现实中的菜单控件, 物体, 虚拟人物等) 使用了本文设计的复合式物体检测算法。示例程序很好的模拟了真实的虚拟现实使用环境, 复合式物体检测算法在示例程序模拟的真实的虚拟现实使用环境中, 进行测试和验证, 取得了很好的效果。



图 4.10 示例程序中的用户视角

5. 结论

根据本文设计实验场景的测试数据和虚拟现实示例程序的用户体验，以及数据可视化分析软件对测试数据的分析结果，可以得出以下结论。

1. 使用了聚焦有效性检测的复合式物体检测算法有效降低控件的误触发率，并使得触发阈值可以安全的设定为较小值，降低了控件触发失败率，提高系统的流畅度。
2. 复合式物体检测算法，相较传统的直接选中物体检测算法，在鲁棒性上有很大的提升。物体距离、疏密程度和遮挡程度等影响物体检测算法准确度的因素，对复合式物体检测算法的影响要小很多。测试数据表明，复合式物体检测算法在三维空间复杂的场景下，可以有效的提高物体选中准确率。

综上所述，所设计的复合式物体检测算法，降低了虚拟现实眼球追踪人机交互过程中误触发率，提高了三维空间复杂场景下用户关注物体的检测和选中的准确率，提升了虚拟现实眼球追踪系统精度，改善了虚拟现实眼球追踪的人机交互体验。

尽管为减小虚拟现实眼球追踪设备误差做了大量的研究，但设备自身相机的拍摄误差，瞳孔检测算法的误差对虚拟现实眼球追踪系统精度的影响还是存在，在一些场景中，计算出来的焦点位置依旧会有较大的误差，偏离用户的实际关注点。通过实验发现，这些误差是有规律的，可修正减小的。在今后的研究中可以设计合适数学模型和校准算法，对设备误差进行分析和修正，进一步提高虚拟现实眼球追踪系统精度。

参考文献

1. Nishizawa K, Kadonosono K, Itoh N, et al. Eye Tracking for Personal Visual Analytics. [J]. IEEE Computer Graphics & Applications, 2015, 35(4):64-72.
2. Alam S, Jianu R. Analyzing Eye-Tracking Information in Visualization and Data Space: from Where on the Screen to What on the Screen[J]. IEEE Transactions on Visualization and Computer Graphics, 2017, 23(5):1492-1505.
3. Stengel M, Grogorick S, Eisemann M, et al. An Affordable Solution for Binocular Eye Tracking and Calibration in Head-mounted Displays[C]. Proceedings of the 23rd ACM international conference on Multimedia, 2015. 15-24.
4. Fuhl W, Santini T, Kasneci E. Fast camera focus estimation for gaze-based focus control[J/OL]. arXiv preprint, 2017, arXiv:1711.03306.
5. He G, Oueida S, Ward T. Gazing into the Abyss: Real-time Gaze Estimation[J/OL]. arXiv preprint, 2017, arXiv:1711.06918.
6. Zhang B, Luan S, Chen C, et al. Latent Constrained Correlation Filter[J]. IEEE Transactions on Image Processing, 2018, 27(3): 1038 - 1048 .
7. Arar N M, Thiran J P. Robust Real-Time Multi-View Eye Tracking[J/OL]. arXiv preprint, 2017, arXiv:1711.05444.
8. Santini T, Fuhl W, Kasneci E. PuRe: Robust pupil detection for real-time pervasive eye tracking[J/OL]. arXiv preprint, 2017, arXiv:1712.08900.
9. Fuhl W, Santini T, Kasneci G, et al. PupilNet v2.0: Convolutional Neural Networks for CPU based real time Robust Pupil Detection[J/OL]. arXiv preprint, 2017, arXiv:1711.00112.
10. Pfeiffer T, Latoschik M E, Wachsmuth I. Evaluation of Binocular Eye Trackers and Algorithms for 3D Gaze Interaction in Virtual Reality Environments[J]. Journal of Virtual Reality & Broadcasting, 2008, 5(16).
11. Stellmach S, Nacke L, Dachselt R. 3D attentional maps: aggregated gaze visualizations in three-dimensional virtual environments[C]. Proceedings of the International Conference on Advanced Visual Interfaces. 2010. 345-348.
12. <http://www.tobiidynavox.cn/devices/>
13. <https://www.smivision.com>
14. <https://www.7invensun.com/>
15. <https://www.getfove.com/>
16. <http://www.aglass.com/>

致谢

本论文在写作过程中得到了复旦大学姜忠鼎教授和上海中学吴奕明老师的指导和关心, 在此致以诚挚的谢意。同时, 也感谢班主任许建丽老师, 张智顺老师在本论文写作过程中给予的鼓励与支持。

本参赛团队声明所提交的论文是在指导老师指导下进行的研究工作和取得的研究成果。尽本团队所知,除了文中特别加以标注和致谢中所罗列的内容以外,论文中不包含其他人已经发表或撰写过的研究成果。若有不实之处,本人愿意承担一切相关责任。

参赛队员: 梅毅轩 指导老师: 吴奕明

2018 年 9 月 11 日