

# 浙江大学

## 硕士研究生读书报告



题目 VR 人机交互技术

作者姓名 杨天泽

作者学号 21860421

授课教师 李启雷

学科专业 计算机技术

所在学院 工程师学院

提交日期 2018-12-25

Reading Report on *VR Human computer interaction technology*

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Computer Technology

Advisor: Li QiLei

By

Tao Jiaheng

Zhejiang University, P.R. China

2018

## 摘要

虚拟现实 (virtual reality, 简称 VR), 是利用计算机模拟产生一个三维空间的技术, 为用户提供视觉、听觉、触觉等感官的模拟, 让使用者可以身临其境般及时地、没有限制地观察三维空间内的事物。VR 是多种技术的综合, 包括实时三维计算机图形技术, 广角 (宽视野) 立体显示技术, 对使用者头、眼和手等的跟踪技术, 以及触觉反馈、立体声、网络传输、语音输入与输出技术等。

当前在市场上或研究室中, 常见的人机交互技术有空间定位、动作识别 (捕捉)、语音识别、眼动识别、脑电波识别等。姿态识别技术是发展得较快的领域之一, 姿态识别技术分为基于数据手套 (传感器) 和基于计算机视觉两大类。基于数据手套 (传感器) 的手势 (姿态) 识别需穿戴传感设备, 具有灵敏性 (高灵敏度, 刷新速度快) 的特点。但不足的是使用者必须依赖于设备, 并且各种平台设备之间的连接需要定制。相对而言, 基于计算机视觉识别技术对使用者没有限制, 实现的交互方式更贴近日常行为方式, 但其处理数据庞大、算法复杂、技术门槛高。

**关键词:** 虚拟现实, 人机交互, 三维, 传感。

## Abstract

Virtual reality (VR) is a technology that uses computer simulation to generate a three-dimensional space, providing users with visual, auditory, tactile and other sensory simulation, so that users can observe things in the three-dimensional space in time and without restrictions. VR is a synthesis of many technologies, including real-time three-dimensional computer graphics technology, wide-angle (wide-field) stereo display technology, tracking technology of user's head, eye and hand, tactile feedback, stereo, network transmission, voice input and output technology, etc.

At present, in the market or research room, common human-computer interaction technologies include spatial positioning, motion recognition (capture), speech recognition, eye movement recognition, EEG recognition and so on. Attitude recognition technology is one of the fast developing fields. Attitude recognition technology divided into two categories: data glove (sensor) based and computer vision based. The gesture recognition based on data gloves (sensors) needs to wear sensor equipment, which has the characteristics of sensitivity (high sensitivity, fast refresh speed). What is insufficient is that users must rely on devices, and the connection between devices on various platforms needs to customize. Comparatively speaking, Computer Vision recognition technology has no restrictions on users, and the interactive mode is closer to the daily behavior mode, but its processing data is huge, the algorithm is complex, and the technical threshold is high.

**Keywords:** Virtual reality, human-computer interaction, 3d, sensing.

# 1、引言

毫无疑问，2016 年是 VR 爆发的元年，人机交互技术也随之成为人们关注的聚焦点。当前而言，人机交互已经从以计算机为中心转移到以人为中心的发展阶段。人们开始将人机技术与虚拟环境结合应用于各行业领域，但成熟的虚拟现实技术需要三维人机交互高度自然化，人机交互技术的应用潜力巨大且已展现。

## 2、动作捕捉在 VR 中应用

在目前的消费级 VR 设备中，除了三大头显（Oculus rift、HTCvive、PS VR）外，大部分的 VR 头显都不具备配套的体感交互（需要第三方设备），而正因为缺少了体感交互，使得这些设备未能构成完善的虚拟现实体验。

支持体感交互的 VR 设备能有效降低晕动症的发生，并大大提高沉浸感，其中最关键就是可以让用户的身体与虚拟世界中的各种场景互动。在体感交互技术中又可以细分出各种类别及产品，如体感座椅、跑步机、体感衣服、空间定位技术、动作捕捉技术等

如图 1 所示：



图 1 体感交互技术

下面主要讲述目前市面上关于 VR 常见的动作捕捉及空间定位技术。

### 2.1 激光定位技术

基本原理就是在空间内安装数个可发射激光的装置，对空间发射横竖两个方向扫描的激光，在被定位的物体上放置多个激光感应接收器，通过计算两束光线到达定位物体的角度差，从而得到物体的三维坐标。物体在移动时三维坐标也会跟着变化，便得到了动作信息，完成动作的捕捉。

代表：HTC Vive - Lighthouse 定位技术。

HTC Vive 的 Lighthouse 定位技术就是靠激光和光敏传感器来确定运动物体的位置，通过在空间对角线上安装 2 个高约 2 米的“灯塔”，灯塔每秒能发出 6 次激光束，内有 2 个扫描模块，分别在水平和垂直方向轮流对空间发射激光扫描定位空间。

## 2.2 红外光学定位技术

这种技术的基本原理是通过在空间内安装多个红外发射摄像头，从而对整个空间进行覆盖拍摄，被定位的物体表面则安装了红外反光点，摄像头发出的红外光再经反光点反射，随后捕捉到这些经反射的红外光，配合多个摄像头工作再通过后续程序计算后便能得到被定位物体的空间坐标。

代表：Oculus Rift 主动式红外光学定位技术+九轴定位系统。

## 2.3 可见光定位技术

可见光定位技术的原理和红外光学定位技术有点相似，同样采用摄像头捕捉被追踪物体的位置信息，只是其不再利用红外光，而是直接利用可见光，在不同的被追踪物体上安装能发出不同颜色的发光灯，摄像头捕捉到这些颜色光点从而区分不同的被追踪物体以及位置信息。

代表：PS VR。

## 2.4 计算机视觉动作捕捉技术

这项技术基于计算机视觉原理，由多个高速相机从不同角度对运动目标进行拍摄。当目标的运动轨迹被多台摄像机获取后，通过后续程序的运算，便能在计算机中得到目标的轨迹信息，也就完成了动作的捕捉。

代表：Leap Motion 手势识别技术。

## 2.5 基于惯性传感器的动作捕捉技术

采用这种技术，被追踪目标需要在重要节点上佩戴集成加速度计、陀螺仪和磁力计等惯性传感器设备，这是一整套的动作捕捉系统，需要多个元器件协同工作，其由

惯性器件和数据处理单元组成。数据处理单元利用惯性器件采集到的运动学信息，当目标在运动时，这些元器件的位置信息被改变，从而得到目标运动的轨迹，之后再通过惯性导航原理便可完成运动目标的动作捕捉。

代表：诺亦腾 Perception Neuron。

### 3. 三维手势交互

目前在 VR/AR 的人机交互场景中，人们迫切希望看到自己的真手，并使用自己的双手和 VR/AR 中的虚拟物体进行自然地直接互动。三维手势交互首先需要骨骼的三维深度信息，可以直接使用这些信息驱动物理模型，直接和虚拟物体产生捏取、抓握等各种自然交互；也可以选取多帧三维数据，提取模式，定义为手势，达到手势识别的效果。手势识别分二维手势识别和三维手势识别。相比二维手势识别技术，三维手势识别需要使用多个摄像头，因为单个普通摄像头无法提供深度信息。双目摄像头实际上就是在模拟人眼工作的原理，只有使用两台摄像机对当前环境进行拍摄，才能得到两幅针对同一物体的不同视角照片。

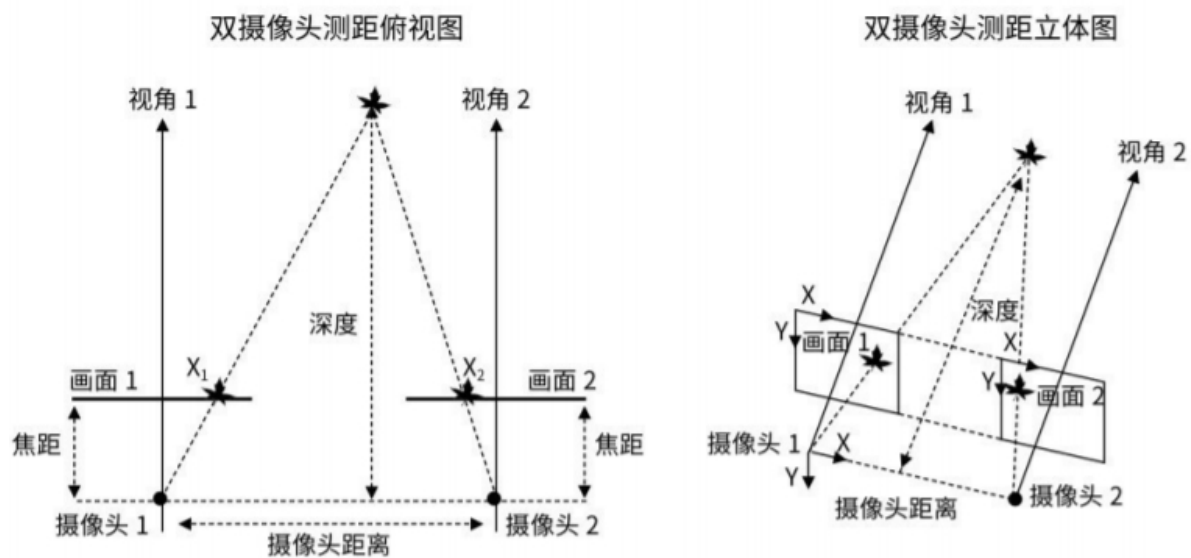


图 2 三维深度信息

例如，我们拍一片枫叶，两台摄像机的各项参数以及它们之间相对位置的关系是已知的，只要找出相同物体（枫叶）在不同画面中的位置，我们就能通过算法计算出这个物体（枫叶）距离摄像头的深度了，如图 2 所示。

如何取得手部骨骼的三维坐标，是三维交互技术的核心。uSens 凌感人机三维手势

交互技术以 3D 骨架模型为手势分析对象，通过 22 个手部关节点、26 个自由度的识别和跟踪，辨别单双手的各种手部姿态，包括点击、滑动、抓取、抚摸、翻转、握拳、鼓掌、赞、OK 手势等。

### 3.1 手势识别技术发展的三个阶段

手势识别技术是三维手势交互技术的重要组成部分，从发展过程来说，也经历了从低到高的发展阶段。

第一阶段主要是对“静态手势”的识别，从技术上来说，实现的可能性相对容易。“静态”是这种二维手势识别技术的重要特征，这种技术只能识别手势的“状态”，而不能感知手势的“持续变化”。例如，如果将这种技术用在猜拳上的话，它可以识别出石头、剪刀和布的手势状态。但是对除此之外的手势，它就一无所知了。所以，这种技术是一种模式匹配技术，通过计算机视觉算法分析图像，和预设的图像模式进行比对，从而理解这种手势的含义。这种技术的不足之处在于只可以识别预设好的状态，拓展性差、控制感弱，使用者只能实现最基础的人机交互功能。虽然看起来它处于初级阶段，但是它是识别复杂手势的第一步，也是起点。而且从应用上来说，我们的确能通过手势和计算机互动。想象一下，你一边吃饭一边用计算机看视频，但距离远也不想起身。这时只要做个手势，计算机识别后就可以切换视频，比你起身去滑动鼠标更为方便。

第二阶段是对“动态手势”的识别，从技术上来说，实现的可能性比静态稍难一些，但仍不包含深度信息提取，因为深度信息提取仍然停留在二维层面识别上。这种技术不仅可以识别手型，还可以识别一些简单的二维手势动作，如对着摄像头挥挥手。其代表公司是以色列的 PointGrab、EyeSight 和 ExtremeReality。二维手势识别拥有动态的特征，可以追踪手势的运动，进而识别将手势和手部运动结合在一起的复杂动作。这样一来，我们就把手势识别的范围真正拓展到二维平面。我们不仅可以通通过手势来控制计算机视频的播放、暂停，还可以实现前进、后退、向上翻页、向下滚动这些需要二维坐标变更信息的复杂操作。这种技术虽然在硬件要求上和二维手型识别并无区别，但是得益于更加先进的计算机视觉算法，可以获得更加丰富的人机交互内容。在使用体验上也能得到提高，纯粹的状态控制，变成比较丰富的平面控制，这种技术已经被集成到电视里。

第三阶段是实现自然手势识别——三维手势识别技术实现。三维手势识别需要的



输入包含深度的信息，可以识别各种手型、手势和动作。相比于前两种二维手势识别技术，三维手势识别不能只使用单个普通摄像头，因为单个普通摄像头无法提供深度信息。要得到深度信息需要特别的硬件，目前主要有 3 种硬件实现方式。再加上新的先进的计算机视觉软件算法就可以实现三维手势识别。

### 3.2 三维手势交互技术与 VR 结合

目前而言，VR 头显是 VR 领域发展最为迅速的产品形式。我们曾调查过很多初次体验 VR 头显的用户，当他们初次置身于一个陌生的虚拟环境中时，通常来说首先想做的动作就是用手去触摸感知。但现实却是，大部分 VR/AR 头显设备完全不具备这个功能，个别设备支持有限的手部交互功能，这就是三维手势交互技术在 VR 领域的价值所在。计算机是否能够通过各种技术实现对手的定位和识别，决定了虚拟现实技术是否能到达下一个层次的体验。通过手势来进行人机交互，符合人体的原始认知，就能创造深度沉浸感。正如上文所说，目前实现这种交互的方式主要有两种：一种是使用传感器设备，用户需要穿戴装置或手持外设；另一种是直接利用自然手势交互。前一种方式，现在主要应用于重度游戏玩家、射击竞速等大型游戏。VR/AR 要能呈现好的效果就必须有较强的沉浸感（临场感），而沉浸感较强的 VR/AR 一般具有知觉渠道多、情节统合完整、参与感强等特点。

三维手势交互技术则有助于实现超强的临场感。例如，如果我们要在 VR 中去拿一个苹果。在去拿这个苹果之前，我们就有心理和认知预设，知道苹果是圆的，我们需要把手掌张大，五个指头一起去抓。但如果我们不用手去抓取，而要用手柄去抓苹果，就与人本身的行为习惯是相悖的，与真实世界的动作逻辑不符合，我们就会感觉“跳戏”，这种“跳戏”的感觉会提醒我们，我们现在看到的 VR/AR 环境都不是真的！大大降低了 VR/AR 的沉浸感。另外，还有重要的一点就是力反馈问题。在现实生活中，我们抓到苹果后，苹果光滑的表皮、温度都会给我们的手一个力反馈，但是目前在 VR/AR 里还不能完全达到。因此，自然交互将成为 VR/AR 将要攻克的技术难题，三维手势交互技术在 VR/AR 中的应用与结合也势在必行。

## 参考文献

- [1]杰伦·拉尼尔. 虚拟现实：万象的新开端[M]. 中信出版集团, 2018.
- [2]Long L, Huai Y. Virtual Flower Visualization System Based on Somatosensory Interaction[J]. International Journal of Pattern Recognition & Artificial Intelligence, 2018(10).
- [3]苏凯, 赵苏砚. VR 虚拟现实与 AR 增强现实的技术原理与商业应用[M]. 北京: 人民邮电出版社, 2017.