

第3章 虚拟现实的关键技术和引擎

虚拟现实是一种基于可计算信息的沉浸式交互环境,具体地说,就是采用以计算机技术为核心的现代高科技生成逼真的视、听、触觉一体化的特定范围的虚拟环境,用户借助必要的设备以自然的方式与虚拟环境中的对象进行交互作用、相互影响,从而产生“沉浸”于等同真实环境的感受和体验。VR带来了人机交互的新概念、新内容、新方式和新方法,使得人机交互的内容更加丰富、形象,方式更加自然、和谐。

在虚拟现实的关键技术方面,系统强调实物虚化、虚物实化和高性能的计算机处理,这些技术是VR的3个主要方面。

1. 实物虚化

实物虚化是指通过技术手段来生成具有真实感的虚拟世界,并且在虚拟环境中对用户的操作进行检测和准确获取用户的操作数据。要实现该目标,应该包括如下基本关键技术。

- (1) 基本模型的构建技术。
- (2) 空间跟踪技术。
- (3) 声音跟踪技术。
- (4) 视觉跟踪与视点感应技术。

2. 虚物实化

确保用户在虚拟环境中获取视觉、听觉、力觉和触觉等感官认知的关键技术,是虚物实化的主要研究内容。

- (1) 视觉感知。
- (2) 听觉感知。
- (3) 力觉和触觉感知。

3. 高性能计算处理技术

- (1) 服务于实物虚化和虚物实化的数据转换和数据预处理。
- (2) 实时、逼真图形图像生成与显示技术。
- (3) 多种声音的合成与声音空间化技术。
- (4) 多维信息数据的融合、数据转换、数据压缩、数据标准化以及数据库的生成。
- (5) 模式识别,如命令识别、语音识别,以及手势和人的面部表情信息的检测、合成和识别。

- (6) 高级计算模型的研究,如专家系统、自组织神经网络、遗传算法等。
- (7) 分布式与并行计算,以及高速、大规模的远程网络技术。

下面仅对部分关键技术予以探讨。



3.1 立体高清显示技术

作为虚拟现实系统实现沉浸交互的方式之一,立体高清显示可以把图像的纵深、层次、位置全部展现,参入者可以更直观、更自然地了解图像的现实分布状况,从而更全面地了解图像或显示内容的信息。从技术方面看,需要通过光学技术构建逼真的三维环境和立体的虚拟物体对象,这就要求根据人类双眼的视觉生理特点来设计,使得人们在虚拟现实环境中看到的景观与日常生活中的场景比较时,在质量、清晰度和范围方面应该是无法区分的,从而产生身临其境的沉浸感。但就目前的技术而言,往往需要借助一些昂贵的硬件设备,如数据手套、头盔显示器及其他高档图形工作站等。

3.1.1 立体视觉的形成原理

立体视觉是人眼在观察事物时所具有的立体感。人眼对获取的景象有相当的深度感知能力(Depth Perception),而这些感知能力又源自人眼可以提取出景象中的深度要素(Depth Cue)。之所以可以具备这些能力,主要依靠人眼的如下几种机能。

- (1) 双目视差(Binocular Parallax)。
- (2) 运动视差(Motion Parallax)。
- (3) 眼睛的适应性调节(Accommodation)。
- (4) 视差图像在人脑中的融合(Convergence)。

除了以上几种机能外,人的经验和心理作用也对景象的深度感知能力有影响,如图像的颜色差异、对比度差异、景物阴影甚至是所观看显示器的尺寸和观察者所处的环境等,但这些要素相对上述机能来讲,在建立立体感上是比较小的。

当人们的双眼同时注视某物体时,双眼视线交叉于某个物体对象上,叫注视点。从注视点反射回到视网膜上的光点是对应的,但由于人的两只眼睛相距约 65mm,因此两眼观察物体对象时的角度是不一样的,从这两点返回的信号也就有了差异,再转入大脑视中枢合成一个物体完整的图像时,不但看清了该物体对象,而且该物体对象与周围物体间的距离、深度、凸凹等都能辨别出来,这样所获取的图像就是一种具有立体感的图像,这种视觉也就是人的双眼立体视觉。

实际上,人们在观察事物时,不仅是双眼看物会产生立体感,同样,用单眼看物也会产生三维效果,如果一个物体对象有一定的景深效果,单眼观察时会自动进行调节,也就是对物体的远近差异引起眼睛内的晶状体焦距及瞳孔直径的调节;如果物体是运动的,单眼会产生移动视差,因物体位置的前后不同引起移动时的差异。

总之,人类对世界万物的认知从心理到生理都留下了深深的三维轮廓,不可变更。

3.1.2 立体图像再造

人们对现实世界的观察印象是三维的,因此在虚拟现实系统中,借助现代科技对视觉



生理的认识和电子科技的发展,目前光学设备主要采用下面四种原理来重构三维环境。

1. 分光

常见的光源都会随机发出自然光和偏振光,分光技术是用偏光滤镜或偏光片滤除特定角度偏振光以外的所有光,让 0° 的偏振光只进入右眼, 90° 的偏振光只进入左眼(也可用 45° 和 135° 的偏振光搭配)。两种偏振光分别搭载着两套画面,观众须带上专用的偏光眼镜,眼镜的两片镜片由偏光滤镜或偏光片制成,分别可以让 0° 和 90° 的偏振光通过,这样就完成了第二次过滤。目前,分光技术的应用还主要停留在投影机上,早期必须使用双投影机加偏振光滤镜的方案,现在已经可以用单投影机来实现,不过都必须配合不破坏偏振光的金属投影幕才能使用。

2. 分时

分时技术是将两套画面在不同的时间播放,显示器在第一次刷新时播放左眼画面,同时用专用的眼镜遮住观看者的右眼,下一次刷新时播放右眼画面,并遮住观看者的左眼。按照上述方法将两套画面以极快的速度切换,在人眼视觉暂留特性的作用下就合成了连续的画面。目前,用于遮住左右眼的眼镜用的都是液晶板,因此也被称为液晶快门眼镜,早期还曾用过机械眼镜。

3. 色分法

分色技术是另一种3D立体成像技术,现在也比较成熟,有红蓝、红绿等多种模式,但采用的原理都是一样的。色分法会将两个不同视角上拍摄的影像分别以两种不同的颜色印制在同一幅画面中。这样视频在放映时仅凭肉眼观看就只能看到模糊的重影,而通过对应的红蓝等立体眼镜就可以看到立体效果。以红蓝眼镜为例,红色镜片下只能看到红色的影像,蓝色镜片只能看到蓝色的影像,两只眼睛看到的不同影像在大脑中重叠呈现出3D立体效果。

4. 光栅

光栅技术和前三种差别较大,它是将屏幕划分成一条条垂直方向上的栅条,栅条交错显示左眼和右眼的画面,如1、3、5...显示左眼画面,2、4、6...显示右眼画面。然后在屏幕和观众之间设一层“视差障碍”,它也是由垂直方向上的栅条组成的。对于液晶这类有背光结构的显示器来说,视差障碍也可设在背光板和液晶板之间。视差障碍的作用是阻挡视线,它遮住了两眼视线交点以外的部分,使左眼看到的栅条右眼看不到,右眼看到的栅条左眼又看不到。不过,如果观看者的位置改变,那么视差障碍位置也要随之改变。为了方便移动视差障碍,小型光栅显示器都是采用液晶板来作为视差障碍的。而检测观看者位置的方法主要有两种,一种是在观看者头上戴一个定位设备,另一种是用两个摄像头像人眼一样地定位。

光栅式自由立体显示器主要是由平板显示屏和光栅精密组合而成,左右眼视差图像按一定规律排列并显示在平板显示屏上,然后利用光栅的分光作用将左右眼视差图像的光线向不同方向传播,观看者位于合适的观看区域时其左右眼分别观看到左右眼视差图像,经过大脑融合便可观看到有立体感的图像。根据采用的光栅类型可分为狭缝光栅式自由立体显示和柱透镜光栅式自由立体显示两类。狭缝光栅式自由立体显示器又分为前置狭缝光栅和后置狭缝光栅两种,其结构与原理图分别如图3.1(a)和图3.1(b)所示。

柱透镜光栅式自由立体显示器的结构与原理如图3.2所示,利用柱透镜阵列对光线



的折射作用,将左右眼视差图像分别提供给观看者的左右眼,经过大脑融合后产生具有纵深感的立体图像。

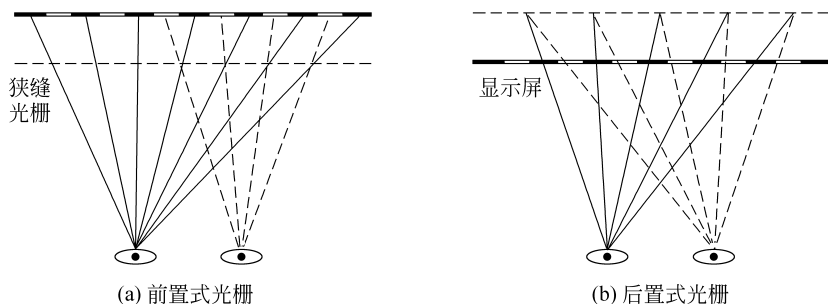


图 3.1 光栅式自由立体显示器原理

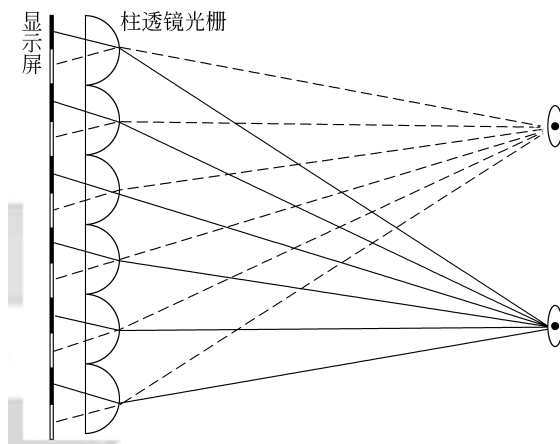


图 3.2 柱透镜光栅自由立体显示器原理

3.1.3 其他新型立体显示技术

除了上述传统的立体显示技术外,目前还有一些新设备或新技术模式也能够很好地展现三维环境效果。

1. 全息技术

全息技术的发展最早在 1947 年,英国的匈牙利裔物理学家丹尼斯·盖伯发明了全息投影术并获得了 1971 年的诺贝尔物理学奖。第一张实际记录了二维物体的光学全息投影照片是在 1962 年由苏联科学家尤里·丹尼苏克拍摄的。1969 年,本顿发明了彩虹全息术,20 世纪 60 年代末期,古德曼和劳伦斯等人提出了新的全息概念——数字全息技术,开创了精确全息技术的时代。2001 年,德国国家实验室首创研发了全息膜技术,而魔幻效果的技术则由丹麦公司 ViZoo 在 2006 年研发出来。2008 年,美国亚利桑那州大学打造了展现物体对象的可更新的 3D 全息显示屏,这是世界上首批 3D 全息显示屏之一。他们用全息膜搭建了一个倒金字塔形的几何模型,利用四台投影机投射的视频图像,在金字塔里经过一系列的光学衍射后汇合成了一幅闪亮的全息图像,该 3D 图像的效果就



像实物飘浮在空中一样。如图 3.3 所示,至此全息投影技术开始逐步走向成熟。

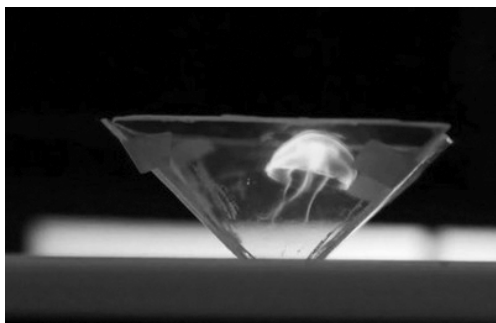


图 3.3 全息 3D 图像图

全息技术的理论源于光的物理学基础,光的物理属性具有波粒二重性,既有波的属性,也有粒子的特点。当光照射在某种介质的物体表面时,光会产生反射、折射和透射,这是光的粒子属性表现;而当两束光发生相干叠加效应时,会产生干涉和衍射的情况,这是光的波属性的表现特征。

全息投影技术就是利用了光波的干涉和衍射原理记录并再现了物体真实的三维图像的技术。

(1) 利用干涉原理记录物体光波信息。该过程也即拍摄过程。被摄物体在激光辐照下形成漫射式的物光束,另一部分激光作为参考光束射到全息底片上,和物光束叠加产生干涉,把物体光波上各点的位相和振幅转换成在空间上变化的强度,从而利用干涉条纹间的反差和间隔将物体光波的全部信息记录下来。记录着干涉条纹的底片经过显影、定影等处理程序后,便成为一张诺利德全息图,或称全息照片。

(2) 利用衍射原理再现物体光波信息。该过程即为成像过程。全息图犹如一个复杂的光栅,在相干激光照射下,一张线性记录的正弦形全息图的衍射光波一般可给出两个像,即原始像(又称初始像)和共轭像。再现的图像立体感强,具有真实的视觉效应。全息图的每一部分都记录了物体上各点的光信息,故原则上它的每一部分都能再现原物的整个图像,通过多次曝光还可以在同一张底片上记录多个不同的图像,而且能互不干扰地分别显示出来。全息投影技术原理如图 3.4 所示。

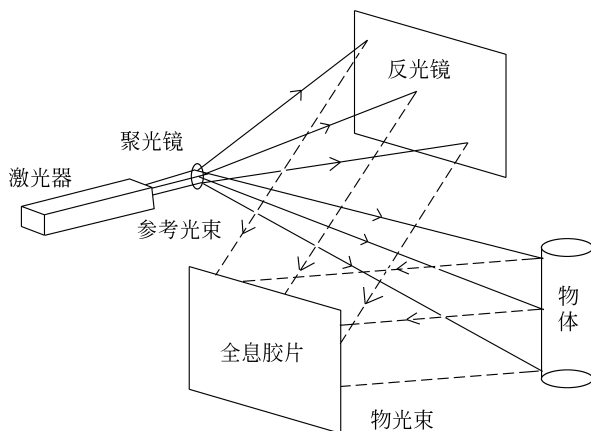


图 3.4 全息投影技术原理图



近二十年来,高分辨率 CCD 电荷耦合器件的性能快速提高,运用 CCD 器件代替传统的胶片存储影像信息,用计算机模拟取代光学衍射来实现物体再现,实现了全息图的记录、存储、处理和再现全过程的数字化,同时也极大地改善了全息投影的技术性能,特别是全息投影技术经过了最近一段时间的快速发展。全息投影按技术特征可以分为透射式、反射式、像面式、彩虹式、合成式、模压式、运算式等七大类。

① 透射式全息显示图像属于一种应用最多、形式最基本的全息显示图像。透射式全息显示图像清晰逼真,景深较大,观看效果较好。但为确保光的相干性,需用激光记录与再现。采用激光也会带来其特有的散斑效应的弊病,即再现像面上附有微小而随机分布的颗粒状结构。

② 为克服透射式全息显示图像无法利用普通白光(非相干光)再现的缺陷,人们又发展了反射式全息显示图像。反射式全息显示图像便可用普通白光扩展光源再现。这是其一大优点,同时也消除了激光的散斑效应。缺点是景深不太大,不适合制作屏幕较大的反射式全息显示图像。

③ 人们采用将物体通过透镜成像于全息板的附近,同时引入参考光波与其干涉的办法来记录全息显示图像,这样记录的全息显示图像称为像面全息显示图像,它可用普通白光扩展光源再现。同样,该全息显示图像模式的景深也是有限的,距全息板平面愈远的像点愈模糊不清。

④ 20 世纪 70 年代末,一种新型全息显示图像即彩虹式全息显示图像(Rainbow Hologram)问世,它可采用白光再现,图像清晰明亮,尤其适用于立体三维显示,倍受人们的重视。它的一大特点是加入了一个狭缝,能够限制光波,以免光波再次出现的时候因重复导致的图像显示不清楚。由于缝隙能显示不同的颜色,不同的人从不同的角度选择一个颜色去观察的时候得到的结果就会不一样。所以,当某个人逐步选择用不同的颜色去观察的时候,就会发现宛如彩虹般美丽的颜色在变化。这也就是彩虹式影像的由来。

⑤ 合成式全息显示图像是指将一系列由普通拍摄物体的二维底片借助全息方法记录在一块全息软片(或干板)上,再现时实现原物体的准立体三维显示的一种技术。它可制成圆筒式,也可制成平面式。利用该方法在平面全息板上再现环视或立体活动图像具有极大的优势。

⑥ 前述各种全息显示图像的共同缺陷是复制较为烦琐,通常需采用激光源及光学器件,还皆需曝光、显影和定影等过程。20 世纪 80 年代开发出一种可像印书一样大批量快速复制的模压式全息显示图像,仅需三步即可完成操作:记录原版全息显示图像,制作金属压模,压印复制。

⑦ 运算式又称计算机全息(Computer Generated Hologram,CGH)。由于全息显示图像属于一种光学干涉图像,于是人们设想可以利用计算机直接产生出这种图像模式,则无须再采用光学设备实地记录了。这种方法既可完全节省光源及要求相当精密的光路设置,又能模拟实际上并不存在的各种物体,故具有明显的简易性与灵活性。目前已在多领域获得了较好的应用。

3D 全息投影技术的创新效果在于它改变了人们对那些传统展示艺术的表现模式,对于未来的全息电影、全息动漫、全息计算机游戏及科技探索都具有划时代的促进意义。



2. 体显示

体显示技术又称为(真)立体显示技术,是近年来新兴的具有真实物理景深感的三维显示技术。与计算机平板显示技术不同,该显示系统的特殊性表现为体显示系统仿佛能够直接在三维空间中产生图像,这一新颖的计算机显示系统已经引起人们的关注,它在科学分析、科学计算可视化以及国防建设、生物医学工程等领域都具有十分广泛的应用前景。

体显示技术的特点是希望能通过一个 3D 显示器来直接显示三维图像,从而使得表现出的三维物体既有心理景深,更有物理景深,而且多个观察者不需要任何辅助设备(如偏光眼镜),就可以从多个角度直接观察三维物体,就像人们在观赏一个金鱼缸里的金鱼一样。

真立体显示技术可分为用于动态物体的体扫描技术和静态体成像技术。体扫描技术中屏幕的运动方式又可分为平移运动和旋转运动。

以旋转体扫描显示系统为例,其结构主要包括显示单元和图像引擎两部分。

(1) 旋转体扫描显示系统的显示单元。

任何立体显示单元都包括三个主要的子系统:成像空间产生子系统、体素生成子系统、体素激活子系统。

构造成像空间就是产生一个透明的、可编址的立体空间。在这个立体空间里,体素被激活生成图像。成像空间的构造方式对图像质量会有较大的影响。

一般情况下,真立体显示系统采用模块化设计。系统的主要组成部分如下。

- ① 显示体(包括投影面)。
- ② 投影单元。
- ③ 投影单元驱动和电动机能量供应。
- ④ 有 3D 接口的控制 PC。

真立体显示系统原理结构图如图 3.5 所示。

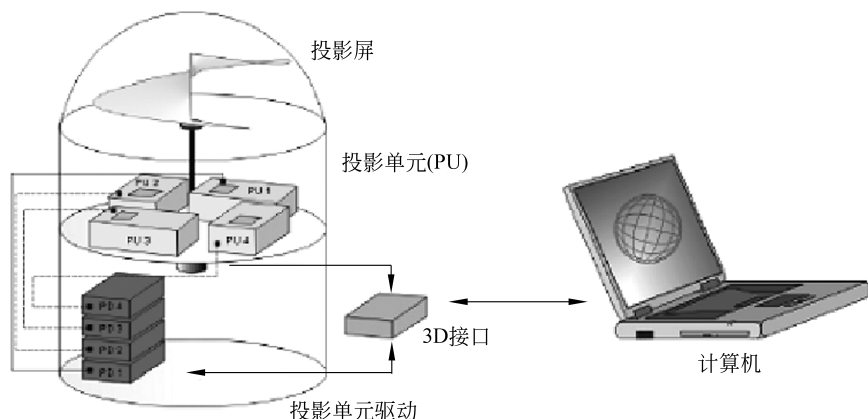


图 3.5 旋转体扫描的系统结构图

(2) 旋转体扫描显示系统的图像引擎

图像引擎是十分关键的部分,在系统的模块化设计中,有 3D 接口的控制 PC,主要是指图像引擎的设计。图像引擎的功能就是将原始空间的图像数据转换成既符合显示单



元的几何特性,又按照体扫描显示要求的顺序排列的体素说明符,然后输入显示单元,测量和校准显示单元的信息。

图像引擎主要组成部分如图 3.6 所示。

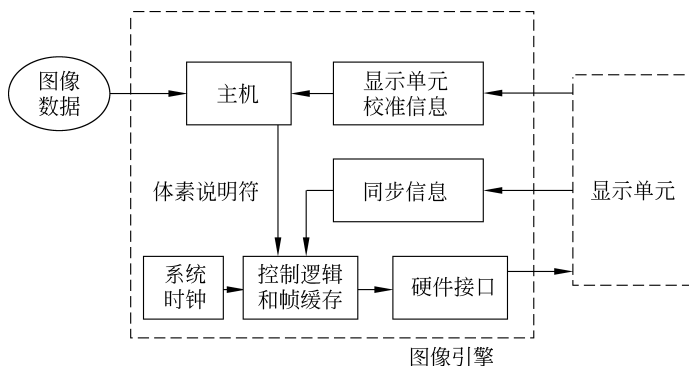


图 3.6 体扫描显示系统的图像引擎模块图

上述体显示方法可供多个观看者同时从不同角度观看同一立体场景,且兼顾了人眼的调节和会聚特性,不会引起视觉疲劳。

3.2 三维建模技术

虚拟环境的建立是虚拟现实技术的核心内容。在三维模型的建立过程中,人们不仅要求模型的几何外观逼真可信,部分对象还需要具有较为复杂的物理属性和良好的交互功能。此外,VR 系统对实时性的要求较高,而场景中的模型数据和类型又通常较多,因此对模型数据的简化和优化技术也极为重要。目前,存在多种较为成熟的建模技术,但由于各种应用领域都有其特殊性,因此 VR 建模系统并无完全统一的规范。通常,建模技术可分为几何建模、物理建模和运动建模。几何建模是基于物体的几何和形状等信息的表示,研究图形数据结构等问题;物理建模是给一定几何形状的物体对象赋予特定的物理属性;运动建模用于处理对物体对象的运动和行为的描述,通常称之为动画。

3.2.1 几何建模

几何建模是指一种技术,它可将物体的形状存储在计算机内,形成该物体的三维几何模型,并能为各种具体对象应用提供信息,如能随时在任意方向显示物体形状,计算体积、面积、重心、惯性矩等。这个模型是对原物体的确切的数学描述或是对原物体某种状态的真实模拟。然而,现实世界中的物体是复杂多样的,不可能用某一种方法就能描述各种不同特征的所有物体。为了产生景物的真实感显示,需要使用能精确地建立物体特征的表示,如使用多边形和二次曲面能够为诸如多面体和椭圆体等简单欧氏物体提供精确描述;样条曲面可用于设计机翼、齿轮及其他有曲面的机械结构;特征方程的表示方法,如分形几何和微粒系统,可以给出诸如树、花、草、云、水、火等自然景物的精确表示。



目前,在计算机内部,表示三维形体数据结构有3种存储模式,同时也就决定了形体的3种表达模型:线框模型、表面模型和实体模型。

1. 线框模型

三维线框模型是在二维线框模型的基础上发展起来的。线框模型采用顶点表和边表两个表的数据结构来表示三维物体,顶点表记录各顶点的坐标值,边表记录每条边所连接的两个顶点。由此可见,三维物体可以用它的全部顶点及边的集合来描述,“线框”一词由此而来。线框模型的优点主要是可以产生任意视图,视图间能保持正确的投影关系。线框模型的缺点也很明显,物体的真实形状须由人脑的解释才能理解,因此容易出现二义性。

2. 表面模型

表面模型通常用于构造复杂的曲面物体,构形时常常利用线框功能,先构造一个线框图,然后用扫描或旋转等手段变成曲面,当然也可以用系统提供的许多曲面图素来建立各种曲面模型。与线框模型相比,数据结构方面多了一个面表。记录了边、面间的拓扑关系,但仍旧缺乏面、体间的拓扑关系,无法区别面的哪一侧是体内,哪一侧是体外,依然不如实体模型那么直观。

3. 实体模型

实体模型与表面模型的不同之处在于确定了表面的哪一侧存在实体这个问题。实体模型的数据结构当然比较复杂,可能会有许多不同的结构。但有一点是肯定的,即数据结构不仅记录了全部几何信息,而且记录了全部点、线、面、体的拓扑信息,这是实体模型与线框或表面模型的根本区别。

4. 几何建模的常用方法

主要有两种方法:①通过人工的几何建模方法;②采用更便捷的自动的几何建模方法。

(1) 人工的几何建模方法。

① 利用相关程序语言来进行建模,如 OpenGL、Java3D、VRML 等。这类方法主要针对虚拟现实技术的特点而编写,编程相对容易,效率较高。

② 直接从某些商品图形库中选购所需的几何图形,这样可以避免直接用多边形或三角形拼构某个对象外形时烦琐的过程,也可节省大量的时间。

③ 利用常用建模软件来进行建模,如 AutoCAD、3ds Max、SoftImage、Pro/E 等。用户可交互式地创建某个对象的几何图形。这类软件的一个问题是并非完全为虚拟现实技术所设计,由 AutoCAD 或其他工具软件所产生的文件取出三维几何并不困难,但问题是并非所有要求的数据都以虚拟现实要求的形式提供,实际使用时必须要通过相关程序或手工导入。

④ 自开发的工具软件。尽管有大量的通用工具软件可供选择使用,但可能由于建模速度缓慢、周期较长、用户接口不便、不灵活等方面的原因,使得建模成为一项比较繁重的工作。多数实验室和商业动画公司宁愿使用自开发的建模工具软件,或在某些情况下用自开发的建模工具与市场销售的建模工具软件相结合的方法来解决问题。

(2) 自动的几何建模方法。

自动建模的方法有很多,最典型的是采用三维扫描仪对实际物体进行三维建模。它



能快速、方便地将真实世界的立体彩色物体信息转换为计算机能直接处理的数字信号,而不需要进行复杂、费时的建模工作。

除此之外,在虚拟现实技术中,还可采用基于数字照片的建模技术。该方法是借助数码相机,直接对需要建模的物体对象进行多个不同角度的拍摄,得到有关物体对象各个角度的照片后,采用照片建模软件进行建模。现在技术比较成熟的照片建模软件有 REALVIZ 公司的 ImageModeler、Discreet 公司的 Plasma 等。

建模时,至少需要对建模对象环绕拍摄三张以上的照片,根据透视学和摄影测量学原理,标志和定位对象上的关键控制点,建立三维网格模型。与大型 3D 扫描仪相比,这类软件有很大的优势,使用简单、节省人力、成本低、速度快,但实际建模效果一般,常用于大场景中建筑物的建模。

几何模型的表示方法是计算机图形学的基础理论,但对于虚拟现实系统而言,主要是借助于这些基础理论来研究如何更快、更好地开发几何建模对象,不论是通过图形软件进行人工建模,还是利用一些成熟的硬件设备,例如 3D 扫描仪等。需要注意的是,这些软件和硬件都有自己特定的文件格式,在导入虚拟现实系统时需要做适当的文件格式转换。

3.2.2 物理建模

虚拟现实系统中的模型不是静止的,而是具有一定的运动方式。当与用户发生交互时,也会有一定的响应方式。这些运动方式和响应方式必须遵循自然界中的物理规律,例如,刚体之间的碰撞反弹、物体的自由落体、物体受到用户外力时会朝预期方向移动等。又如,实体物对象不能相互穿插通过、软体物质对象遇到硬体物体对象时会被压缩、布料物体移动时会有飘逸的感觉。上述这些内容就是物理建模技术需要解决的问题:如何描述虚拟场景中的物理规律以及几何模型的物理属性。物理建模技术需要重点解决如下问题。

1. 设计数学模型

数学模型是指描述虚拟对象行为和运动的一组参数方程,它用来建立虚拟对象的视觉属性(如大小、形状、颜色等)、物理属性(如质量、硬度等)和物理规则(如引力、阻力等)。建立数学模型往往并不困难,但设计引入这些行为的接口程序,使物理属性和行为与几何数据库联系起来却比较复杂。

2. 创建物理效果

对虚拟对象创建物理效果的方法是从几何模型出发,将时间、长度、质量和力等过程抽象处理后,与图形学中的元素,如帧、绝对坐标、节点和面等结合起来,搭建出一个表现基本物理量的三维场景。具体来说,首先确定物理过程,即作用在虚拟对象上的物理现象,接着利用软件仿真算法描述上述物理过程,最后通过计算机程序语言实现上述仿真算法,由此表达出模型质量、密度等物理属性和力的概念。

3. 实时碰撞检测

精确的碰撞检测对提高虚拟环境的真实性、增强虚拟环境的沉浸感有着至关重要的作用。碰撞检测技术不仅要能随时检测出虚拟场景中是否有碰撞发生,还要检测出碰撞



发生的位置、时间,以及根据数学模型和物理属性计算出碰撞发生后的不同反应。因而对于碰撞检测系统来说,其技术难度要求很高。由于虚拟现实系统中的碰撞检测通常都是三维虚拟环境中发生的,其自身的复杂性和实时性又对碰撞检测提出了更高的要求,因此碰撞检测始终是物理运动中的一个关键问题。现阶段碰撞检测主要有三种方式,分别是静态碰撞检测、伪动态碰撞检测和动态碰撞检测。静态碰撞检测是判断活动对象在某一特定的位置和方向是否与环境对象相交;在静态碰撞检测的应用中,一般没有实时性的要求,因此,在计算几何中应用比较广泛。伪动态碰撞检测则是根据物体活动对象的运动路径检测它是否在某一离散的采样位置方向上与环境对象相交;因此,对于伪动态碰撞检测中关于时间点和运动参数之间的信息,可以通过开发时空相关性来获得较好的性能。动态碰撞检测则是检测活动对象扫过的空间区域是否与环境对象相交;动态碰撞检测的研究通常考虑到四维时空或结构空间精确的建模问题,因此该方法计算量相对较大。目前,较成熟的碰撞检测算法有层次包围盒法和空间分解法等。

层次包围盒法的基本思想是利用体积略大而几何特性简单的包围盒将复杂几何对象包裹起来。在进行碰撞检测时,首先进行包围盒之间的相交测试,只有包围盒相交时,才对其所包裹的对象做进一步求交计算。在构造碰撞体的包围盒时,若引入树状层次结构,可快速剔除不发生碰撞的元素,减少大量不必要的相交测试,从而提高碰撞检测效率。比较典型的包围盒类型有沿坐标轴的包围盒、包围球、方向包围盒、固定方向凸包等。层次包围盒方法应用得较为广泛,适用于复杂环境中的碰撞检测。

空间分解法是将整个虚拟空间划分成相等体积的规则单元格,只对占据同一单元格或相邻单元格的几何对象进行相交测试。比较典型的方法有 K-D 树、八叉树、BSP 树、四面体网、规则网等。

空间分解法通常适用于稀疏的环境中分布比较均匀的几何对象间的碰撞。传统的八叉树有空间非均匀网格剖分算法和层级边界盒算法。传统算法适合于静态场景,对于动态场景,采用较多的是基于面向对象的动态八叉树结构,它是对原算法的改进。动态八叉树的构造和碰撞检测策略是将场景表示为等体积的规则单元格的组合。

BSP 树包含的是平面的层级,其每一个平面都将一个区域的空间分割成两个子空间。BSP 的碰撞检测策略为:在两个对象间找出分割的平面以确定两个对象是否相交;若存在分割平面则无碰撞发生。当有相交时再与包围盒中对象的多边形进行精确检测。

3.2.3 运动建模

在虚拟现实环境中,除了要观察一个对象的 3D 几何形状,还必须考虑该对象的具体位置,并以此位置为基点,进行平移、碰撞、旋转和缩放等变化。这些内容的数据建模描述表达了对对象的运动属性,所以称为运动建模或者行为建模。

几何建模与物理建模结合,可以部分实现虚拟对象“看起来真实,动起来也真实”的特征和效果,但要真正构造一个能够逼真表现虚拟世界的运动环境,必须采用更加有效的行为描述方法,才能客观、自然地模拟虚拟对象的本质特征。

运动建模的目的就是要赋予虚拟对象仿真的行为与自然的反应能力,并服从客观世界的运动规律。例如,当一个虚拟对象被抛射出去后,它将沿着一个抛物线自然回落到



地面。在对运动建模的数据描述中,与以下四个要素相关。

1. 对象的物理位置

在虚拟现实的运动建模过程中,物体对象的位置是需要首先关注的内容,通常以三维坐标系来表示对象的物理空间位置。当物体对象运动时,一般可依据计算机图形学的几何变换理论进行计算,首先物体对象按照其几何图形,可获得该图形顶点坐标的集合矩阵,再将该矩阵转变为相应的规范化的齐次坐标矩阵,然后与特定的变换矩阵相乘,即可完成物体对象的几何图形的平移计算。

例如某三维物体对象 $P(X, Y, Z)$ 的坐标位置平移到了 $P'(X', Y', Z')$, 则 P 的规

范化齐次坐标矩阵为:
$$\begin{pmatrix} x_0 & y_0 & z_0 & 1 \\ x_1 & y_1 & z_1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & z_n & 1 \end{pmatrix}$$
, 三维平移变换矩阵为:
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ T_x & T_y & T_z & 1 \end{pmatrix},$$

式中 T_x, T_y, T_z 为三个坐标轴上的平移参数。

平移计算可以通过下面的矩阵相乘完成,如图 3.7 所示。

$$\begin{pmatrix} x'_0 & y'_0 & z'_0 & 1 \\ x'_1 & y'_1 & z'_1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x'_n & y'_n & z'_n & 1 \end{pmatrix} = \begin{pmatrix} x_0 & y_0 & z_0 & 1 \\ x_1 & y_1 & z_1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & z_n & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ T_x & T_y & T_z & 1 \end{pmatrix}$$

图 3.7 三维几何图形的平移计算

如果反过来,物体对象从 $P'(X', Y', Z')$ 点位置反向平移到了 $P(X, Y, Z)$ 的坐标位置,则三维平移变换矩阵为:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -T_x & -T_y & -T_z & 1 \end{pmatrix}$$

同样,计算公式如图 3.8 所示。

$$\begin{pmatrix} x_0 & y_0 & z_0 & 1 \\ x_1 & y_1 & z_1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & z_n & 1 \end{pmatrix} = \begin{pmatrix} x'_0 & y'_0 & z'_0 & 1 \\ x'_1 & y'_1 & z'_1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x'_n & y'_n & z'_n & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -T_x & -T_y & -T_z & 1 \end{pmatrix}$$

图 3.8 三维几何图形的反向平移计算

确定了坐标系与物体对象的相对位置,就可以通过运算或矩阵变换,获得物体对象的运动效果。

2. 对象的层次

对象的层次定义了一个作为一个整体一起运动的一组对象,各部分也可以独立运行。假设不考虑对象层次,就会出现对象在运动时只能是整体运动。例如,有一个虚拟手,没有层次划分,手的指头就不能单独运动,而为了实现对手指的独立运动,就必须对手的三维模型进行分段设计,并进行分层控制。

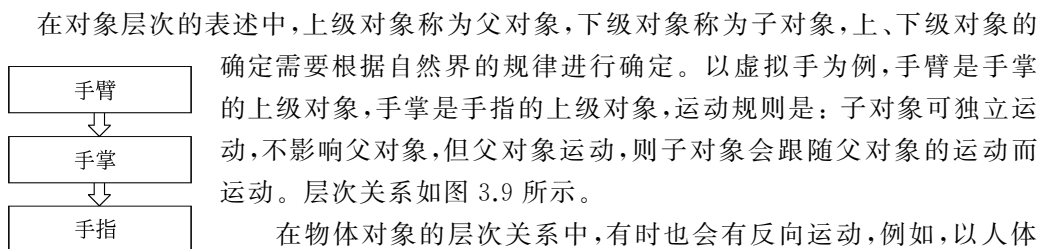


图 3.9 虚拟手的层次结构

在对象层次的表述中,上级对象称为父对象,下级对象称为子对象,上、下级对象的确定需要根据自然界的规律进行确定。以虚拟手为例,手臂是手掌的上级对象,手掌是手指的上级对象,运动规则是:子对象可独立运动,不影响父对象,但父对象运动,则子对象会跟随父对象的运动而运动。层次关系如图 3.9 所示。

在物体对象的层次关系中,有时也会有反向运动,例如,以人体为例,人的身子是上层对象,四肢、头是身子的下层对象,当身体往前运动时,下层对象必须跟随移动,下层对象如头、四肢均可独立运动。但该人体做拉单杠运动时,身子就要跟随手的运动而运动,这就是反向运动,要描述运动物体对象,分清物体对象的层次关系是必需的。

3. 虚拟摄像机

三维世界通常采用摄像机的坐标系来观察,摄像机坐标系在固定的世界坐标系中的位置和方向称为观察变换。即在观察虚拟对象时,应该通过摄像机窗口来观察对象,所以在实时绘制图形时,需要根据摄像机的坐标系来绘制,并且只能是摄像机能够看到的那部分对象,可视窗口外的部分将被裁剪。

一般情况下,为了优化处理结果,实时图形绘制时,要把视窗口规范化,这样有利于图形显示时的坐标规范化。同时对 Z 轴缓冲区的处理过程中,可依照 Z 数据值的大小进行判别, Z 值大的物体对象距观察者远,反之则近。如果 Z 值大的对象被 Z 值小的对象遮拦,则遮拦的部分可以不必绘制。

4. 人体的运动结构分析

人体是最为复杂的建模对象,也是虚拟现实中最特殊的对象,人体由骨、骨骼连接,其受力和运动均与骨骼的平衡有关。因此,人体的骨骼是构成各种动态姿势的基础。人体的骨骼系统在结构和平衡上是非常复杂和巧妙的,它能做出各种各样的动作。人体的骨骼除了维系肌肉之外,还起到保护内脏的作用。骨骼的形状多种多样,有长有短、有圆有扁,所以能适应许多特殊的动作。当人们观察一个人的运动效果时,需要准确刻画出各个骨骼关节的变化状态,没有骨骼关节的活动,就不能产生动作。人物动画的表现是连贯的、有周期性变化的运动形象,或者说就是表现姿势不断变化、重心不断移动的状态。

3.2.4 关于六自由度

在虚拟现实技术的发展过程中,许多的人机交互设备都具有六自由度的运动功能。自由度(Degree of Freedom, DoF)作为一种面向机械运动属性的评价标准,对于衡量人机交互设备的运动姿态性能具有重要的意义。一个拥有六自由度的人机交互设备表明该设备在虚拟空间里,它拥有沿其 X 、 Y 、 Z 三个直角坐标轴进行平移和环绕 X 、 Y 、 Z 三个坐标轴进行旋转的自由度,如图 3.10 所示。需要说明

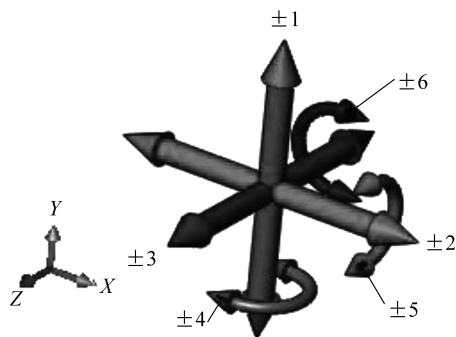


图 3.10 六自由度运动方向图



的是,六自由度的存在,也就表明在虚拟环境空间里,人机交互设备的运动轨迹或姿态至少采用六维坐标($X, Y, Z, \alpha, \beta, \gamma$)进行标定,如果要减少某物体对象运动形式的自由度,可以通过添加一定的约束来消除其中的部分自由度,当某物体对象的自由度为零时,那么该物体对象就完全处于静止的位置。

1. 六自由度的概念

在理论力学体系里,自由度指的是力学系统中的独立坐标的个数。力学系统由一组坐标来描述。例如,一个质点在三维空间中的运动,在笛卡尔坐标系中,由 X, Y, Z 三个坐标来描述;或者在球坐标系中,由 A, B, C 三个坐标描述。一般而言, N 个质点组成的力学系统由 $3N$ 个坐标参数来描述。但力学系统中常常存在着各种约束,使得这 $3N$ 个坐标并不都是独立的。对于 N 个质点组成的力学系统,若存在 M 个完整约束,则系统的自由度运算式为 $S=3N-M$ 。

例如,运动于平面的一个质点,其自由度为 2。而在三维空间中的两个质点,中间以定长直线连接。那么其自由度为 $S=3 \times 2 - 1 = 5$ 。该例说明在三维空间中的无约束两个质点应该是 3×2 , 共有 6 个自由度,但加入一条定长直线将两个质点连接起来,等同于给两个质点添加了一个约束,于是就要减去 1 个自由度为 5 个。物体的运动形式除了平移自由度外,现实中还有旋转自由度和振动自由度的运动姿势,如果要完全确定一个物体在空间位置所需要的独立坐标的数目,叫作这个物体的自由度。力学系统定义由一组坐标来描述。

2. 六自由度的跟踪系统

目前在虚拟现实的人机交互设备中,通常需要进行六自由度的跟踪定位,常用的跟踪技术有光学跟踪、声学跟踪、机械跟踪、惯性位置跟踪和电磁跟踪。在普通的应用中,光学跟踪应用最为广泛,是一种非接触式的位置计算设备,基于三角测量。缺点是会受到视线阻挡的限制。声学跟踪技术的原理就是超声测距。缺点就是会受到声波脉冲的干扰,而且和光学系统一样在系统中不能有障碍物。机械跟踪器的工作原理是通过机械臂上的参考点与被测物体相接触的方法来检测位置变化。对于一个六自由度的跟踪器,机械臂必须有 6 个独立的机械连接部件,分别对应 6 个自由度,可将任何一种复杂的运动用几个简单的平动和转动组合表示。缺点是比较笨重,不灵活而且有惯性。惯性位置跟踪由定向陀螺和加速度计组成。通过计算得出被跟踪物体的姿势,即采用通过运动系统内部的推算,而不依靠外部环境参数得到所需信息。电磁跟踪系统由磁场发射器和接收器组成。优点就是不受视线阻挡的限制,除了导体或导磁体外没有其他物体能够遮挡住电磁跟踪系统的跟踪。基于上述不同跟踪系统的优势,电磁跟踪系统更多地应用于六自由度的人机交互设备当中。

以电磁跟踪系统为例,其六自由度的跟踪算法可分四步:电磁跟踪系统中方位坐标的标定;电磁跟踪系统中的磁场感应参数的采集;电磁跟踪系统中方位坐标的矩阵变换;方位数据的求解。

3. 六自由度的应用

六自由度的机件设备由于具有多变的运动姿势,因而被广泛应用到社会发展的各个领域,如飞行模拟器、舰艇模拟器、海军直升机起降模拟平台、坦克模拟器、汽车驾驶模拟器、火车驾驶模拟器、地震模拟器以及动感电影、娱乐设备等训练、教育及科研部门,甚至还用到了空间宇宙飞船的对接,空中加油机的加油训练;以及在加工制造业可制成六轴



联动机床、灵巧机器人等。利用六自由度概念设计的运动测试平台在制造过程中涉及机械、液压、电气、控制、计算机、传感器、空间运动数学模型、实时信号传输处理、图形显示、动态仿真等一系列高科技领域,因而六自由度运动平台的研制变成了各高等院校、科学院所在液压传动和自动控制领域水平的标志性象征,也被科技人员视作传动及控制技术领域的皇冠级产品。

3.3 三维虚拟声音技术

三维虚拟声音与人们熟悉的立体声音有所不同。立体声虽然有左右声道之分,但就整体效果而言,立体声来自听者面前的某个平面。而三维虚拟声音则来自围绕听者双耳的一个球形中的任何地方,即声音出现在头的上方、后方或者前方。NASA 研究人员通过实验研究证明了三维虚拟声音与立体声的不同感受。他们让实验者戴上立体声耳机,如果采用通用的立体声技术制作声音信息,实验者会感觉到声音在头内回响,而不是来自外界。但如果设法改变声音的混响时间差和混响压力差,实验者就会明显地感觉到声源位置在变化,并开始有了沉浸感,这就是三维虚拟声音。总之,VR 里的声音系统需要满足如下需求。

(1) 3D 定位: 精确地定位虚拟声源。

(2) 音响仿真: 音响空间仿真是再现真实环境的基本要素,要能反映出房间的大小、墙面的特点等。

(3) 速度和效能: 在空间中声音的物理性质的精确仿真和声音实时有效生成间往往存在矛盾,因此一般需要有个折中。同时,实现虚拟环境还需要一定数目的虚拟声源。大多数重要的声音现象可以用计算机引擎仿真。值得注意的一个挑战是,如何在任意虚拟位置将声源映射到一定数目的喇叭,而其实际位置又受到 VR 系统安装时的物理设置所限制。

3.3.1 三维虚拟声音的特征

三维虚拟声音的特征主要包括全向三维定位特性和三维实时跟踪特性。

(1) 全向三维定位特性是指在三维虚拟环境中把实际声音信号定位到特定虚拟声源的能力。它能使用户准确地判断出声源的精确位置,从而符合人们在真实世界中的听觉方式。

(2) 三维实时跟踪特性是指在三维虚拟环境中实时跟踪虚拟声源位置变化和虚拟影像变化的能力。当用户转动头部时,这个虚拟声音的位置也应随之变动,使用户感到声源的位置并未发生变化。而当虚拟发声物体移动位置时,其声源位置也应有所改变,因为只有声音效果与实时变化的视觉一致,才可能产生视觉与听觉的叠加和同步效应。

举例来说,设想在虚拟房间中有一台正在播放节目的电视。如果用户站在距离电视较远的地方,则听到的声音也将较弱,但只要他逐渐走近电视,就会感受到越来越大的声音效果;当用户面对电视时,会感到声源来自正前方,而如果此时向左转动头部或走到电视左侧,他就会立刻感到声源已处于自己的右侧。这就是虚拟声音的全向三维定位特性和三维实时跟踪特性。可以说,一套性能良好的三维声音系统将能使所有虚拟声音的体验与人们在现实生活中取得的经验相同。



3.3.2 头部相关传递函数

在虚拟环境中构建较完美的三维声音系统是一个极其复杂的过程。为了建立三维虚拟声音,一般可以先从一个最简单的单耳声源开始,然后让它通过一个专门的回旋硬件,生成分离的左右信号,便可以使一个戴耳机的实验者准确地确定声源在空间的位置。实际上,在听觉定位过程中,声波要经过头、躯干和外耳构成的复杂外形对其产生的散射、吸收等作用之后,才能传递到鼓膜。当相同入射声波的方向不同时,到达鼓膜的声音频率成分就不同,此改变依赖于入射声波的方向以及人头部、外耳、躯干的形状与声学特性。为此,经研究人员实验证明,首先通过测量外界声音与鼓膜上声音的频率差异,获得了声音在耳部附近发生的频谱变形,随后利用这些数据对声波与人耳的交互方式进行编码,得出相关的一组传递函数,并确定出两耳的信号传播延迟特点,以此对声源进行定位。通常在 VR 系统中,当无回声的信号由这组传递函数处理后,再通过与声源缠绕在一起的滤波器驱动一组耳机,就可以在传统的耳机上形成有真实感的三维声音了。由于这组传递函数与头部有关,故被称为头部相关传递函数。由此看出,头部相关传递函数可视为声音在人体周围位置包含人体特征的函数。当获得的头部相关传递函数能够准确描述某个人的听觉定位过程时,利用它就能够模拟再现真实的声音场景。

由于每个人的头、耳的大小和形状各不相同,头部相关传递函数也会因人而异。但目前已有研究开始寻找对各种类型都通用且能提供良好效果的头部相关传递函数。

3.3.3 语音合成技术

语音合成技术是一门综合性的前沿新技术,该技术相当于给机器装上了人工嘴巴。它涉及声学、语言学、数字信号处理、计算机科学等多个学科技术。在 VR 系统中,语音合成技术与传统的录音-回放设备(系统)有着本质的区别。它实际上包括语音识别、文语转换和电子控制下的机械发声等三大关键技术。

由于交互的需要,用户可以向 VR 系统自由地用语音或者是文字传递信息,而 VR 系统则可通过语音合成技术用声音反馈给用户。

由于语音与普通的声音不同,具有特殊的波形纹理和周期,并且因语言和人的不同有着较大的差异,这一点使得机器在语音识别过程中,需要进行语音信号的预处理、特征提取、模式匹配等几个步骤的数据处理。预处理包括预滤波、采样和量化、加窗、端点检测、预加重等过程。其中,特征参数提取是语音信号识别中最为重要的一环。

文语转换简称为 TTS,该技术可将外部输入的文字信息转变为可识别的语音输出,附属于语音合成技术的一部分,从原理上看,该技术包括语言学处理、韵律建模和声学处理(即合成语音)。

语音合成的理论基础是语音生成的数学模型。该模型表现的语音生成过程是在激励信号的激励下,声波经谐振腔(声道),由嘴或鼻辐射声波。因此,声道参数、声道谐振特性一直是研究的重点。基于上述原理,目前有多种语音合成方法,如共振峰合成法、LPC 参数合成法、PSOLA 合成技术、LMA 声道模型合成法等。



将语音合成与语音识别技术结合起来,可以使用户与计算机所创建的虚拟环境进行简单的语音交流,这在 VR 环境中具有突出的应用价值,特别是当使用者的双手正忙于执行其他任务,双眼无暇注视图像时,这个语音交流的功能就显得尤为重要了。

3.4 情感计算

在人们的传统观念里,机器的运行模式似乎就是一种冷冰冰的机械运动,不懂人情世故。但能否设想一下,如果机器也具有人的智慧属性,能够实时地关注和体会人们的喜怒哀乐,并见机行事呢?正如马云所说,梦想总是要有的,也许有一天真的实现了机器对人的查颜观色,那会带给人们怎样的一种感受?情感计算研究的目标就是试图创建一种能感知、识别和理解人的情感,并能针对人的情感做出智能、灵敏、友好反应的计算系统,即赋予计算机像人一样的观察、理解和生成各种情感特征的能力。

3.4.1 情感计算的提出

在脑科学、认知科学和人工智能发展的很长一段时期内,人们的情感行为一直位于其科学研究者的法眼之外,直到 1997 年,美国麻省理工学院媒体实验室的罗莎琳德·皮卡德(Rosalind W. Picard)教授首次提出了情感计算的概念,使得情感计算一夜之间被世人所关注,并作为一个新兴的研究领域受到了诸多领域专家的重视。

在罗莎琳德·皮卡德于 1997 年正式出版的 *Affective Computing* 专著中,她指出情感计算就是针对人类的外在表现,能够进行测量和分析并能对情感施加影响的计算,并提出了情感强度计算第一定律。在该定律中,她用了一条函数抛物线来揭示情感强度是事物的价值率高差在人的头脑中的主观反映值,虽然事物的价值率高差在根本上决定着人的情感强度,但情感的强度并不与事物的价值率高差成正比,而是一种特殊的函数关系。也许有的时候人们的情感看起来会是喜怒无常、变幻难测的,但那种现象只是在表面上体现出很强的形式多样性、产生随机性和作用模糊性,而实际上却遵循着较为深奥的内在规律性,只要沿着正确的逻辑思路,采用正确的研究方法,就不难发现情感变化的内在规律性。

由于情感计算技术的研究目的,是通过赋予计算机识别、理解、表达和适应人的情感的能力,使计算机能够对人们的音容笑貌“一解风情”,并且通过多种媒体做出智能、友好的反应,最终能够创建一个和谐的人机交互环境,因而情感计算技术对于诸多行业来说,都是一种新的思维模式和未来方向。当人们认真探讨与事物价值率变化的对数成正比的函数规律时,就会发现人类情感的发展变化,不仅为未来的人工智能技术奠定了理论基础,同时也为人机自然交互提供了一个核心的研究课题。

中国科学院自动化研究所的 IEEE 高级会员胡包刚教授通过自己的深入研究,提出了对情感计算的定义:“情感计算的目的是通过赋予计算机识别、理解、表达和适应人的情感的能力来建立和谐人机环境,并使计算机具有更高的、全面的智能。”

情感计算从本质上来说就是一个典型的模式识别问题。智能机器通过多种传感器,



获取人的表情、姿态、手势、语音、语调、血压、心率等各种数据,结合当时的环境、语境、情境等上下文信息,识别和理解人的情感。在实际的自然交互系统中,智能机器还需要对上述信息做出实时的、恰当的、情感化的反应。情感之间距离的定义和计算方法是情感计算的核心问题,例如,需要定义和计算“微笑、冷笑、开心大笑、抑制不住的狂笑”之间的距离,以便把它们分别聚类,从而使系统能够识别出不同程度的笑。遗憾的是,目前情感计算的研究还只能对情感进行粗略的分类,即仅能识别 7 种典型的情感。

自从情感计算的新概念提出以后,全世界的许多实验室都积极投入到了对情感计算相关技术的研究中,首先是美国人工智能协会(AAAI)在 1998 年、1999 年和 2004 年连续组织召开专业的学术会议对人工情感和认知进行研讨。同样,国内的研究学者也开展了许多的研究工作和学术活动。2003 年 12 月,在北京召开了第一届中国情感计算及智能交互学术大会。这次会议集合了世界一流的情感计算、人工情绪和人工心理研究的著名专家学者。由此可见,我国的人工情感研究正逐步展开并向国际水平看齐。

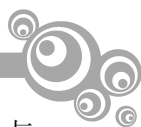
麻省理工学院媒体实验室的情感计算小组研制的情感计算系统,通过记录人面部表情的摄像机和连接在人身体上的生物传感器来收集数据,然后由一个“情感助理”来调节程序以识别人物的情感。如果你对电视讲座的一段内容表现出困惑,情感助理会重放该片段或者给予解释。IBM 公司开始实施“蓝眼计划”和开发“情感鼠标”;Affectiva(情绪识别)公司的 Affectiva 通过网络摄像头,使用计算机视觉和深度学习技术分析面部(微)表情或网络上视觉内容中非语言的线索,从而积累了庞大的数据存储库,用于学习识别更复杂的系统,并将情感人工智能引入到了机器人、医疗、教育和娱乐领域。

日本从 20 世纪 90 年代就开始了感性工学(Kansei Engineering)的研究。所谓感性工学就是将感性与工程结合起来的技术,即在感性科学的基础上,通过分析人类的感性,把人的感性需要加入到商品设计、制造中,目前日本已经形成了举国研究感性工学的高潮。

在欧洲,许多国家也积极地投入到了对情感信息处理技术(表情识别、情感信息测量、可穿戴计算等)的研究中。欧洲许多大学都成立了情感与智能关系的研究小组。其中比较著名的有:日内瓦大学 Klaus Soberer 领导的情绪研究实验室,布鲁塞尔自由大学的 D. Canamero(卡纳梅罗)领导的情绪机器人研究小组以及英国伯明翰大学的 A. Sloman 领导的 Cognition and Affect Project。剑桥大学、飞利浦公司等则通过实施“环境智能”“环境识别”“智能家居”等科研项目来开辟这一领域。

我国对人工情感和认知的理论和技术的研究始于 20 世纪 90 年代,大部分研究工作是针对人工情感单元理论与技术的实现展开的。进入 21 世纪以后,在我国特别是近年来,随着普适计算、人本计算、社会计算等概念和研究方向的提出,人机自然交互也同时日益成为各研究领域的热点研究内容和项目,情感计算自然地成为各学科共同关注的焦点。中国国家自然科学基金委也不失时机地支持了“情感计算理论与方法”的研究。

例如,哈尔滨工业大学机器人技术与系统国家重点实验室实现语音情感交互系统,提出了智能情感机器人进行情感交互的框架,设计实现了智能服务机器人的情感交互系统。北京航空航天大学基于特征参数的语音情感识别并能有效识别语音情感。中国科学技术大学基于特权信息的情感识别,提出了融合用户脑电信号和视频内容的情感视频标注方法,以某一模态特征为特权信息的情感识别和视频情感标注方法。同时,清华大



学信息科学与技术国家实验室、中国科学院心理研究所、行为科学院重点实验室均参与到了情绪识别的相关方面的研究中。2015年,我国的阅面科技(ReadSense)人工智能公司推出了情感认知引擎 ReadFace。由云(利用数学模型和大数据来理解情感)和端(SDK)共同组成,嵌入任何具有摄像头的设备来感知并识别表情,输出人类基本的表情运动单元、情感颗粒和人的认知状态,该系统已经成功应用于互动游戏智能机器人、视频广告效果分析、智能汽车、人工情感陪伴等。

3.4.2 情感计算的系统架构

情感计算研究的重点就在于通过各种高效传感器获取由人的情感所引起的生理及行为特征信号,建立“情感模型”,从而创建感知、识别和理解人类情感的能力,并能针对用户的情感做出智能、灵敏、友好反应的个人计算系统,缩短人机之间的距离,营造真正和谐的人机环境。分析情感计算的系统,情感计算从功能上可以划分为以下四个主要组成部分。

(1) 通过人机交互接口,借助传感器进行高效的用户信息的获取,同时加入上下文环境、语境、情境信息,以及情感机理的基本原理。

(2) 将获取的交互信息构建分析模型和数字化处理(去噪、降维)。

(3) 将得到的结果进行分析、处理、对比学习从而达到正确的理解。

(4) 将计算机所获取和转换的信息通过有效的方式呈现在用户面前,从而完成人机情感交互的全过程。

情感计算的系统框架如图 3.11 所示。

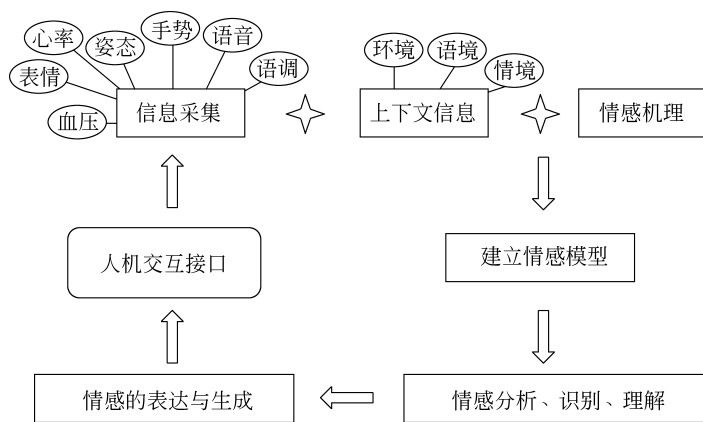


图 3.11 情感计算的系统框架

概括而言,通过传感器直接或间接与人接触获得情感信息;通过建立模型对情感信息进行分析与识别;对分析结果进行推理达到感性的理解;将理解结果通过合理的方式表达出来,从而完成了情感交流的全过程周期。

情感计算系统的四个主要功能组成中,信息的采集、获取十分重要,由于人类情感的复杂性特点,进行情感测量成为首先遇到的困难,因为情感测量需要对包括情感维度、表情和生理指标等三种成分的测量。在实际操作中,目前常常采用的方法是通过



一些采集输入设备提取人的面部表情、语音语调和肢体动作,再进行特征提取。此外,还通过测量人的一些生理反应,包括心率、血压、脉搏、瞳孔是否扩大、呼吸、皮肤导电、肾上腺素、荷尔蒙胆汁的分泌以及皮色体温等用于情感状态的识别理解。总之获取的有效数据越多,对后期的科学分析判断自然就越有利。其次,情感信息的分析和识别主要是对所提取到的信息进行预处理、模式分类。而最后的情感信息的理解就是根据上一步的分类结果和数据库中的模板进行比对判断,把所提取到的情感以最大概率确定出来,然后合成表情。

从目前的研究状况来看,我国在情感计算这一领域的研究仍然主要表现在人脸识别这一方面,究其原因还是因为人脸表情容易获取,易于分析处理,其成果具有重要的应用前景等;同时这一现象也反映了情感计算研究的一个普遍难题,那就是如何通过表情、语言、动作等各种信息的融合,识别和理解人类的情感,因为当前对于多模态情感数据获取、分析、融合、识别和理解,以及情景等上下文信息的获取依然是情感计算研究中最富有挑战性的课题。只有该课题获得真正突破,才能真正实现具有情感反馈的人机自然交互系统目标。

3.4.3 情感计算的相关技术

有关人类情感的研究,实际上很早就已经开始了。当时人们在研究心理学、认知科学的时候,就关注到了情感的存在,同样一些作家在自己的作品中,也描绘了情感是人类特有的精神力量。但是把人类的情感与冰冷的机器联系起来还是在 20 世纪 90 年代。探讨情感计算无疑是一个高度综合化的技术课题,它需要多学科的领域知识,尽可能多地获取人们的多模态信息和生理方面的相关参数变化,加入上下文的环境信息,再通过创新性的建模与分析、识别和情感理解,最终才能制作出具有情感反馈的人机交互环境,满足人的情感需求。可以说,要想真正实现人机的情感交互这一目标,需要完成的计算研究非常庞大。简单归纳,其主要研究内容如下。

1. 情感机理的研究

探究人类的情感机理,首先需要了解心理学、生理学、认知科学等对情感机理的解读观点。从心理学的角度出发,情感是由客观事物引起的,离开了客观事物,人不可能自发地产生情感。情感的实质是以主体的需要为基础,通过认知,明确客观事物与主体需求之间的关系,从而实现个体对客观现实的一种反映形式。而社会学和认知心理学的研究表明,人们在相关外界信息的刺激下,情感能够快速、轻易、自动甚至无意识地唤起。换句话说,人类没有无缘无故的爱也没有无缘无故的恨,同样,人类也没有无因果的高兴和无因果的痛苦,人们的喜怒哀乐缘于外部客观信息的刺激,而人们的情感反映都可能会伴随着人们外部表情和几种生理或行为特征的变化。因此,利用人们的外部表情,以及生理或姿态行为特征来确定不同情感状态之间的对应关系,就是情感机理所要解决的问题,该问题也是情感计算的前提条件之一。

2. 情感信号的获取

情感信号的获取研究的对象是当人们的情感发生改变时,如何尽可能多地获取人的各类视觉和音频数据,同时还包括生理变化的数据以及客观环境方面的数据。在获取信