第3章 虚拟现实的核心技术

导 学

内容与要求

本章介绍了虚拟现实当前流行的核心技术的分类、特点、应用，各种技术的原理和实现方法，以及核心技术在智能医学中的应用。

三维建模技术中要求掌握三种建模技术的建模特点以及区别。

立体显示技术中要求掌握分色技术、分光技术、分时技术、光栅技术、全息投影技术的实现原理。

真实感实时绘制技术中要求了解真实感绘制技术和实时绘制技术如何实现。

虚拟声音的实现技术中要求了解虚拟声音的原理，虚拟声音与传统立体声的区别，如何实现三维虚拟声音。

人机交互技术中要求了解当前人机交互技术的发展现状。

碰撞检测技术中要求掌握碰撞检测技术的技术要求以及实现方法。

重点和难点

本章的重点是虚拟现实系统的核心技术分类、特点、应用。难点是虚拟现实系统相关技术的原理和实现方法。

虚拟现实系统主要包括模拟环境、感知、自然技能和传感设备等方面。是由计算机生成虚拟世界，用户能够进行视觉、听觉、触觉、力觉、嗅觉、味觉等全方位交互。现阶段在计算机的运行速度达不到虚拟现实系统所需要的情况下，相关技术就显得尤为重要。要生成一个三维场景，并使场景图像能随视角不同实时地显示变化，除了必需的设备外，还要有相应的技术理论相支持。随着智能医学的发展，虚拟现实技术逐步呈现与医疗型人工智能相结合的发展方向，并在各个领域已有了展现。

3.1 三维建模技术

虚拟环境建模的目的在于获取实际三维环境的三维数据，并根据其应用的需要，利用获取的三维数据建立相应的虚拟环境模型。只有设计出反映研究对象的真实有效的模型，虚拟现实系统才有可信度。

虚拟现实系统中的虚拟环境，可能有下列3种情况：

(1) 模仿真实世界中的环境(系统仿真)。

(2) 人类主观构造的环境。

(3) 模仿真实世界中人类不可见的环境(科学可视化)。

三维建模一般主要指三维视觉建模。三维视觉建模可分为几何建模、物理建模、行为建模。

3.1.1 几何建模技术

几何建模是在虚拟环境中建立物体的形状和外观,产生实际的或想象的模型。虚拟环境中的几何模型是物体几何信息的表示,设计表示几何信息的数据结构、相关的构造与操纵该数据结构的算法。虚拟环境中的每个物体包含形状和外观两个方面。物体的形状由构造物体的各个多边形、三角形和顶点等来确定,物体的外观则由表面纹理、颜色、光照系数等来确定。因此,用于存储虚拟环境中几何模型的模型文件应该提供上述信息。

通常几何建模可通过以下两种方式实现:

1. 人工几何建模

利用虚拟现实工具软件编程进行建模,如 OpenGL、Java3D、VRML 等。这类方法主要针对虚拟现实技术的建模特点而编写,编程容易、效率较高。直接从某些商品图形库中选取所需的几何图形,可以避免直接用多边形拼构某个对象外形时烦琐的过程,也可节省大量的时间,如图 3.1 所示。

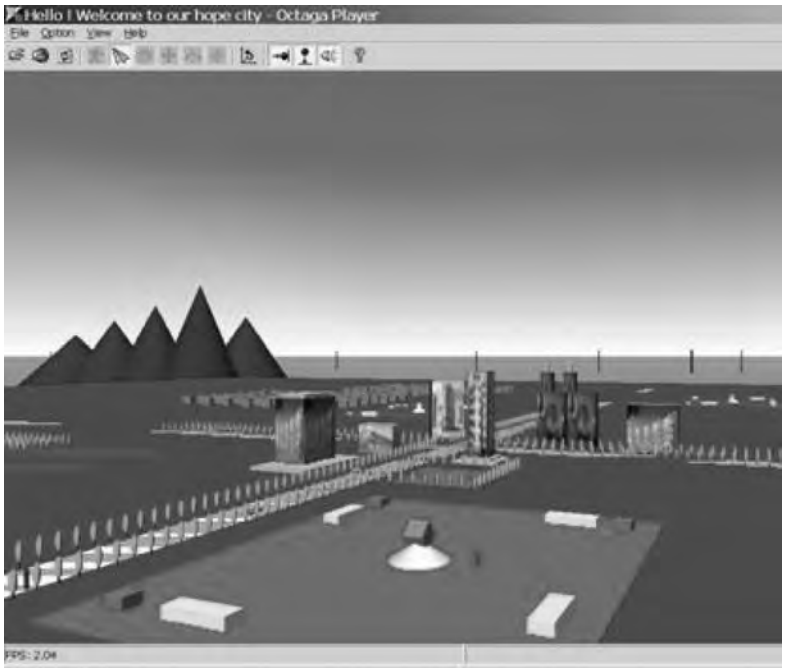

图 3.1 应用 VRML 语言实现的城市环境的模拟

利用交互式建模软件来进行建模,如 AutoCAD、3ds Max、Maya、Autodesk 123D 等,如图 3.2 所示。用户可交互式地创建某个对象的几何图形,但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工具软件。

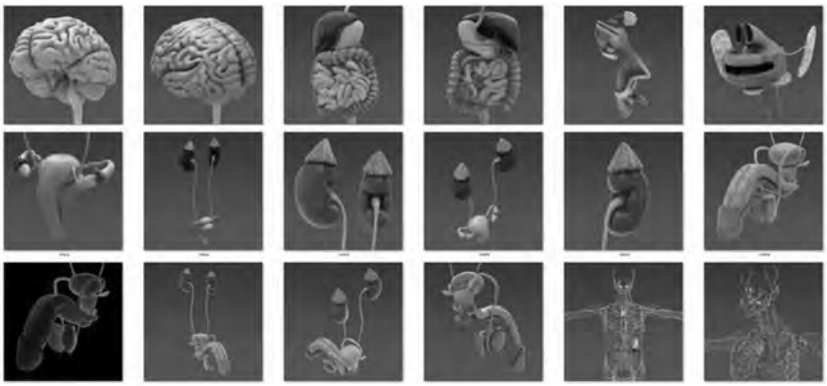

图 3.2　应用 3ds Max 制作的人体器官模型

2. 数字化自动建模

数字化自动建模方法主要是采用三维扫描仪对实际物体进行三维扫描，实现数字化自动建模。例如线激光手持式三维扫描仪（如图 3.3 所示），其原理是自带校准功能，工作时将激光线照射到物体上，两个相机来捕捉这一瞬间的三维扫描数据，由于物体表面的曲率不同，光线照射在物体上会发生反射和折射，然后这些信息会通过第三方软件转换为 3D 模型。在扫描过程中移动扫描仪，哪怕扫描时动作很快，也同样可以获得很好的扫描效果。

图 3.3　激光手持式三维扫描仪

3.1.2　物理建模技术

物理建模是对虚拟对象的质量、重量、惯性、表面纹理（光滑或粗糙）、硬度、变形模式（弹性或可塑性）等特征的建模。物理建模是虚拟现实系统中比较高层次的建模，它需要物理学与计算机图形学配合，涉及力的反馈问题，主要是质量建模、表面变形和软硬度等物理属性的体现。典型的物理建模方法有分形技术和粒子系统。

1. 分形技术

分形技术是指可以描述具有自相似特征的数据集。在虚拟现实系统中一般仅用于静态远景的建模。最常见的自相似例子是树，不考虑树叶的区别，树枝看起来也像一棵大树。当然，由树枝构成的树从适当的距离看时自然也是棵树。与此类似的如一棵蕨类植物，如图 3.4 所示。这种结构上的自相似称为统计意义上的自相似。自相似结构可用于复杂的不

规则外形物体的建模。该技术首先被用于河流和山体的地理特征建模。举一个简单的例子，如图 3.5 所示。可利用三角形来生成一个随机高度的地形模型。取三角形三边的中点并按顺序连接起来，将三角形分割成 4 个三角形，在每个中点随机赋予一个高度值，然后递归此过程，就可产生近似山体的形态。分形技术的优点是用简单的操作就可以完成复杂的不规则物体建模，缺点是计算量太大，不利于实时性。

图 3.4　蕨类植物形态

图 3.5　分形技术模拟山体形态

2. 粒子系统

粒子系统是一种典型的物理建模系统，它是用简单的体素完成复杂的运动建模。体素的选取决定了建模系统所能构造的对象范围。粒子系统由大量称为粒子的简单体素构成，每个粒子具有位置、速度、颜色和生命期等属性，这些属性可根据动力学计算和随机过程得到。常使用粒子系统建模的有：火、爆炸、烟、水流、火花、落叶、云、雾、雪、尘和流星尾迹等。如图 3.6 所示是使用粒子系统建模的烟花效果图。

图 3.6　粒子系统建模的烟花效果图

3.1.3 行为建模技术

行为建模是建立模型的行为能力，并且使模型服从一定的客观规律。虚拟现实的本质就是客观世界的仿真或折射，虚拟现实的模型则是客观世界中物体或对象的代表。而客观世界中的物体或对象除了具有表观特征如外形、质感以外，还具有一定的行为能力和符合客观规律。例如，把桌子上的重物移出桌面，重物不应悬浮在空中，而应当做自由落体运动。因为重物不仅具有一定外形，而且具有一定的质量并受到地球引力的作用。

作为虚拟现实自主性的特性的体现，除了对象运动和物理特性对用户行为直接反应的数学建模外，还可以建立与用户输入无关的对象行为模型。虚拟现实的自主性的特性，简单地说是指动态实体的活动、变化以及与周围环境和其他动态实体之间的动态关系，它们不受用户的输入控制（即用户不与之交互）。例如战场仿真虚拟环境中，直升飞机螺旋桨的不停旋转；虚拟场景中的鸟在空中自由地飞翔，当人接近它们时，它们要飞远等行为。

3.2 立体显示技术

立体显示技术是虚拟现实的关键技术之一，它使人在虚拟世界里具有更强的沉浸感，立体显示的引入可以使各种模拟器的仿真更加逼真。因此，有必要研究立体成像技术并利用现有的计算机平台，结合相应的软硬件系统在平面显示器上显示立体视景。目前，立体显示技术主要以佩戴立体眼镜等辅助工具来观看立体影像。随着人们对观影要求的不断提高，将逐渐发展为裸眼式。目前比较有代表性的技术有：分色技术、分光技术、分时技术、光栅技术、全息投影技术。这些技术通常都以机器为主体，通过机器设备实现。

3.2.1 双目视差显示技术

由于人两眼有 4～6cm 的距离，所以实际上看物体时两只眼睛中的图像是有差别的，如图 3.7 所示。两幅不同的图像输送到大脑后，可以感到一个三维世界的深度立体变化，这就是所谓的立体视觉原理。据立体视觉原理，如果能够让左、右眼分别看到两幅在不同位置拍摄的图像，可以从这两幅图像感受到一个立体的三维空间。结合不同的技术产生不同的立体显示技术。只要符合常规的观察角度，即产生合适的图像偏移，形成立体图像。

图 3.7 立体显示技术原理

1. 分色技术

分色技术的基本原理是让某些颜色的光只进入左眼，另一部分只进入右眼。人眼睛中的感光细胞共有 4 种，其中数量最多的是感觉亮度的细胞，另外 3 种用于感知颜色，分别可以感知红、绿、蓝三种波长的光，感知其他颜色是根据这 3 种颜色推理出来的，因此红、绿、蓝被称为光的三原色。要注意这和美术上讲的红、黄、蓝三原色是不同的，后者是颜料的调和，而前者是光的调和。

显示器就是通过组合这三原色来显示上亿种颜色的，计算机内的图像资料也大多是用

三原色的方式存储的。分色技术在第一次过滤时要把左眼画面中的蓝色、绿色去除，右眼画面中的红色去除，再将处理过的这两套画面叠合起来，但不完全重叠，左眼画面要稍微偏左边一些，这样就完成了第一次过滤。第二次过滤是观众带上专用的滤色眼镜，眼镜的左边镜片为红色，右边镜片是蓝色或绿色，由于右眼画面同时保留了蓝色和绿色的信息，因此右边的镜片不管是蓝色还是绿色都是一样的。分色技术原理如图 3.8 所示。

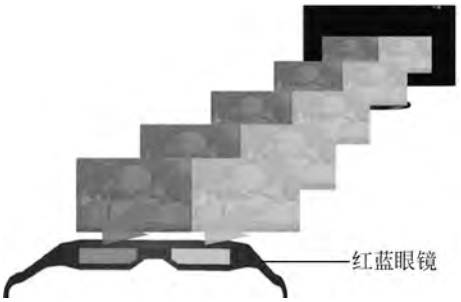

图 3.8　分色技术原理

也有一些眼镜右边为红色，这样第一次过滤时也要对调过来，购买产品时一般都会附赠配套的滤色眼镜，因此标准不统一也不用在意。以红、绿眼镜为例，红、绿两色互补，红色镜片会削弱画面中的绿色，绿色镜片削弱画面中的红色，这样就确保了两套画面只被相应的眼睛看到。其实准确地说是红、青两色互补，青介于绿和蓝之间，因此戴红、蓝眼镜也是一样的道理，如图 3.9 所示。目前，分色技术的第一次滤色已经开始用计算机来完成了，按上述方法滤色后的片源可直接制作成 DVD 等音像制品，在任何彩色显示器上都可以播放。

图 3.9　红蓝立体眼镜、红绿立体眼镜、棕蓝立体眼镜

2. 分光技术

分光技术的基本原理是当观众戴上特制的偏光眼镜时，由于左、右两片偏光镜的偏振轴互相垂直，并与放映镜头前的偏振轴相一致；致使观众的左眼只能看到左像、右眼只能看到右像，通过双眼汇聚功能将左、右像叠合在视网膜上，由大脑神经产生三维立体的视觉效果。如图 3.10 所示为分光技术原理。

图 3.10　分光技术原理

3. 分时技术

分时技术的基本原理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。如图 3.11 所示为分时技术原理。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，早期曾用过机械眼镜。

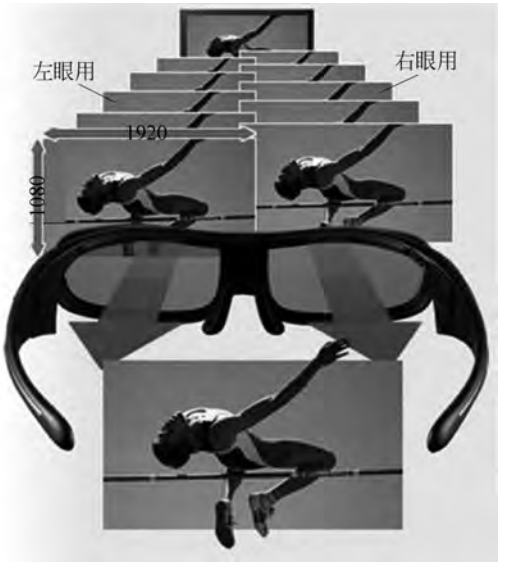

图 3.11　分时技术原理

4. 光栅技术

光栅技术的基本原理是在显示器前端加上光栅，光栅的功能是要挡光，让左眼透过光栅时只能看到部分的画面；右眼也只能看到另外一半的画面，于是就能让左、右眼看到不同影像并形成立体，此时无须佩戴眼镜，如图 3.12 所示。而光栅本身亦可由显示器所形成，也就是将两片液晶画板重叠组合而成，当位于前端的液晶面板显示条纹状黑白画面时，即可变成立体显示器；而当前端的液晶面板显示全白的画面时，不但可以显示 3D 的影像，也可同时相容于现有 2D 的显示器。

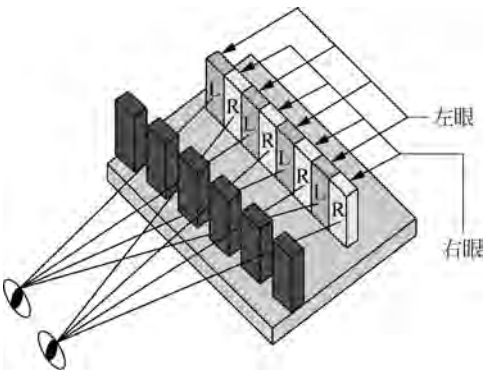

图 3.12　光栅 3D 显示技术原理

3.2.2　全息技术

全息技术是利用干涉和衍射原理记录并再现物体真实的三维图像，从而产生立体效果的一种技术。3D 全息投影技术原理是利用 3D 全息立体投影设备而不是应用数码技术实现的。投影设备将物体以不同角度投影到 MP 全息投影膜上，让观测者看到自身视觉范围内的图像，实现了 3D 全息立体影像。如图 3.13 为全息技术原理。

第一步：利用干涉原理记录物体光波信息，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，便成为一张全息图，或称全息照片。如图 3.14 所示为拍摄过程原理。

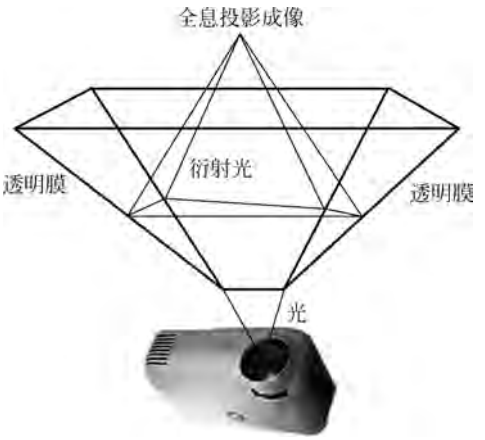

图 3.13　全息技术原理

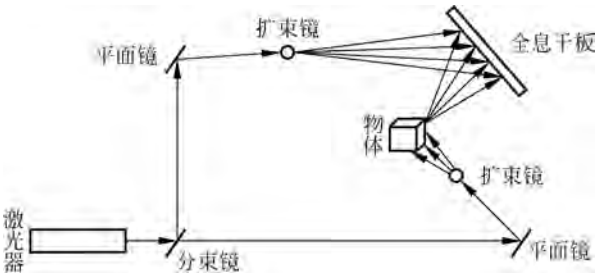

图 3.14　拍摄过程原理

第二步：利用衍射原理再现物体光波信息，这是成像过程。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型全息图的衍射光波一般可给出两个像，即原始像和共轭像。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。如图 3.15 所示为成像过程原理。如图 3.16 所示为在一个密闭容器内将空气电离，利用镭射激光可以呈现 3D 全息影像。

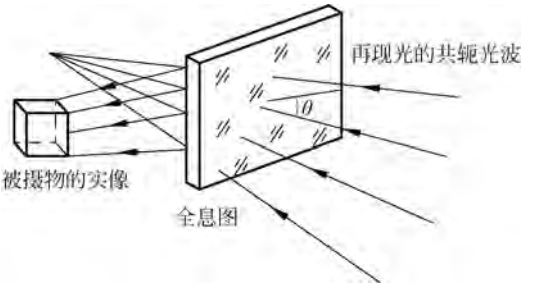

图 3.15　成像过程原理

图 3.16　3D 全息投影

3.3　真实感实时绘制技术

要实现虚拟现实系统中的虚拟世界，仅有立体显示技术是远远不够的，虚拟现实中还有真实感与实时性的要求，也就是说，虚拟世界的产生不仅需要真实的立体感，而且虚拟世界必须实时生成，这就要采用真实感实时绘制技术。

真实感实时绘制是在当前图形算法和硬件条件限制下提出的，在一定时间内完成真实感绘制的技术。“真实感”的含义包括几何真实感、行为真实感和光照真实感。几何真实感指与描述的真实世界中对象具有十分相似的几何外观；行为真实感指建立的对象对于观察者而言在某些意义上是完全真实的；光照真实感指模型对象与光源相互作用产生的与真实世界中亮度和明暗一致的图像。而“实时”的含义则包括对运动对象位置和姿态的实时计算与动态绘制，画面更新达到人眼观察不到闪烁的程度，并且系统对用户的输入能立即做出反应并产生相应场景以及事件的同步。它要求当用户的视点改变时，图形显示速度也必须跟上视点的改变速度，否则就会产生迟滞现象。

3.3.1　真实感绘制技术

真实感绘制技术是在当前图形算法和硬件条件限制下完成模型真实感的技术。其主要任务是要模拟真实物体的物理属性，即物体的形状、光学性质、表面纹理和粗糙程度，以及物体间的相对位置、遮挡关系等。

为了提高显示的逼真度，加强真实性，常采用下列方法：

（1）纹理映射：纹理映射是将纹理图像贴在简单物体的几何表面，以近似描述物体表面的纹理细节，加强真实性。实质上，它用二维的平面图像代替三维模型的局部。如图 3.17 所示为纹理映射前后的对比图。

（2）环境映射：采用纹理图像来表示物体表面的镜面反射和规则透视效果。如图 3.18 所示为环境映射效果图。

（3）反走样：走样是由图像的像素性质造成的失真现象。反走样方法的实质是提高像素的密度。反走样的一种方法是以两倍的分辨率绘制图形，再由像素值的平均值，计算正常

图 3.17　纹理映射前后的对比图

图 3.18　环境映射效果图

分辨率的图形。另一种方法是计算每个相邻接元素对一个像素点的影响，再把它们加权求和得到最终像素值。如图 3.19 所示为线型反走样对比图。

图 3.19　线型反走样对比图

3.3.2　实时绘制技术

实时绘制技术是侧重三维场景的实时性，在一定时间内完成绘制的技术。传统的虚拟场景基本上都是基于几何的，就是用数学意义上的曲线、曲面等数学模型预先定义好虚拟场景的几何轮廓，再采用纹理映射、光照等数学模型加以渲染。大多数虚拟现实系统的主要部分是构造一个虚拟环境，并从不同的方向进行漫游。要达到这个目标，首先是构造几何模型，其次模拟虚拟摄像机在 6 个自由度运动，并得到相应的输出画面。除了在硬件方面采用高性能的计算机，提高计算机的运行速度以提高图形显示能力外，还可以降低场景的复杂度，即降低图形系统需处理的多边形数目。有下面 5 种用来降低场景复杂度的方法。

（1）预测计算：根据各种运动的方向、速率和加速度等运动规律，可在下一帧画面绘制之前用预测、外推的方法推算出手的跟踪系统及其他设备的输入，从而减少由输入设备所带有的延迟。

（2）脱机计算：在实际应用中有必要尽可能将一些可预先计算好的数据进行预先计算并存储在系统中，这样可以加快需要运行时的速度。

（3）3D 剪切：将一个复杂的场景划分成若干个子场景，系统针对可视空间剪切。虚拟环境在可视空间以外的部分被剪掉，这样就能有效地减少在某一时刻所需要显示的多边形数目，以减少计算工作量，从而有效降低场景的复杂度。

（4）可见消隐：系统仅显示用户当前能“看见”的场景，当用户仅能看到整个场景很小部分时，由于系统仅显示相应场景，可大大减少所需显示的多边形的数目。

（5）细节层次模型（Level Of Detail，LOD）：首先对同一个场景或场景中的物体，使用具有不同细节的描述方法得到的一组模型。在实时绘制时，对场景中不同的物体或物体的不同部分，采用不同的细节描述方法。对于虚拟环境中的一个物体，同时建立几个具有不同细节水平的几何模型。通过对场景中每个图形对象的重要性进行分析，使得最重要的图形对象进行较高质量的绘制，而不重要的图形对象采用较低质量的绘制，在保证实时图形显示的前提下，最大程度地提高视觉效果。

例如，皮克斯动画公司员工展示了一款名为 Presto 的软件，拥有基于 OpenGL 开发的细分曲面技术。可以在实时运算速度下细分十几层，让角色动作、光影微调变成实时性。其原理是基于 CPU 与 GPU 的配合，GPU 加速计算能将应用程序计算密集部分的工作负载转移到 GPU，同时仍由 CPU 运行其余程序代码。从用户的角度来看，应用程序的运行速度明显加快。以往动画制作过程中微幅调整角色表情、动作、调整光源位置，都需要消耗很长的重新渲染时间。然而当 GPU 演算技术导入后，不仅节省了时间、资源，更让角色的动态、材质、光影更为细腻、精致、充满艺术韵味又不失真实感，丰富了 3D 影像，图 3.20 为实时绘制 3D 影像。

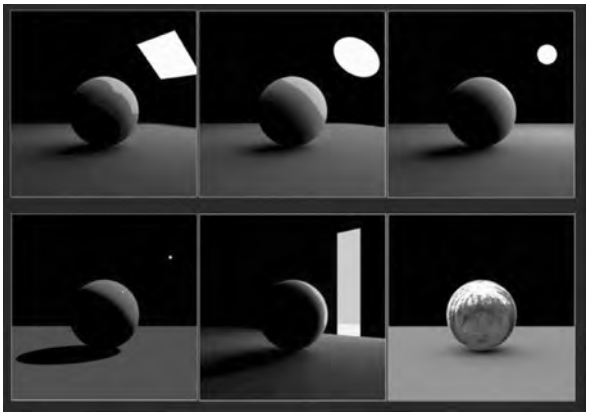

图 3.20　实时绘制 3D 影像

3.4　三维虚拟声音

三维虚拟声音技术的价值在于使用两个音箱模拟出环绕声的效果，虽然不能和真正的家庭影院相比，但是在最佳的听音位置上效果是可以接受的，其缺点是普遍对听音位置要求较高。

1. 三维虚拟声音

三维虚拟声音是在虚拟场景中能使用户准确地判断出声源精确位置、符合人们在真实境界中听觉方式的声音系统。

立体声是指具有立体感的声音。自然界发出的声音是立体声，但我们如果把这些立体声经记录、放大等处理后而重放时，所有的声音都从一个扬声器放出来，这种重放声就不是立体的了。这时由于各种声音都从同一个扬声器发出，原来的空间感也消失了。这种重放声称为单声。如果从记录到重放整个系统能够在一定程度上恢复原发生的空间感，那么，这种具有一定程度的方位层次等空间分布特性的重放声，称为音响技术中的立体声。

三维虚拟声音与人们熟悉的立体声音有所不同，但就整体效果而言，立体声来自听者面前的某个平面，而三维虚拟声音则是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、后方或前方。虚拟声音是在双声道立体声的基础上，不增加声道和音箱，把声场信号通过电路处理后播出，使聆听者感到声音来自多个方位，产生仿真的立体声场。例如，战场模拟训练系统中，当听到了对手射击的枪声时，就能像在现实世界中一样准确而且迅速地判断出对手的位置，如果对手在我们身后，听到的枪声就应是从后面发出的。

如图 3.21 所示，耳机中的立体声音听上去好像是从用户头的里面发出来的，而真实的声音应该在头的外面。当用户戴着简易的立体声耳机时，随着用户头部的移动，小提琴的声音也跟着改变了方向。而从同一个耳机或扬声器中的三维声音则包含着重要的心理信息，这个心理信息可以改变用户的感觉，使用户相信这些三维声音真的来自于用户外面的环境。三维声音是使用头部跟踪器的数据合成的，虚拟小提琴在空间中的位置保持不变，因此声音听起来好像来头的外面。

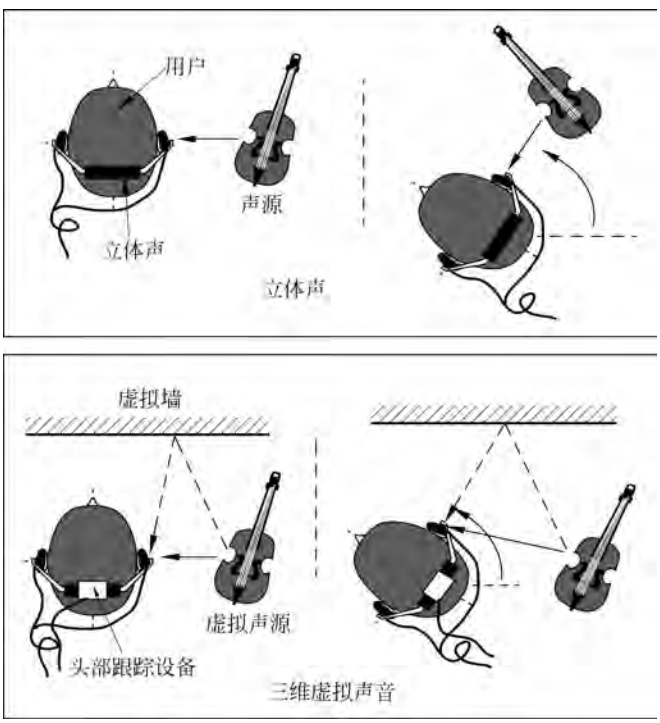

图 3.21　立体声音与三维声音的对比

2. 三维虚拟声原理

三维虚拟声的关键是声音的虚拟化处理，依据了人的生理声学和心理声学原理专门处理环绕声道，制造出环绕声源来自听众后方或侧面的幻象感觉。应用了人耳听音原理的4种效应。

(1) 双耳效应。英国物理学家瑞利于1896年通过实验发现人的两只耳朵对同一声源的直达声具有时间差(0.44～0.5μs)、声强差及相位差，而人耳的听觉灵敏度可根据这些微小的差别准确判断声音的方向、确定声源的位置，但只能局限于确定前方水平方向的声源，不能解决三维空间声源的定位。

(2) 耳廓效应。人的耳廓对声波的反射以及对空间声源具有定向作用。借此效应，可判定声源的三维位置。

(3) 人耳的频率滤波效应。人耳的声音定位机制与声音频率有关，对20～200Hz的低音靠相位差定位，对300～4000Hz的中音靠声强差定位，对高音则靠时间差定位。据此原理可分析出重放声音中的语言、乐音的差别，经不同的处理而增加环绕感。

(4) 头部相关传输函数。人的听觉系统对不同方位的声音产生不同的频谱，而这一频谱特性可由头部相关传输函数(Head Related Transfer Function，HRTF)来描述。

总之，人耳的空间定位包括水平、垂直及前后三个方向。水平定位主要靠双耳，垂直定位主要靠耳壳，而前后定位及对环绕声场的感受靠HRTF函数。虚拟杜比环绕声依据这些效应，人为制造与实际声源在人耳处一样的声波状态，使人脑在相应空间方位上产生对应的声像。

3.5 人机交互技术

在计算机系统提供的虚拟空间中，人可以使用眼睛、耳朵、皮肤、手势和语音等各种感觉方式直接与之发生交互，这就是虚拟环境下的人机自然交互技术。在虚拟相关技术中嗅觉和味觉技术的开发处于探索阶段，而恰恰这两种感觉是人对食物和外界最基础的需要。并且，随着智能移动设备的普及，人们的各种基础需求会不断得到满足。因此，气味传送或嗅觉技术的现实应用空间将会很大，也更能引起人们的兴趣。在虚拟现实领域中较为常用的交互技术主要有手势识别、面部表情的识别、眼动跟踪以及语音识别等。

3.5.1 手势识别技术

手势识别技术是用户可以使用简单的手势来控制或与设备交互，让计算机理解人类的行为。其核心技术为手势分割、手势分析以及手势识别。在计算机科学中，手势识别可以来自人的身体各部位的运动，但一般是指脸部和手的运动。

手势识别系统的输入设备主要分为基于数据手套的识别和基于视觉(图像)的识别系统两种，如图3.22所示。基于数据手套的手势识别系统，就是利用数据手套和位置跟踪器来捕捉手势在空间运动的轨迹和时序信息，对较为复杂的手的动作进行检测，包括手的位置、方向和手指弯曲度等，并可根据这些信息对手势进行分析。基于视觉的手势识别是从视觉通道获得信号，通常采用摄像机采集手势信息。由摄像机连续拍下手部的运动图像后，先采用轮廓的办法识别出手上的每一个手指，进而再用边界特征识别的方

法区分出一个较小的、集中的各种手势。手势识别技术主要有模板匹配、人工神经网络和统计分析技术。

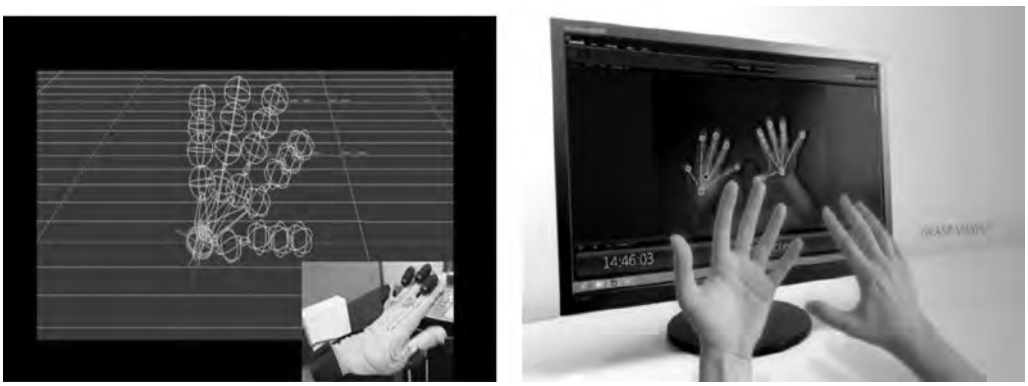

图 3.22　基于数据手套和基于视觉(图像)的两种手势识别技术

虚拟现实治疗(VRT)能创造出一种多感觉环境,让大脑实现超越药物治疗效果的生物结果。最初,VRT 只被用于治疗恐惧症、抑郁症、成瘾等心理问题,不过很快它就能被运用在神经康复治疗上了。例如,瑞士创业公司 MindMaze 推出了一套 VR 神经康复治疗系统 MindMotionPro,如图 3.23 所示。其原理是通过虚拟现实场景来刺激患者的大脑对身体做出相应的活动,让身体慢慢重新回到大脑的控制。由于 VR 可以模拟各种各样的运动场景,患者可以进行标准康复计划 10～15 倍的运动量,并且具有良好的趣味性,使患者忘记自己是在医院接受治疗,目前很多中风患者已经使用该平台进行康复治疗。

图 3.23　Mindmaze 公司的沉浸式治疗方式

Facebook 利用手势识别技术研发了一种可将周边日常物品作为手柄的全新体感交互系统:Gripmarks。其原理是利用手势识别技术通过与预设手势进行对比,识别使用者握住易拉罐等物体的手势,并计算手中物体的形状,为该物体生成支持交互的虚拟界面。当握住一样物体时,需要做出特定手势来将它作为手柄激活,而后使用方式类似于手机触屏,可以在该物体上进行多种不同功能的选择,比如播放媒体、计算等。目前,Gripmarks 可识别多种物体,包括书、钱包、圆锥形、子弹形、笔、苹果等,如图 3.24 所示。

图 3.24　体感交互系统 Gripmarks

3.5.2　面部表情识别技术

面部表情识别技术是用机器识别人类面部表情的一种技术。人可以通过脸部的表情表达自己的各种情绪，传递必要的信息。面部表情识别技术包括：人脸图像的分割、主要特征（如眼睛、鼻子等）定位以及识别。

一般人脸检测问题可以描述为：给定一幅静止图像或一段动态图像序列，从未知的图像背景中分割、提取并确认可能存在的人脸，如果检测到人脸，提取人脸特征。在某些可以控制拍摄条件的场合，将人脸限定在标尺内，此时人脸的检测与定位相对容易。在另一种情况下，人脸在图像中的位置是未知的，这时人脸的检测与定位将受以下因素的影响：人脸在图像中的位置、角度、不固定尺度以及光照的影响；发型、眼镜、胡须以及人脸的表情变化等；图像中的噪声等。

人脸检测的基本思想是建立人脸模型，比较所有可能的待检测区域与人脸模型的匹配程度，从而得到可能存在人脸的区域。

BinaryVR 是一款可以实现人脸识别虚拟现实的工具。在 VR 世界中，要把人脸转换成 3D 模型需要大量的建模工作，而且要实现低延时更是难上加难。而 BinaryVR 脸部识别技术则是通过将面部重要的信息点数据进行采集，快速转换成简单的 3D 模型，最后再通过立体动画的形式展现给观众。虽然在虚拟场景中展现的不是真正的脸庞，但依然可以栩栩如生地传达表情和情绪。其实现原理是通过头显内置摄像机并且结合应用软件来实现表情追踪。即使 VR 头显遮住了面部，也可以使用深度摄像头“分析”出我们的脸部表情，并在建模的时候用 20 种表情完成视觉的反馈，并实时复制到游戏或动画的人物中，让主人公的神情与玩家神情同步，如图 3.25 所示。

图 3.25　面部表情识别技术

VIPKID公司深度融合运用人脸识别技术实现课堂表情数据分析，优化教学方式。其原理是在教学过程中通过人脸识别、情绪识别等技术，抓取用户上课数据，对师生的表情进行分析，计算分析用户的视线关注情况。了解在有效的在线远程学习过程中，促成用户专注度形成与知识习得的关键因素，进而让学生提高关注度，更高效地学习，如图3.26所示。

图3.26　VIPKID通过面部识别技术分析学生课堂学习表现

3.5.3　眼动跟踪技术

人们可能经常在不转动头部的情况下，仅仅通过移动视线来观察一定范围内的环境或物体。为了模拟人眼的功能，在虚拟现实系统中引入眼动跟踪技术，如图3.27所示。

眼动跟踪技术是利用图像处理技术，使用能锁定眼睛的特殊摄像机。通过摄入从人的眼角膜和瞳孔反射的红外线连续地记录视线变化，从而达到记录、分析视线追踪过程的目的。表3.1归纳了5种主要的眼动跟踪技术及特点。

图3.27　眼动跟踪原理示意图

表 3.1　眼动跟踪技术及特点

视觉追踪方法	技 术 特 点
眼电图	高带宽，精度低，对人干扰大
虹膜-巩膜边缘	高带宽，垂直精度低，对人干扰大，误差大
角膜反射	高带宽，误差大
瞳孔-角膜反射	低带宽，精度高，对人无干扰，误差小
接触镜	高带宽，精度最高，对人干扰大，不舒适

3.5.4　语音识别技术

语音识别技术（Automatic Speech Recognition，ASR）是将人说话的语音信号转换为可被计算机程序所识别的文字信息，从而识别说话者的语音指令以及文字内容的技术，包括参数提取、参考模式建立和模式识别等过程。与虚拟世界进行语音交互是实现虚拟现实系统中的一个高级目标。语音技术在虚拟现实中的关键是语音识别技术和语音合成技术。

1. 语音识别方法主要是模式匹配法

（1）在训练阶段，用户将词汇表中的每一词依次说一遍，并且将其特征矢量作为模板存入模板库。

（2）在识别阶段，将输入语音的特征矢量依次与模板库中的每个模板进行相似度比较，将相似度最高者作为识别结果输出。

2. 语音识别的主要问题

（1）对自然语言的识别和理解。首先必须将连续的讲话分解为词、音素等单位，其次要建立一个理解语义的规则。

（2）语音信息量大。语音模式不仅对不同的说话人不同，对同一说话人也是不同的，例如，一个说话人在随意说话和认真说话时的语音信息是不同的。一个人的说话方式随着时间变化。

（3）语音的模糊性。说话者在讲话时，不同的词可能听起来是相似的。这在英语和汉语中常见。

（4）单个字母或词、字的语音特性受上下文的影响，以致改变了重音、音调、音量和发音速度等。

（5）环境噪声和干扰对语音识别有严重影响，致使识别率低。

随着人工智能的发展，语音识别技术在聊天机器人方面的研发和产品越来越多，并陆续推出了相关产品，比如苹果的 Siri、微软的 Cortana 与小冰、Google 的 Now、百度的“度秘”、亚马逊的蓝牙音箱 Amazon Echo 内置的语音助手 Alexa、Facebook 推出的语音助手 M、Siri 创始人推出的新型语音助手 Viv 等。对于聊天机器人技术而言，包括几种主流的技术，其原理如下：

基于人工模板的技术通过人工设定对话场景，并对每个场景写一些针对性的对话模板，模板描述了用户可能的问题以及对应的答案模板。这个技术路线的好处是精准，缺点是需要大量人工工作，而且可扩展性差，需要一个场景一个场景去扩展。应该说目前市场上各种类似于 Siri 的对话机器人中都大量使用了人工模板的技术，主要是其精准性是其他方法还无法比拟的。

基于检索技术的聊天机器人则走的是类似搜索引擎的路线，首先存储好对话库并建立索引，根据用户问句，在对话库中进行模糊匹配找到最合适的应答内容。

基于机器翻译技术的聊天机器人把聊天过程比拟成机器翻译过程，就是说将用户输入聊天 Message(信息)，然后聊天机器人 Response(应答)的过程识别为把 Message 翻译成 Response 的过程。基于这种假设，就完全可以将统计机器翻译领域里相对成熟的技术直接应用到聊天机器人开发领域中来。

基于深度学习的聊天机器人技术是在 Encoder-Decoder 深度学习技术框架下进行改进的。使用深度学习技术来开发聊天机器人相对传统方法来说整体思路是简单可扩展的。

Starship Commander 是一款被称为“星舰指挥官”的 VR 语音互动游戏。在虚拟世界中，可以像在现实生活中一样进行语音交流，自然地发表疑问、表达自己的想法，如图 3.28 所示。相比于设定好反馈模式的游戏角色，加入 AI 元素的全息指挥官能够更加智能地与玩家扮演的飞行员进行对话。玩家与 NPC 之间更加自然的互动关系，会加强游戏体验的真实感，让玩家更好地沉浸在 VR 世界之中。

图 3.28　全息指挥官与玩家扮演的飞行员进行对话

3.6　碰撞检测技术

碰撞检测是用来检测对象甲是否与对象乙相互作用。在虚拟世界中，由于用户与虚拟世界的交互及虚拟世界中物体的相互运动，物体之间经常会出现碰撞的情况。为了保证虚拟世界的真实性，就需要虚拟现实系统能够及时检测出这些碰撞，产生相应的碰撞反应，并及时更新场景输出，否则就会发生穿透现象。正是有了碰撞检测，才可以避免诸如人穿墙而过等不真实情况的发生，影响虚拟世界的真实感。如图 3.29 所示为虚拟现实系统中两辆车发生碰撞反应前后的状态。

1. 碰撞检测技术原理

在虚拟世界中关于碰撞，首先要检测到有碰撞的发生及发生碰撞的位置，其次是计算出发生碰撞后的反应。在虚拟世界中通常有大量的物体，并且这些物体的形状复杂，要检测这些物体之间的碰撞是一件十分复杂的事情，其检测工作量较大，同时由于虚拟现实系统中有较高实时性的要求，要求碰撞检测必须在很短的时间(如 30～50ms)完成，因而碰撞检测成了虚拟现实系统与其他实时仿真系统的瓶颈，碰撞检测是虚拟现实系统研究的一个重要技术。

图 3.29　虚拟现实中的碰撞检测技术

2. 碰撞检测技术的要求和实现方法

为了保证虚拟世界的真实性，碰撞检测要有较高的实时性和精确性。所谓实时性，基于视觉显示的要求，碰撞检测的速度至少要达到 24Hz；而基于触觉要求，速度至少要达到 300Hz 才能维持触觉交互系统的稳定性，只有达到 1000Hz 才能获得平滑的效果。精确性的要求取决于虚拟现实系统在实际应用中的要求。

最简单的碰撞检测方法是对两个几何模型中的所有几何元素进行两两相交测试。这种方法可以得到正确的结果，但当模型的复杂度增大时，计算量过大，十分缓慢。对两物体间的精确碰撞检测的加速实现，现有的碰撞检测算法主要可划分为两大类：层次包围盒法和空间分解法。

本 章 小 结

虚拟现实技术是由计算机产生，通过视觉、听觉、触觉等作用，使用户产生身临其境感觉的交互式视景仿真，具有多感知性、存在感、交互性和自主性等特征。本章阐述了虚拟现实系统核心技术的分类、特点及应用，重点介绍了三维建模、立体显示、真实感实时绘制、三维虚拟声音、人机自然交互以及碰撞检测几种技术。本章的学习要点是各种技术的原理和实现方法，以及一些技术在智能医学领域的应用。从虚拟现实技术的发展历程来看，虚拟现实技术在今后的发展过程中依然会遵循“低成本、高性能”的原则，主要的发展方向为：动态环境建立技术、实时三维图像生成与显示、研制新型交互设备、智能语音虚拟建模、应用大型分布式网络虚拟现实。随着它在越来越多的领域的应用与发展，必将给人类带来巨大的经济和社会效益。

【注释】

曲率：针对曲线上某个点的切线方向角对弧长的转动率，通过微分来定义，表明曲线偏离直线的程度。数学上表明曲线在某一点的弯曲程度的数值。曲率越大，表示曲线的弯曲程度越大。曲率的倒数就是曲率半径。

体素（Volumepixel）：用来构造物体的原子单位，包含体素的立体可以通过立体渲染或者提取给定阈值轮廓的多边形等值面表现出来。

三维扫描仪（3D scanner）：一种科学仪器，用来侦测并分析现实世界中物体或环境的形状（几何构造）与外观数据（如颜色、表面反照率等性质）。搜集到的数据常被用来进行三维

重建计算，在虚拟世界中创建实际物体的数字模型。

电压传感器：一种将被测电量参数转换成直流电流、直流电压并隔离输出模拟信号或数字信号的装置。

结构光：激光从激光器发出，经过柱面透镜后汇聚成宽度很窄的光带。

摄影测量学：研究利用摄影手段获得被测物体的图像信息，从几何和物理方面进行分析处理，对所摄对象的本质提供各种资料的一门学科。

景深：指在摄影机镜头或其他成像器前沿能够取得清晰图像的成像所测定的被摄物体前后距离范围。

自然光：又称“天然光”，不直接显示偏振现象的光。天然光源和一般人造光源直接发出的光都是自然光。它包括垂直于光波传播方向的所有可能的振动方向，所以不显示出偏振性。从普通光源直接发出的天然光是无数偏振光的无规则集合，所以直接观察时不能发现光强偏于哪一个方向。这种沿着各个方向振动的光波强度都相同的光叫作自然光。

偏振光（Polarized light）：光学名词。光是一种电磁波，电磁波是横波。而振动方向和光波前进方向构成的平面叫做振动面，光的振动面只限于某一固定方向的，叫作平面偏振光或线偏振光。

光栅（Grating）：由大量等宽等间距的平行狭缝构成的光学器件。一般常用的光栅是在玻璃片上刻出大量平行刻痕制成，刻痕为不透光部分，两刻痕之间的光滑部分可以透光，相当于一狭缝。

干涉（Interference）：物理学中指两列或两列以上的波在空间中重叠时发生叠加从而形成新的波形的现象。

振幅：指振动的物理量可能达到的最大值，通常以 A 表示。它是表示振动的范围和强度的物理量。

OpenGL：一个跨编程语言、跨平台的编程接口规格的专业的图形程序接口。

杜比：杜比是英国 R. M. DOLBY 博士的中译名，他在美国设立的杜比实验室，先后发明了杜比降噪系统、杜比环绕声系统等多项技术，对电影音响和家庭音响产生了巨大的影响。家庭中常常用到的杜比技术主要包括杜比降噪系统和杜比环绕声系统。

生理声学：声学的分支，主要研究声音在人和动物引起的听觉过程、机理和特性，也包括人和动物的发声。

心理声学（Psychoacoustics）：研究声音和它引起的听觉之间关系的一门边缘学科。心理声学就是指“人脑解释声音的方式”。压缩音频的所有形式都是用功能强大的算法将我们听不到的音频信息去掉。

声强差效应：如果一个声音来自听者正前反正前方的中轴线上，那么，声音到达双耳的声音大小是一样的，于是听者就觉得这个声音处在前方；倘若声音来自听者人的左侧，听者人就觉得声源偏左。

模板匹配：数字图像处理的重要组成部分之一。把不同传感器或同一传感器在不同时间、不同成像条件下对同一景物获取的两幅或多幅图像在空间上对准，或根据已知模式到另一幅图中寻找相应模式的处理方法就叫做模板匹配。

人工神经网络（Artificial Neural Network，ANN）：从信息处理角度对人脑神经元网络进行抽象，建立某种简单模型，按不同的连接方式组成不同的网络。在工程与学术界也常

直接简称神经网络或类神经网络。

统计分析：指运用统计方法及与分析对象有关的知识，从定量与定性的结合上进行的研究活动。它是继统计设计、统计调查、统计整理之后的一项十分重要的工作，是在前几个阶段工作的基础上通过分析从而达到对研究对象更为深刻的认识。

Encoder-Decoder：编码-解码模型。这是一种应用于 seq2seq 问题的模型。

包围盒算法：包围盒检测法就是将物体简化为多面体或球体，计算两个待测实体中心点的距离与它们半径之和的关系，以此来判定两物体是否可能碰撞。

空间分解法：空间分解法则将包含实体的空间划分为多个子空间，子空间中的实体按照一定顺序进行排列，碰撞检测只限制在某个子空间中进行。