**1 引言**

**1.1 增强现实概述**

增强现实（Augmented Reality，简称AR）技术是一种通过计算摄像机影像的位置和角度，再配合上图像分析技术，使得显示设备上的虚拟物体能够与现实场景进行结合与互动的技术。AR以一种更为直观高效、生动有趣的方式将虚实结合的信息通过显示设备传递给用户，使用户与计算机之间的交互更为自然，产生更为真实的体验。

对于增强现实的具体定义，目前有两种主流的声音。一种是由Paul Milgram和Fumio Kishino于1994年在论文中提出的“现实-虚拟连续体”概念。Paul Milgram和Fumio Kishino对虚拟信息和现实物体相结合的显示系统进行分类，现实环境和虚拟环境位于分类标准的两端，在这之间更接近现实环境的是增强现实技术（Augmented Reality），位于中间的是混合现实技术（Mixed Reality），更接近虚拟环境的是虚拟现实技术（Virtual Reality）。

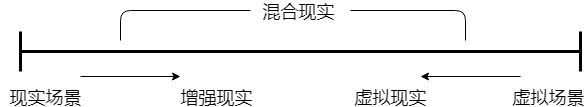


图1

另一种主流的定义是由R.T.Azuma于1997年在休斯实验室提出的，Ronald Azuma归纳了增强现实技术的三个特征：1）三维空间注册：增强现实应实时跟踪相机的位置并实时地计算出虚拟图像在真实场景中的注册位置；2）实时交互：增强现实应使用户在获得现实世界的信息的同时获得相应的虚拟反馈信息（如移动相机镜头，增强现实所生成的虚拟物体在显示画面中的相对位置也应实时发生改变）；3）虚实结合：增强现实技术依赖现实场景产生虚拟物体并使之与现实场景完美融合。

增强现实主要运用了实时跟踪注册技术、系统显示技术等。而实时跟踪注册技术是增强现实系统最关键最核心的技术。跟踪注册技术通过算法实时计算虚拟空间与现实空间坐标的映射关系，并使虚拟空间坐标与现实空间坐标精准对齐，实现虚拟信息在现实空间的叠加。该技术直接影响着增强现实的最终效果，实时计算虚拟空间与现实空间坐标的映射关系速度太慢，会导致效果输出的帧数过低，严重影响增强现实的效果。目前常用的跟踪注册技术主要是基于计算机视觉的，通过摄像机获取实时图像再进一步分析构建坐标映射，主要有两类：一类是基于标识物的方法，通过对摄像机获取的实时图像进行边缘检测识别标识物，从而以标识物的坐标为基础进一步建立虚拟空间坐标，最终建立坐标映射关系；另一类则是基于自然特征的方法，分析实时图像，提取相匹配的自然特征点，并以自然特征点的坐标为基础建立虚拟空间坐标，最终实现虚拟空间与现实空间的坐标映射。

随着相关技术的不断进步和相关设备的不断升级，增强现实的众多技术瓶颈不断被击破，用户体验获得较大提升，应用场景逐渐扩大。美国知名第三方调研机构 Manatt Digital Media 和Digi-Capital 投资银行发布一组数据指出，全球虚拟现实及增强现实市场规模将于2020年达到1500亿美元，可以说增强现实技术是信息科技产业未来发展的大趋势。虽然增强现实技术在硬件发展方面屡屡受挫（google公司于2015年1月19日停止了谷歌眼镜的“探索者”项目，英特尔公司于2018年4月18日正式停止Vaunt智能眼镜项目），但是随着越来越多的开发者进入AR领域，越来越多优质AR内容被产出，Vuforia、Metaio、ARToolKit等AR开发框架被发展和完善，AR与Web的结合，增强现实技术终将成为我们的未来。

**1.2 增强现实阅读系统需求分析**

互联网时代尤其是移动互联网时代的到来引起了信息接受方式的变革，互联网中海量的多媒体信息带给人们全新的接受信息的方式，给人们的阅读习惯和阅读偏好带来巨大改变。在这个信息爆炸的时代，生活节奏不断加快，人们的时间进一步碎片化，传统的文字加图片的阅读方式已经不能满足人们利用碎片化的空闲时间快速获取信息的需求。

相比于传统的阅读方式，人们更加倾向于选择多媒体的信息接收方式，而增强现实的阅读方式在多媒体的基础上提供给受众更生动和更“真实”的体验，增强受众的信息接收印象。这也是增强现实技术在广告营销、教育、游戏娱乐等领域越来越流行的原因。所以构建一个增强现实阅读系统，将文字、图片、视频等多媒体信息通过浏览器以三维立体的效果呈现给受众，符合受众期望获得更为“真实”、生动的信息的需求，具有非常实用的现实意义。同时对增强现实技术的发展和应用也是一次有价值的尝试和探索，具有一定的前瞻性研究意义。

**1.3 增强现实阅读系统可行性分析**

HTML5技术作为新一代Web开发标准，优越的跨平台性使其在技术运用上的限制变得更小，同时HTML5在数据存储、网络通信、音视频播放、网页绘图、地理定位等方面都有着自己的优势。与之前的版本相比，HTML5的新特性使得开发和使用网页更为简单便捷，还优化了浏览器对资源的占有率和对插件的依赖。基于HTML5构建一个增强现实的阅读系统，不仅开发时间短、成本低，还能通过浏览器直接使用，无需安装相关应用，便于使用和分享推广。

WebGL（Web图形库）是一种Web图形标准，用于在不使用插件的情况下在任何兼容的Web浏览器中呈现交互式2D和3D图形。WebGL可以为HTML5 Canvas提供硬件3D加速渲染，这样Web开发人员就可以借助系统显卡来在浏览器里更流畅地展示3D场景和模型了。而three.js是利用 JavaScript代码实现3D效果的WebGL第三方库，相当于浏览器的3D引擎，能够实现很多Web3D效果。这些为在浏览器中实现增强现实效果提供了3D图形基础。

WebRTC（Web实时通信）是一个支持浏览器实时获取设备视频流和语音流的开源技术，它使得网页在无需安装插件或者应用程序的前提下实现音频和视频的通信。所以WebRTC能够使网页实时获得摄像机的视频流，并在此基础上建立真实场景坐标系并计算虚拟物体的坐标等数据，再进一步将真实场景和虚拟物体相融合，实现增强现实效果。

所以我们能够基于Augmented Reality的开源HTML5和JavaScript代码，构建一个增强现实阅读系统，并使之能在任意一款支持WebGL和WebRTC的浏览器上运行，将文字、图片、视频等多媒体信息以三维立体的效果传递给受众。

**1.4 小结**

增强现实技术能够以一种更为直观高效、生动有趣的方式将虚实结合的信息通过显示设备传递给用户，使用户与计算机之间的交互更为自然，产生更为真实的体验。在信息爆炸的新时代，增强现实阅读系统能够更好地满足受众的新需求，提供更为真实生动的信息接收体验。而新的技术标准的不断发展（如WebGL和WebRTC）使得构建一个基于HTML5的轻量级增强现实阅读系统成为可能。

**2 增强现实系统相关研究进展**

**（尤其是关于阅读的增强现实系统，总结已有系统的差异）**

**2.1 增强现实阅读系统研究现状**

虽然增强现实技术的应用广泛，但大部分都是处于试验探索阶段，离消费级应用都还有一定的距离。其中较为成熟且形成一定规模的AR应用主要集中在图书出版、教育和游戏等领域。所以我将重点从这三个领域对增强现实阅读系统的研究现状进行梳理。

在图书出版领域，早在2008年德国的ARS Edition公司就推出了无需基于特征标识物直接识别图像的AR图书《Aliens&UFOs》，此后AR图书越来越频繁地出现在我们眼前。目前来说实现增强现实效果的图书大都采取相同的方法：图书出版商在图书上印上能够被算法识别的图像标记，并与提前准备好的3D模型内容相链接，当读者通过手机或者其他设备的摄像头对图书进行扫描时，图像标记被识别出来，然后链接到提前准备好的3D模型内容，读者获得增强现实的效果体验。

在教育领域，增强现实技术主要被用来增加趣味性和增加空间立体感知。如为儿童设计制作的增强现实单词卡、动物卡等，还有利用AR技术立体地展示电波、磁场、原子、几何等那些抽象或肉眼不可见的学习内容的课件。这些往往也是通过移动设备扫描卡片呈现对应的3D模型来实现的。

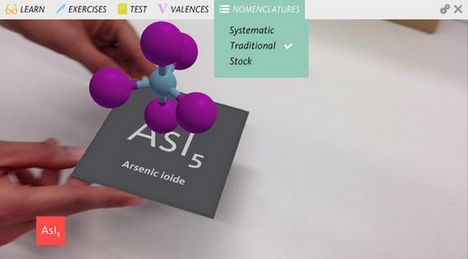
 

图 AR Flashcards-animal Alphabet展示和AR展示化学元素

在游戏领域，《Pokemon Go》2016年的火爆让全世界看到了AR游戏的魅力。AR游戏由于其需要与玩家进行更广泛的互动，需要设备获取更多的信息，所以更多地会结合LBS技术（Location Based Service基于位置服务）和陀螺仪（又叫角速度传感器）技术。

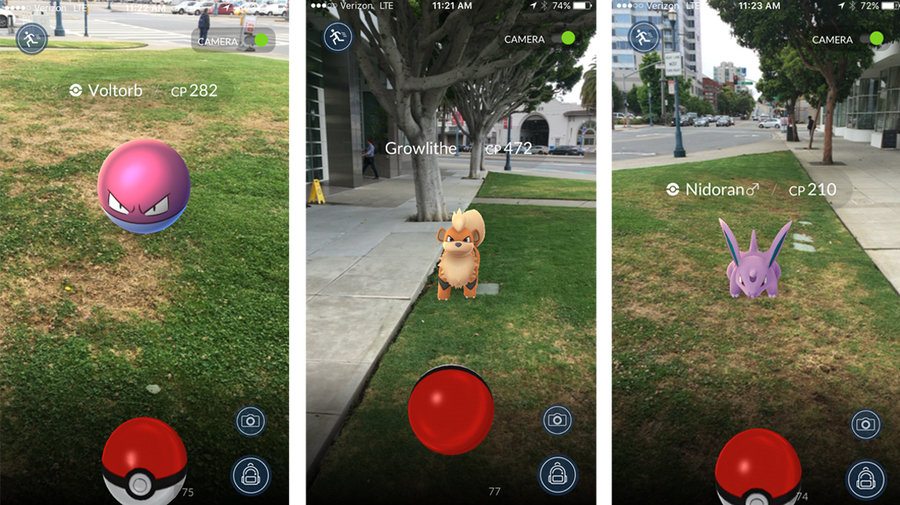


图 Pokemon Go游戏截图

在医疗领域，增强现实技术常作为一种可视化的辅助工具，如将诊断时在病人的身体部位提供对应的内在医学数据、在手术时指导医生进行手术操作等。但由于医学应用的严谨性，增强现实技术在医疗领域的应用还需要进一步实验优化完善。在制造业领域，增强现实技术可以辅助技术人员对大型复杂机械进行更高效的组装、检修和维护。但由于大规模使用需要一定的硬件和软件成本，使得增强现实技术在制造业的推广使用受到一定限制。

**2.2 增强现实阅读系统差异分析**

一般而言，大部分实现增强现实效果的应用（如图书出版的AR应用和教育领域的AR应用）都比较强调虚拟物体定位的精确性，所以一般采用基于标识物的跟踪注册技术。而AR游戏更侧重于游戏与玩家的互动性，所以对使用灵活性有一定的要求，一般采用的是算法更复杂、但使用更加灵活自由的基于自然特征的跟踪注册技术。

对比分析增强现实技术在各领域的应用，可以发现增强现实技术的应用主要起到提供辅助信息的作用，对提高人们的学习工作效率、提升生活品质有着较大帮助。

**3 增强现实阅读系统概要设计**

**这个系统的顶层设计，可以画一个数据流程图，关于这个系统的功能、操作流程内容。二级标题可以在后边写论文和代码的时候再确定**

**3.1系统顶层设计**

Bimber和Raskar在《Spatial Augmented Reality Merging Real and Virtual Worlds》[[1]](#footnote-1)一书中描述了如下图所示的增强现实系统的基本架构。总结分析对比可以发现目前绝大部分的增强现实应用都符合该基本架构，故该基本架构对本系统也有很大的参考价值。

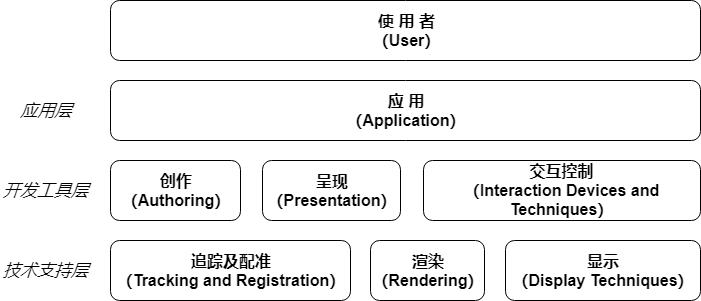
****

图 AR系统的基本架构

**3.1.1系统业务流程**

本系统是基于HTML5构建的增强现实阅读系统，故采用B/S结构（Browser/Server，浏览器/服务器模式）,整个系统作为网站页面文件部署在服务器上。用户利用支持WebGL和WebRTC的浏览器通过网页链接访问本系统的应用页面，页面首先会调用设备的摄像头并获取视频流，然后对视频流中的帧进行标识物检测。若检测出标识物，浏览器便会向服务器请求调用与该标识物对应的3D模型数据文件。最后浏览器将3D模型与基于标识物建立的空间坐标系联系起来，实现增强现实的视觉效果。根据以上描述，可绘制本系统业务流程图，以梳理该系统业务逻辑。

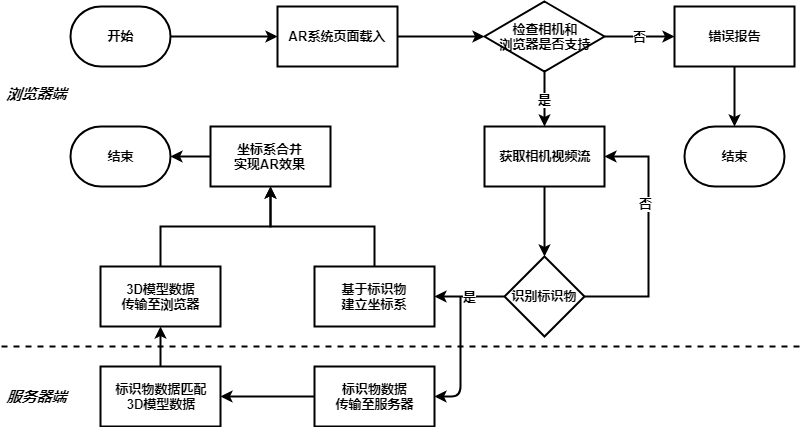


图 系统业务流程图

**3.1.2系统数据流程**

数据流图（Data Flow Diagram，简称DFD）是分析结构化系统的主要工具，它从数据传递和加工角度，以图形方式来表达系统的逻辑功能、数据在系统内部的逻辑流向和逻辑变换过程，是结构化系统分析方法的主要表达工具及用于表示软件模型的一种图示方法。

绘制DFD数据流程图，通过自顶向下、逐层分解的方法，利用分解和抽象这两个基本手段控制系统复杂性，把大问题分解成小问题，然后分别解决。分层绘制DFD数据流程图，能够清晰地分析、调整系统的各个功能模块，最终完善整个系统。

（1）顶层DFD

顶层DFD概括地反映了系统最主要的逻辑功能、最主要的外部实体、数据存储和数据处理过程。本系统的外部数据来源为提前准备好的标识物，经过整个系统识别、注册、渲染等处理后最终在设备屏幕上呈现AR效果。

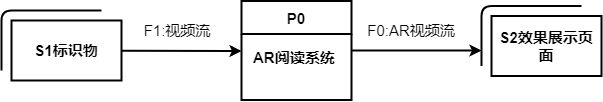


图 系统顶层DFD

（2）第一层DFD

第一层DFD是在系统顶层DFD的基础上将AR阅读系统进行分解，主要将整个系统分为视频采集、识别注册、AR效果合成三个数据处理过程。S1标识物被设备的摄像头拍摄采集，经过P1视频采集过程进入系统。随后P2识别注册过程利用数字图像处理算法，从P1传来的视频数据中识别出标识物的特征信息。确定了标识物后，AR系统便利用相关的矩阵变换算法计算得出真实场景中摄像机相对于标识物的位置和角度，求出摄像机与标识物的相对位置矩阵。之后P2识别注册模块便在虚拟世界中注册一个与现实相机位置和角度一样的虚拟相机，再结合摄像机与标识物的相对位置矩阵确定对应的虚拟物体的位置坐标，实现跟踪注册。最后通过OpenGL等图形渲染技术对3D场景进行渲染。

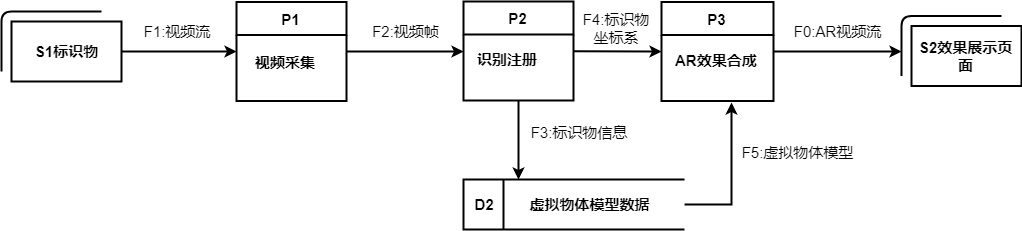


图 第一层DFD

（3）第二层DFD

第二层DFD是对第一层DFD中的视频采集、识别注册、AR效果合成三个数据处理过程分别进一步分解得来的，通过分析第二层DFD能够更清晰地了解三个模块的处理细节。

P1视频采集过程主要是摄像机将视频流暂存为视频数据，然后提取其中的视频帧供系统识别标识物，流程如图。

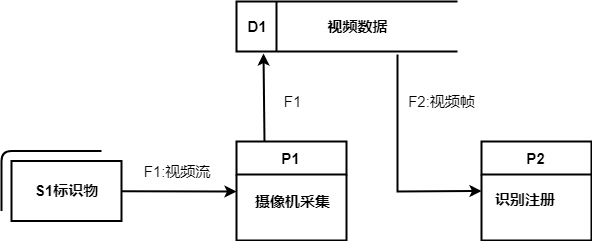


图 P1视频采集第二层DFD

而P2识别注册过程是从视频帧中识别标识物，然后从服务器中调用虚拟物体模型数据，同时基于标识物建立空间坐标系，最后将以上两部分数据传递给P3，流程如图。

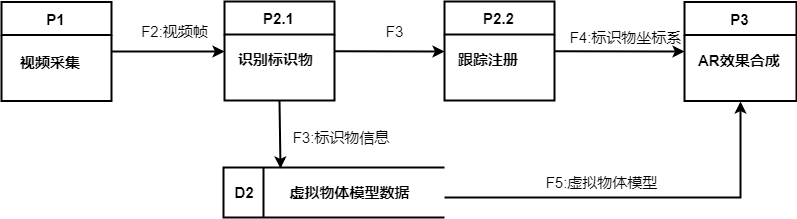


图 P2识别注册第二层DFD

P3 AR效果合成处理过程收到从P2传来的两部分数据，将两个部分的坐标系合并后通过WebGL技术生成AR立体图像并传递给显示页面，具体流程如图。

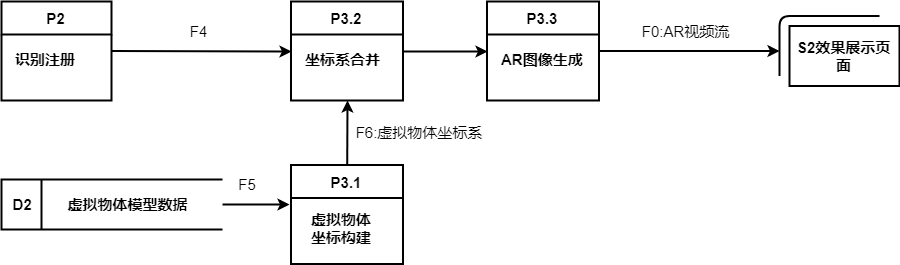


图 P3AR效果合成第二层DFD

**3.2系统功能设计**

本系统主要利用增强现实技术，以一种更为直观高效、生动有趣的方式将虚实结合的信息通过显示设备传递给用户，为用户提供更真实生动的体验。主要有以下功能：

a）平面媒体上动态构建虚拟3D模型

可以将系统所能识别的标识物提前印在平面媒体上，然后通过浏览器打开本系统页面，调用摄像机便能识别标识物并以标识物为基础生成一个3D模型场景，最后展现在设备屏幕上，实现增强现实的效果。

b）锁定摄像头画面固定生成3D模型

增强现实功能需要使用摄像机，且该增强现实阅读系统的使用场景大多数是基于智能手机移动端。而移动端设备一般是由使用者手持，在使用过程中不可避免地会出现设备晃动导致相机画面晃动，影响增强现实效果和使用体验。因此可以在上述a功能的基础上锁定摄像头画面，在固定画面上生成3D模型，展现固定的增强现实场景，便于使用者进一步观察3D模型。

c）增强现实视频

能够基于标识物位置在需要的虚拟空间位置播放视频文件，让用户轻松体验从不同角度观看视频。

**3.3操作流程**

本系统为基于HTML5的增强现实阅读系统，故需要使用浏览器调用系统html页面文件，大致操作流程为：

1）用户打开浏览器并输入本系统页面网址。

2）按照提示打开摄像机，将摄像机对准提前准备好的标识物。

3）观察屏幕，调整相机角度，等待虚拟物体加载。

4）若等待较长时间并未加载出任何虚拟物体，建议重新调整相机角度或者调整标识物附近光照情况。

5）若用户需要体验上述功能b，可点击屏幕右上角的切换模式按钮，摄像头画面将被固定，用户便可仔细观察虚拟物体模型。

6）使用结束，关闭系统网页页面。

**4 增强现实阅读系统详细设计**

**这个系统使用的HTML5技术、js技术等，增强现实技术涉及的各个要素等内容，二级标题可以自己后面再写论文的时候再细化**

**4.1 HTML5技术细节**

**4.1.1 WebRTC**

WebRTC（Web Real-Time Communication，网页实时通信）是一个支持网页浏览器进行实时语音对话和视频对话的技术。Google于2011年5月开放了其工程源代码，使其得到了广泛的支持和应用，也使其成为HTML5的技术标准之一。

WebRTC提供了浏览器进行视频对话所需要的核心技术，包括音视频的采集、编解码、网络传输、显示等功能。其中与本系统密切相关的技术为视频采集video\_capture、视频图像处理video\_processing和视频显示video\_render技术。视频采集video\_capture技术能够调用设备摄像头并采集视频数据，视频图像处理video\_processing技术则可以对获取的视频图像按帧进行处理，利用图像匹配技术实现对标识物的识别检测。视频显示video\_render技术则可以对每一帧进行重新渲染，在已有视频图像的基础上渲染加载虚拟物体。

**4.1.2 WebGL**

WebGL（Web Graphics Library）是一种Web端的3D绘图技术，将JavaScript和OpenGL[[2]](#footnote-2)结合起来，为Web页面进行复杂3D渲染提供支持。WebGL可以为HTML5的Canvas画布提供硬件3D渲染加速，免去了使用网页3D渲染插件的麻烦，同时其基于OpenGL的特性又带来了使用的便利性和跨平台性。利用WebGL的API接口，可以将需要渲染的3D物体快速加载在页面画面中。

**4.2 JavaScript技术细节**

**4.2.1 Three.js**

Three.js是利用JavaScript编写的WebGL第三方库，相当于是运行在浏览器中的3D引擎。利用Three.js可以创建各种复杂三维场景，其内部封装了摄像机、物体材质、光影效果等各种对象，可以方便快捷地在页面画面中渲染加载需要的3D模型。

使用Three.js的方法很简单，只需要在html文件中引用three.js即可。但是要将物体渲染到网页页面中，有三个组件是不可或缺的：场景（scene）、相机（camera）和渲染器（renderer），具体代码将在5.1.1中阐述。

**4.2.2 AR.js**

AR.js是GitHub上的一个开源webAR项目，能够在Web端实现高效的AR效果。ARToolkit是一个基于C/C++的开源AR开发框架，能够快速构建增强现实应用。而AR.js就是基于ARToolkit开发而来的，通过Emscripten编译器将ARToolkit中的C/C++代码编译成asm.js[[3]](#footnote-3)，形成了一个基于JavaScript的AR开发框架，为Web端的AR应用提供各种关键功能。

**4.3增强现实技术细节**

**4.3.1 跟踪注册技术**

跟踪注册技术是增强现实系统中的关键技术。跟踪注册技术通过算法实时计算虚拟空间与现实空间坐标的映射关系，并使虚拟空间坐标与现实空间坐标相互结合，实现虚拟物体在现实空间的叠加。跟踪注册技术主要有两种，一种是基于硬件传感器的跟踪注册技术，这种跟踪注册技术需要利用相关硬件设备来测量获取相应的运动状态数据和位置数据，然后基于这些数据计算出AR设备相对于现实环境的空间坐标和运动状态。如利用GPS系统可以获取AR设备的大致位置，提供基于位置信息的服务，利用设备中的陀螺仪、重力感应器等传感器可以获取设备的运动状态。

另一种是基于计算机视觉的跟踪注册技术，这种跟踪注册技术主要是利用摄像头获取现实场景的实时图像，然后利用算法分析获取的实时图像，从中构建现实环境坐标和虚拟空间坐标的映射关系。而基于计算机视觉的跟踪注册技术大致又可以分为两类。一类是基于标识物的方法，通过对摄像机获取的实时图像进行边缘检测识别标识物，从而以标识物的坐标为基础建立虚拟物体坐标，最终建立与现实空间坐标的映射关系；另一类则是基于自然特征的方法，分析实时图像，提取相匹配的自然特征点，并以自然特征点的坐标为基础建立虚拟空间坐标，最终实现虚拟空间与现实空间的坐标映射。

基于计算机视觉的跟踪注册技术和基于硬件设备的跟踪注册技术各有侧重，前者能够更精确地建立虚拟物体与真实环境的坐标映射，后者能够更快地计算出设备的运动状态，提供更多交互手段。随着增强现实技术的不断进步优化，目前已经能够根据功能需要将两种跟踪注册技术结合在一起，实现更为及时精确的跟踪注册。本系统主要功能是提供信息展示，对精确性有较大需求，对交互性没有很大需求，为了降低系统复杂性提高可用性，故主要采用基于标识物的计算机视觉跟踪注册技术。

**4.3.2显示技术**

增强现实显示技术也是增强现实系统的重要技术之一，显示技术能够直接影响用户使用体验，它还影响着系统的交互性。目前增强现实显示技术一般按显示设备分为：屏幕显示技术、头戴式显示技术、投影显示技术。

屏幕显示技术是目前实现最简单也是最可靠的增强现实显示技术。增强现实系统在实现虚拟物体的渲染后可以直接通过屏幕将效果呈现出来，虽然其提供的沉浸感较差，但不需要依靠额外的显示设备，使用简单，不会产生3D眩晕等不舒适感。

头戴式显示技术是增强现实最理想的呈现方式，目前一般是利用光的反射原理通过半透明的光学器件将虚拟信息传入用户眼中，同时真实世界的信息也会通过半透明的光学器件传入用户眼中，从而实现虚拟世界与现实世界的融合，能够让用户感受到沉浸式的体验。但是由于目前相关技术瓶颈尚未被突破，如快速运动时画面的实时处理问题，头戴式显示器可能会让用户产生不适感，目前还没有得到规模化商用。

投影显示技术是将生成的虚拟物体信息投影到现实场景中从而增强现实的一种显示技术。投影显示技术的优点主要是可以应用于大范围场景中，但是其图像投影焦点不会随着用户视角移动而改变，也需要额外的设备支持，所以广泛应用受到一定限制，比较适合工厂、商场等应用场景。

对比分析可以发现屏幕显示技术更适合本增强现实系统，其不要额外的设备支持，设备成本低，用户通过自己的智能手机便能使用，使用场景也很契合。

**5 增强现实阅读系统实施**

**可以放一些系统的截图、核心代码、具体操作流程解释等。二级标题可以自己后面再写论文的时候再细化**

**5.1 系统实施概述**

**5.2 具体操作流程**

**5.2.1 three.js**

var scene = new THREE.Scene(); //创建场景

var camera = new THREE.Camera(); //创建相机

var renderer = new THREE.WebGLRenderer(); //创建渲染器

**5.2.2 AR.js**

**5.2.3 标识物匹配训练**

**5.2.4 虚拟物体模型**

**6 增强现实阅读系统评估**

**系统的功能特点、优势、劣势，可用性、易用性等评价内容**

**6.1需求评估**

**6.2 功能性测试**

**6.3 非功能性测试**

标记识别黑边，尝试变细，识别效果大大降低。

**7 总结**

1. 《空间增强现实：融合现实与虚拟世界》Bimber. O.& Raskar. R. Spatial Augmented Reality Merging Real and Virtual Worlds[M]. August 08, 2005,A.K. Peters Ltd/CRC Press. [↑](#footnote-ref-1)
2. OpenGL（Open Graphics Library）是一个跨编程语言、跨平台的专业图形程序接口。它用于三维和二维图像，是一个功能强大，调用方便的底层图形库。 [↑](#footnote-ref-2)
3. asm.js是JavaScript语言的一种变体，它的变量一律都是静态类型，并且取消垃圾回收机制，除了这两点asm.js与JavaScript并无差异，可以说asm.js是JavaScript的一个严格子集，只能使用JavaScript的一部分语法。 [↑](#footnote-ref-3)