

# 对AR的进一步思考

李坤桐

大体框架与演讲相同，但补充了许多内容，并且思考的也更加深入；同时针对老师的建议，搜集了一些国内有关AR技术的资料，并进一步思考。

## 一. AR的基本定义

增强现实（Augmented Reality）是指透过摄影机影像的位置及角度精算并加上图像分析技术，让屏幕上的虚拟世界能够与现实世界场景进行结合与交互的技术。

下面重点区分一下vr、ar、mr、xr、cr这几个概念。

Augmented Reality(AR)，增强现实是一种对真实世界物理环境的实时、直接或间接观察，其中的元素通过计算机生成的感官输入(如声音、视频、图形或GPS数据)得到增强(或补充)。由于AR存在于我们自己的世界之上，它提供了与你在正常生活中所获得的一样的自由。增强现实利用你现有的现实，并利用某种设备加以补充。手机和平板电脑是目前最流行的增强现实媒体，通过摄像头，应用程序将数字内容叠加到环境中。

Virtual Reality(VR)，虚拟现实是一种身临其境的体验，也称为计算机模拟现实。它指的是计算机技术使用现实头盔来产生真实的声音、图像和其他感觉，复制一个真实的环境或创造一个想象的世界。VR是一种让用户沉浸在完全虚拟世界中的方式。一个真正的VR环境将涉及所有五种感官(味觉、视觉、嗅觉、触觉、听觉)。经过多年在游戏行业的流行，我们现在看到这项技术进入更实际的应用，市场和行业仍然对这一技术趋势感到兴奋，预计在不久的将来会有进一步的进展。

Mixed Reality(MR)，混合现实，有时被称为混合现实，是真实世界和虚拟世界的结合，以产生新的环境和可视化，其中物理和数字对象共存并实时交互。这意味着将新图像放置在真实空间中，使新图像能够在一定程度上与我们所知的物理世界中的真实事物互动。MR的主要特点是合成内容与真实内容能够实时反应。可能和AR容易混淆，区分它们的一个技巧是虚拟物体的相对位置，是否随设备的移动而移动。

Cinematic Reality（CR），影像现实，这是Magic Leap提出的一个概念，从字面上看就可以看出，其主要是强调要使虚拟场景达到与影像生动逼真。主要是通过光波传导棱镜的设计，多角度将画面直接呈现到我们的视网膜上，以尝试解决视野太窄及眩晕等问题，

实际理念与MR是相似的。

Extended Reality(XR)，扩展现实，是一个新添加到字典的技术词汇。扩展现实是指所有由计算机技术和可穿戴设备产生的真实与虚拟结合的环境和人机交互。扩展现实包括AR、VR、MR等所有描述形式。换句话说，XR可以被定义为一把雨伞，它将所有三种现实(AR、VR、MR)集中在一个术语下，从而减少了公众的困惑。扩展现实在部分传感器输入的虚拟性方面为沉浸式虚拟性提供了广泛的种类和大量的级别。

此外，AR还容易与全息投影、裸眼3D等混淆，区别主要体现在观看方式上，全息投影和裸眼3D不需要借助辅助设备就能直接观看，而AR所展示出的效果则需要借助辅助设备才能实现顺利观看。

## 二. AR的工作原理

### 2.1 AR技术原理

实现 AR 需要识别、追踪和渲染三步，A从其技术手段和表现形式上，可以分为大约两类：Vision based AR，即基于计算机视觉的 AR和LBS based AR，即基于地理位置信息的AR。

#### 2.1.1 Vision based AR

基于计算机视觉的 AR 是利用计算机视觉方法建立现实世界与屏幕之间的映射关系，使我们想要绘制的图形或是 3D 模型可以如同依附在现实物体上一般展现在屏幕上。本质上来讲就是要找到现实场景中的一个依附平面，然后再将这个 3 维场景下的平面映射到我们 2 维屏幕上，然后再在这个平面上绘制你想要展现的图形，从技术实现手段上又可以分为 Marker-Based AR 和Marker-Less AR 2两类。

##### (1) Marker-Based AR

这种实现方法需要一个事先制作好的 Marker，然后把 Marker 放到现实中的一个位置上，相当于确定了一个现实场景中的平面，然后通过摄像头对 Marker 进行识别和姿态评估，并确定其位置，然后将该 Marker中心为原点的坐标系称为 Marker Coordinates 即模板坐标系，我们要做的事情实际上是要得到一个变换从而使模板坐标系和屏幕坐标系建立映射关系，这样我们根据这个变换在屏幕上画出的图形就可以达到该图形依附在Marker上的效果。从模板坐标系变换到真实的屏幕坐标系需要先旋转平移到摄像机坐标系然后再从摄像机坐标系映射到屏幕坐标系。

在实际的编码中，所有变换都是一个矩阵，在线性代数中矩阵代表一个变换，对坐标进行矩阵左乘便是一个线性变换。公式如下：

$$h \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ 0 & C_{22} & C_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_1 \\ R_{21} & R_{22} & R_{23} & T_2 \\ R_{31} & R_{32} & R_{33} & T_3 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} = C T_m \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix}$$

图2-1 线性变换公式

矩阵C的学名叫摄像机内参矩阵，矩阵Tm叫摄像机外参矩阵，其中内参矩阵是需要事先进行摄像机标定得到的，而外参矩阵是未知的，需要我们根据屏幕坐标和事先定义好的 Marker 坐标系以及内参矩阵来估计Tm，然后绘制图形的时候根据Tm来绘制。

## (2) Marker-Less AR

基本原理与Marker based AR 相同，不过它可以用任何具有足够特征点的物体(例如：书的封面)作为平面基准，而不需要事先制作特殊的模板，摆脱了模板对AR应用的束缚。它的原理是通过一系列算法(如：SURF，ORB，FERN 等)对模板物体提取特征点，并记录或者学习这些特征点。当摄像头扫描周围场景，会提取周围场景的特征点并与记录的模板物体的特征点进行比对，如果扫描到的特征点和模板特征点匹配数量超过阈值，则认为扫描到该模板，然后根据对应的特征点坐标估计Tm矩阵，之后再根据Tm进行图形绘制。

### 2.1.2 LBS-Based AR

这种 AR 技术利用设备的 GPS 功能及传感器来实现，摆脱了应用对 Marker 的依赖，用户体验方面要比 Marker-Based AR 更好，而且由于不用实时识别 Marker 姿态和计算特征点，性能方面也好于 Marker-Based AR和Marker-Less AR，因此对比 Marker-Based AR 和 Marker-Less AR，LBS-Based AR 可以更好的应用到移动设备上。

其基本原理是通过 GPS 获取用户的地理位置，然后从某些数据源（比如 wiki，google）等处获取该位置附近物体的 POI 信息，再通过移动设备的电子指南针和加速度传感器获取用户手持设备的方向和倾斜角度，通过这些信息建立目标物体在现实场景中的平面基准（相当于marker），之后坐标变换显示等的原理与 Marker-Based AR 类似。

LBS+AR 就是融合了基于地理位置和增强现实，此前其应用在游戏之中，其中例如去年火遍全球的《Pokemon Go》，正是这一应用的最佳代表。游戏在定位玩家的地理

位置后，系统设定分布在该地域的妖怪品种以及出现几率，玩家跟着导航就能找到各种口袋妖怪，并且游戏中还运用AR技术，让玩家捕获妖怪的扔球动作原汁原味再现于现实。

LBS-Based AR 现在也广泛应用在导航类应用，不仅能根据用户的位置变化动态地提供各种用户所需的空間信息,而且还可以将这种空間信息通过虚拟三维的方式动态的叠加在用户的视频信息上,增加了用户与系统的交互性,使得用户界面更加智能化,这是未来空間信息服务和移动定位服务相结合的一种新方式。

## 2.2 主要硬件技术

### 2.2.1 交互技术

手势操控：构建指尖、指向以及手掌平面等基本操作型手势交互的三维交互模式,进而将从真实世界提取的三维交互信息对齐到虚拟世界,融合渲染并触发交互,从而实现空间一致的操作型手势自然。微软HoloLens是利用手势进行交互的、最有特点的AR硬件。戴上Holo Lens眼镜后，可通过手指在空中点选、拖动、拉伸来控制虚拟物体、功能菜单界面。比如利用Air tap手势打开全息图，利用Bloom 手势打开开始菜单。

语音操控：手势操控固然解放了双手，但是它也有缺陷，那就是频繁的抬手会造成手臂酸软。而语音操控便是更好的人机交互方案。现在微软Cortana、Google Now、苹果Siri、亚马逊Echo都是优秀的语音识别助手，但是他们的识别率还是不高，只能作为辅助操作工具，智能程度也远远达不到AR交互需求。

体感操控：假设有一天全息通话成为了现实，那么除了语音、视觉交流之外，想要获得更加完美的增强现实体验，体感外设显然是非常重要的一环。现在，已经有不少厂商推出了体感手套、体感枪等外设。只是这些设备功能还很单薄，还有着极大的改进空间。

### 2.2.2 镜片成像技术

无论是增强现实还是虚拟现实，FOV 都是影响使用体验的最重要因素之一。现在的AR眼镜的可视广角普遍不高，HoloLens有30°，Meta One只有23°，而公众最为熟悉的Google Glass视角仅有12°。这是由于镜片成像技术和光学模组不成熟造成的，现在还没有太好的解决方案，但太窄的视角显然让增强现实效果大打折扣。

光学镜片还是存在着色散和图形畸变的问题。智能眼镜成像时，视场周边会出现红绿蓝色变，这就是棱镜反射光线时常见的色散现象，可以通过软件进行色彩补偿或者通过多材料镜片来消除。前者会增加硬件负担并降低图像帧率。后者的成品率低，这也是造成AR眼睛昂贵的原因之一。

### 2.2.3 SLAM技术

SLAM 即指同步定位与建图技术。几年前，扫地机是就是它的代言人。扫描室内布局结构，并构建、规划扫地路线的扫地机器人是SLAM技术的最好代表。其实，这项技术也可以被运用在AR领域，现阶段基于SLAM技术开发的代表性产品就有微软Hololens，谷歌Project Tango以及Magic Leap。

以HoloLens为例，它在启动的时候，会对用户所处空间进行扫描，从而建立房间内物体摆设的立体模型。这就需要SLAM技术。

## 2.3 AR SDK

AR SDK是AR应用的技术核心，驱动着AR应用的发展和创新。AR SDK的作用是将数字内容和信息与现实世界融合。SDK的功能最终将决定AR应用程序中的特性和功能，因此根据项目要求选择正确的平台至关重要。AR SDK包含许多应用程序组件，像内容呈现，AR跟踪和场景识别等。内容呈现意味着可以将数字信息和3D对象覆盖在现实世界物体之上，跟踪表示“应用程序的眼睛”，场景识别充当应用程序的中枢神经系统。每个AR SDK都配备自己独特的属性，使AR开发人员能够以最佳方式开发出具有识别，渲染和跟踪功能的AR应用程序。常用的SDK有苹果推出的ARKit、谷歌推出的ARCore、高通推出的Vuforia以及视辰信息科技（上海）有限公司的增强现实解决方案系列的子品牌EasyAR等等，其中最广为人知的就是ARKit和ARCore。

### 2.3.1 ARKit

ARKit 是一个移动端AR平台，用于在 iOS 上开发增强现实APP。可以在目前的数千万台 iOS 设备上运行。但为了获得 ARKit 的完整功能，需要 A9 及以上芯片。

ARKit提供了以下功能：SLAM tracking and sensor fusion(SLAM场景识别)、Ambient lighting estimation(环境光评估)、Scale estimations(刻度估量)、Vertical and horizontal plane estimation with basic boundaries(基本边界的垂直和水平面检测)、Stable and fast motion tracking(快速稳定的动作捕捉)、Multiple Face Tracking(多重人脸追踪)、Collaborative Sessions(多人会话)、Simultaneous Front and Back Camera(同步前置、后置摄像头)等。

### 2.3.2 ARCore

ARCore是Google专有的增强现实SDK。它使开发人员能够在部分支持ARCore的智能手机和平板电脑上启动并运行AR应用。ARCore最值得注意的功能之一是它还支持部分iOS的设备。ARCore拥有多个重要功能：

**光估测：**ARCore 可以检测其环境光线的相关信息，并提供给定摄像头图像的平均光强度和色彩校正。此信息能够使用与周围环境相同的光照来照亮虚拟物体，提升它们的真实感。

**环境理解：**ARCore 会通过检测特征点和平面来不断改进它对现实世界环境的理解。ARCore 可以查找看起来位于常见水平或垂直表面上的成簇特征点，并让这些表面可以由应用用作平面。ARCore 也可以确定每个平面的边界，并将该信息提供给应用。由于 ARCore 使用特征点来检测平面，因此可能无法正确检测像白墙一样没有纹理的平坦表面。

**运动跟踪：**当手机在现实世界中移动时，ARCore 会通过一个名为并行测距与映射的过程来理解手机相对于周围世界的位置。ARCore 会检测捕获的摄像头图像中的视觉差异特征，并使用这些点来计算其位置变化。这些视觉信息将与设备 IMU 的惯性测量结果结合，一起用于估测摄像头随着时间推移而相对于周围世界的姿态。

**用户交互：**ARCore 利用命中测试来获取对应于手机屏幕的 (x,y) 坐标，并将一条射线投影到摄像头的视野中，返回这条射线贯穿的任何平面或特征点以及交叉位置在现实世界空间中的姿态。这让用户可以选择环境中的物体或者与它们互动。

**共享：**借助 ARCore Cloud Anchor API，可以创建适用于 Android 和 iOS 设备的协作性或多人游戏应用。使用云锚点，一台设备可以将锚点和附近的特征点发送到云端进行托管。可以将这些锚点与同一环境中 Android 或 iOS 设备上的其他用户共享。这使应用可以渲染连接到这些锚点的相同 3D 对象，从而让用户能够同步拥有相同的 AR 体验。

**定向点：**借助定向点，可以将虚拟物体置于倾斜的表面上。当执行会返回特征点的命中测试时，ARCore 将查看附近的特征点并使用这些特征点估算表面在给定特征点处的角度。然后，ARCore 会返回一个将该角度考虑在内的姿态。

**锚点和可跟踪对象：**姿态会随着 ARCore 改进它对自身位置和环境的理解而变化。当放置一个虚拟物体时，需要定义一个锚点来确保 ARCore 可以跟踪物体随时间推移的位置。

**增强图像：**使用增强图像，可以构建能够响应特定 2D 图像的 AR 应用。用户可以在将手机的摄像头对准特定图像时触发 AR 体验，例如，他们可以将手机的摄像头对准电影海报，使人物弹出，然后引发一个场景。

从本质上讲，ARCore 在做两件事：在移动设备移动时跟踪它的位置和构建自己对现实世界的理解。

ARCore 的运动跟踪技术使用手机摄像头标识兴趣点（称为特征点），并跟踪这些点随着时间变化的移动。将这些点的移动与手机惯性传感器的读数组合，ARCore可以在手机移动时确定它的位置和屏幕方向。

除了标识关键点外，ARCore 还会检测平坦的表面，并估测周围区域的平均光照强度。这些功能共同让 ARCore 可以构建自己对周围世界的理解。

## 2.4 WebAR

在web端，无论采用哪种实现手段，最主要是运行环境对webRTC、webGL两个特性的支持程度。并且，web端目前只有基于标记图Marker的应用有着比较良好的体验，WebAROnARCore 和 WebAROnARKit所支持的MarkerLess方案都还仅仅是测试阶段，并且国内外对webAR 的讨论和demo案例目前都比较少，总的来说使用WebAR来实现一些简单的功能是可以的。

### 2.4.1 AR.js

AR.js是由Jerome Etienne开发维护的一款WebAR库，代码托管在GitHub。其实，AR.js就是将WebAR所需的各种组件与库进行整理、维护，可以流畅地以高帧率运行在各种设备上。其中，AR.js主要封装了以下几个库：

**JSARToolKit:** 老牌开源AR库ARToolKit的JavaScript版本,主要提供了识别和追踪marker的功能。

**Three.js:**Three.js是一款开源的主流3D绘图JS引擎，我们知道WebGL是一种网页3D绘图标准，和jQuery简化了HTML DOM操作一样，Three.js可以简化WebGL编程。

**aframe-ar.js:**A-Frame团队发布的库，其中包括了一些 Web AR 组件，用 A-Frame 的各种组件可以让你用很少的代码构建出 AR 所需要的 3D 立体世界。A-Frame.js是一个用来构建虚拟现实（VR）应用的网页开发框架（部分API也可以开发AR应用），是当下用来开发WebVR内容主流技术方案。A-Frame基于HTML，容易上手。但是A-Frame不仅仅是一个3D场景渲染引擎或者一个标记语言。其核心思想是基于Three.js来提供一个声明式、可扩展以及组件化的编程结构。

在AR.js中，分为两种写法：使用Three.js结合JS库以及使用A-Frame.js，若选择使用Three.js结合JS库的方式，需要学习了解很多JS库的用法，代码量比较大，学习成本较高；而使用A-Frame.js，则可以大量减少代码量，比如著名的10行代码实现AR。

### 2.4.2 WebARonARKit, WebARonARCore

谷歌的AR团队提供了 WebARonARKit 和 WebARonARCore 两个库，以便开发者能用 Web 技术来基于 ARKit 和 ARCore 开发，提供的功能也类似：运动追踪、环境感知和光线感应，从而实现 WebAR。这两个库虽然功能强大，但因为是ARKit和ARCore的Web衍生版本，所以国内支持这两个库的移动设备仍然很少。

#### 2. 4. 3 EasyAR

EasyAR是国内首款免费全平台AR引擎，提供了EasyAR SDK以及WebAR等解决方案，可以快速构建AR应用。

### 三. AR现状及未来发展趋势

目前在国内外，AR技术都已经开始运用到了各个领域，例如，在医疗领域，作为工具在诊疗过程和日常工作上帮助医生解决问题，如利用AR技术轻易地进行手术部位的精确定位，用于物理治疗及恐惧症的治疗（如恐高症等），通过虚拟网络使患者更易就诊；在教育领域，变革学生在基础教育和高等教育阶段的受教方式，教师可以利用虚拟现实或增强现实技术让学生们在3D环境中与物体进行互动，提高学生的3D几何思维能力，Google为学校免费提供Cardboard来推进这一市场，目前，其已开展了逾百次“模拟实地考察”；在网络视频通讯领域，使用增强现实和人脸跟踪技术，在通话的同时在通话者的面部实时叠加一些如帽子、眼镜等虚拟物体，在很大程度上提高了视频对话的趣味性；在娱乐、游戏领域及旅游业等也有显著的应用。

AR的显示技术目前越来越离不开头盔显示技术，目前的增强现实系统主要使用透视式头盔显示器。透视式头盔显示器分为两种：视频透视式头盔显示器和光学透视式头盔显示器，各有优缺点。

一体式解决方案是目前最前沿的AR技术，最具代表性主要是微软的HoloLens头显，被称为AR行业之光的标杆性产品，是首个不受线缆限制的全息计算机设备，能让用户与数字内容交互，并与周围真实环境中的全息影像互动，2019年初发布了第二代HoloLens，包括投射新闻信息流、模拟游戏、收看视频和查看天气、辅助3D建模等功能，但由于续航问题、应用有限、价格昂贵的问题没法满足普通消费者，所以只提供了企业版和开发者版。

国内的AR也在快速发展。2020年4月8日，在华为公司2020年春季新品发布会上，余承东隆重推出华为AR地图，据称华为AR地图基于华为河图，实现每平方公里40亿三维信息点，运用高精度空间计算，实现厘米级3D地图，1:1 还原真实世界；2019年6月，在国内



5G 牌照发放这个特别的时间点，VIVO率先展示了一款颇具创造力的产品——vivo AR 眼镜，是一款依靠手机驱动的头戴式AR设备；而现在也有越来越多款手机上线了AR的相关功能

但AR技术本身依然存在着许多问题。首先软件方面，底层算法(输入、输出算法)还需要加强。这需要精确的图像识别技术来判断物体所处的位置以及3D坐标等信息。不同于其他3D定位，增强现实领域的物体位置，必须结合观测者的相对位置、三维立体坐标等信息进行定位，难度要高很多。而如何利用叠加呈像算法，将相关信息叠加显示在视网膜上也是个技术难点。硬件技术方面，上述一些关键的技术都不够完善，行业内没有一个比较统一的标准。

总的来说，AR还是一个新兴行业，许多技术都不完善，但随着5G 商用，AR行业及产品可能会“重获新生”，新兴行业升级期间，无论是小的创业公司还是巨头公司都会面临“鲜有人问津”的问题，但是如果能抓住这个机会的话，就可以成为这个领域的领头羊了。

## 四. 参考文献

- [1]陈靖,王涌天,闫达远.增强现实系统及其应用[J].计算机工程与应用,2001(15):72-75.
- [2]薛立明. 增强现实技术在移动终端中的应用研究[D].山东师范大学,2010.
- [3]孙超. 增强现实环境下基于手势的自然交互[D].浙江大学,2010.
- [4]İbili E, Çat M, Resnyansky D, et al. An assessment of geometry teaching supported with augmented reality teaching materials to enhance students' 3D geometry thinking skills[J]. International Journal of Mathematical Education in Science and Technology, 2020, 51(2): 224-246.
- [5]Ogawa H, Kurosaka K, Sato A, et al. Does an augmented reality-based portable navigation system improve the accuracy of acetabular component orientation during THA? A randomized controlled trial[J]. Clinical orthopaedics and related research, 2020, 478(5): 935.