中山大學
SUN YAT-SEN UNIVERSITY

人机交互技术

1.理论作业（视觉交互）

学 院 名 称	：	数据科学与计算机学院
专业（班级）	：	17 软件工程 1 班
学 生 姓 名	：	**曾峥**
学　　号	：	17343006
时　　间	：	2020 年 6 月 17 日

目录

作业内容：

1、所选交互手段的人类感知机制（可以阅读教材及相关资料论述）
2、所选交互手段的方法综述（可以阅读相关书籍和论文，进行总结书写）
3、所选交互手段的优势和劣势（要与至少其他两种交互方式做对比）
4、要求：提交 word 或者 pdf 电子文档，图文并茂，逻辑清晰，格式规整。其中第 1 和第 3 部分各 25 分，第 2 部分 50 分。

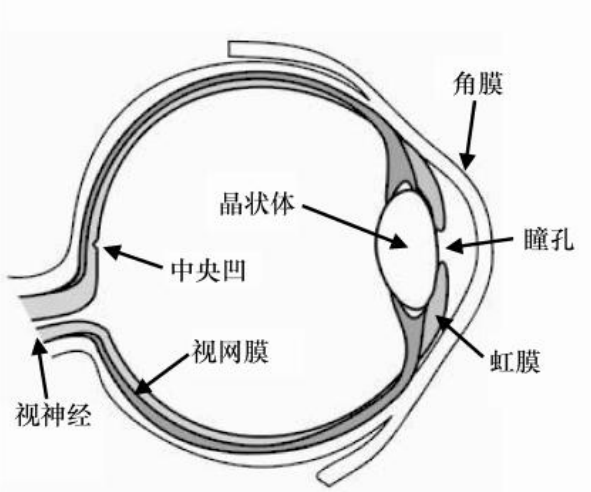
1.视觉交互的人类感知机制

视觉是人类感知外界信息的主要途径，据研究，人类从周围世界获取的信息约有 80%是通过视觉得到的。

视觉感知可以分为两个阶段：

- 接受信息阶段：收到外部刺激
- 解释信息阶段

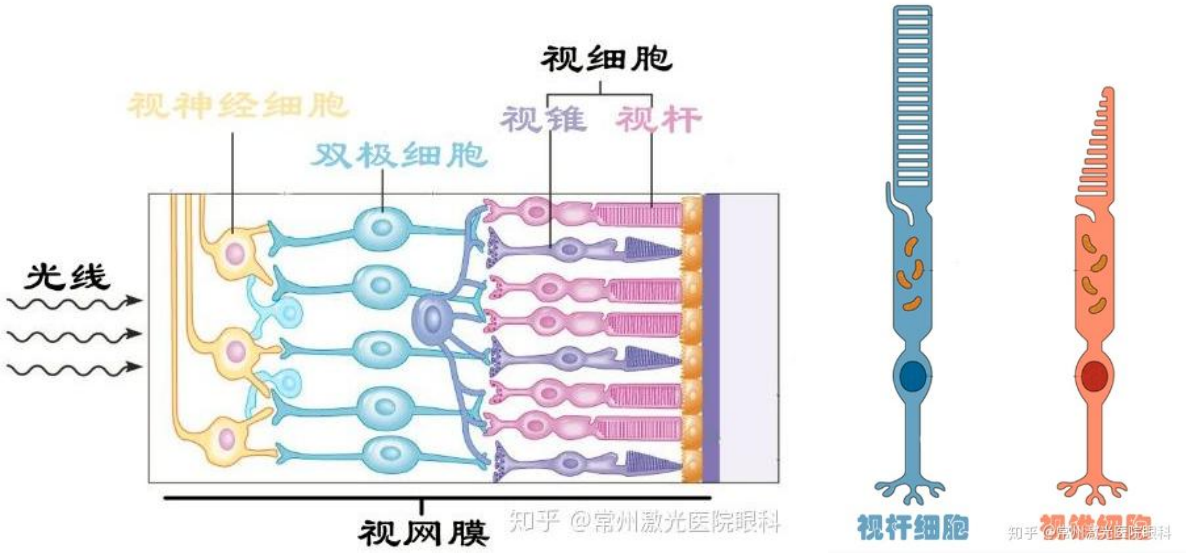
1.1 人眼结构



以上为人眼结构，其机制和相机相似，通过透视系统把外景光源聚焦在眼部后方的视网膜上成像；

1.2 亮度、色彩的感知

在视网膜周围有被称为感光器的感光细胞，这里有两种不同的细胞：



- 锥状体，即视锥细胞，有色觉，主要有集中辨别颜色的锥形感光细胞，分别对黄绿色、绿色和蓝紫色（或称紫罗兰色）的光最敏感。视锥细胞形成的视觉信号复合后为人呈现了色彩缤纷的世界。
- 柱状体，即视柱细胞，无色觉，对灰度敏感，并能将光转换成神经信号。主导夜间视觉的视杆细胞虽然对光的敏感程度要远远强于视锥细胞，但是由于没有像视锥细胞那样感知颜色的能力，所以在幽暗的夜色中我们虽然能看到东西，但却往往感觉不到物体的色彩。

三种可以产生颜色感觉的锥状体，分别对于波长在 420 纳米、534 纳米、564 纳米这三个值附近的光线最敏感，会让大脑分别产生蓝、绿、红三种颜色感觉。而如果蓝绿两种感受细胞同时受刺激，就产生了青色的感觉；同理，蓝红产生品红色，绿红产生黄色。而如果三种细胞都受到差不多的刺激，我们感受到的就是灰色或者白色。

在人眼感知色彩亮度的基础上，可以建立两种色彩模型且用于生活，如下：

1.3 颜色模型

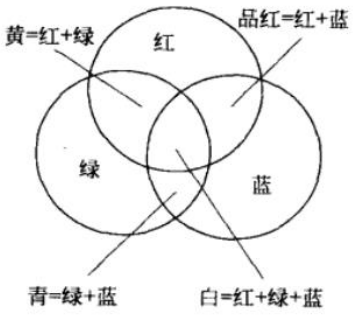


图 2-4　RGB 三原色混合效果

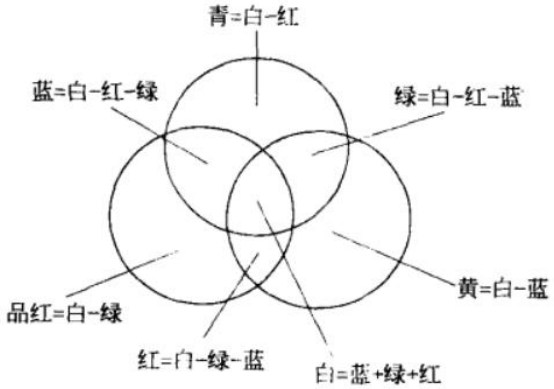


图 2-6　CMYK 原色的减色效果

RGB 是加色模式，CMYK 是减色模式 。

一般来说，**RGB** 用于显示器。**CMYK** 用于打印。原因如下：

- 显示器是一种自发光的装置，RGB 原色吸收两种颜色，反射自身的颜色。根据人对颜色感受的产生原理，它只要用红绿蓝（RGB）三种颜色按不同比例配比，就能不同程度地刺激人的三种感色细胞，产生各种颜色的感觉。
- 但自然界的万物（包括印刷品）不能发光，只能反射光线。不透明物体的颜色是由它反射的光决定的。以太阳光下的物体为例（因为太阳光是全色谱的，也就是说包含了全部波长的光线），如果一个物体是红色的，是因为它吸收了大部分波长的光线，只剩下了能够使人感受红色的光线。用三原色模型来说就是它吸收了蓝光和绿光，反射了红光。
  - 只吸收红反射蓝绿的油墨就是**青色（C）**；
  - 只吸收绿反射红蓝的就是**品红（M）**；
  - 只吸收蓝反射红绿的就是**黄色（Y）**

CMYK 吸收一种光，反射两种光，因此色彩上看更加亮。当然打印机一般都有 4 个墨盒，多出来的一个颜色是**黑色（K）**，可用于调低亮度。

1.4 大小、深度、相对距离的感知

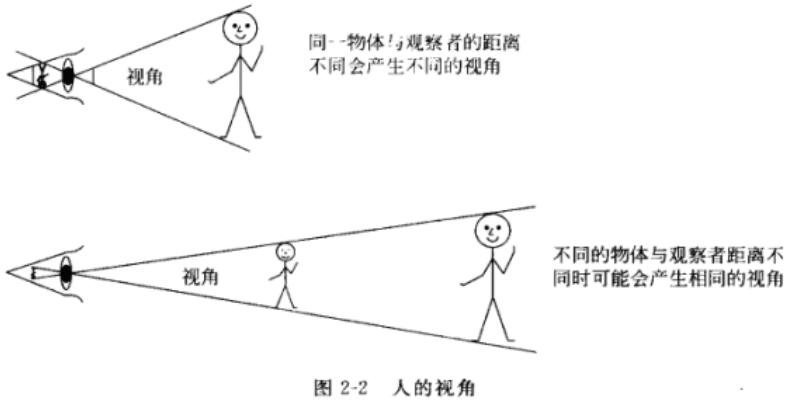


图 2-2　人的视角

大小：‘近大远小’是人对于物体大小最直接的感知；

深度：人的双目产生视差，视差越大，距离我们的距离越小。同一个物体如果在左右两眼单独观察中呈现出来的图像相差越大，说明这个物体离我们越近。

相对距离：人眼具有感知相对距离的功能。大小不同的物体（如一只老虎和一只猫）如果在视觉上呈现出来的图像大小差异不大，那么老虎距离我们的距离必然较大。

在深度特点下，可以研发出 3D 成像设备，输出距离图像（也称为景深映射），如下：

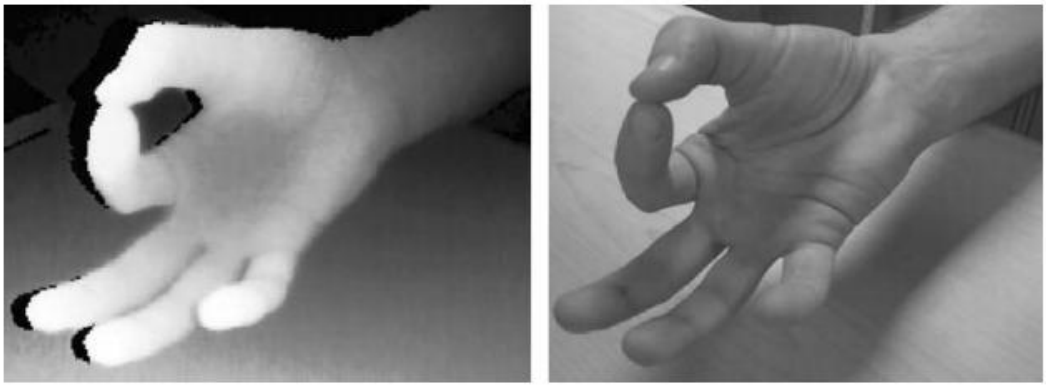
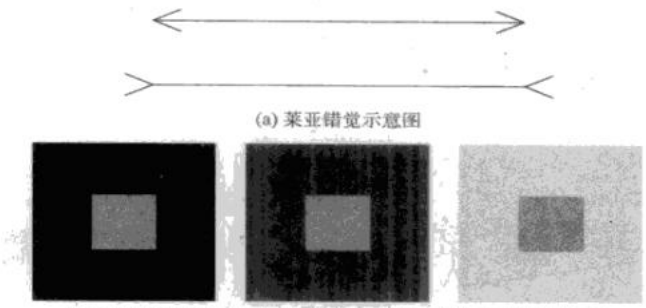


图 4.5　3D 成像设备的输出。左图为在 3D 传感设备前一只手的距离影像，又称为景深映射。景深映射上的灰度值随着图像上的点远离传感器而降低，较近的物体亮度较高。右图为对应的彩色图像。注：两幅图分辨率不同，且经过缩放和裁剪

原理即离传感器越近，物体的亮度越高。在这一技术下，我们还可以进一步识别人体肢体语言，以便更好的交互。

1.5 视错觉



(a) 莱亚错觉示意图

(b) 具有同样亮度值的小方块在不同亮度的背景下感觉亮度不同

图 2-3　视错觉例子

对于图(a)中的两条直线，根据人眼的感知特性，会认为下面的直线更长。

对于图(b)中的 3 个亮度相同的方块，在不同背景下，感知出来的亮度不同。原因是人眼睛无法判断出视场中目标的绝对亮度，人类视觉对亮度的主观响应与目标物的背景亮度有着密切的关系。

2.视觉交互的手段

交互手段除了传统的 **2D** 图像传感，还有近年来的 **3D** 成像技术，后者主流技术主要是 3 种：**结构光、双目、ToF**。

双目技术（即立体成像技术）尚未成熟，结构光和 ToF 技术复杂。2015 年之前，我国在 3D 传感领域一直都是空白，不论是芯片，还是关键零部件都未能完成自主研发。

不过这项技术空白很快被来自深圳的奥比中光填补。2015 年 7 月，奥比中光自主研发了我国首款 3D 深度感知计算芯片—MX400，打破了苹果、微软等国外巨头对 3D 传感技术的垄断。2015 年底，奥比中光正式量产了 Astra 3D 摄像头，这标志着我国消费级 3D 传感技术的一大进步。目前，奥比中光 3D 传感摄像头已应用于智能手机、智慧客厅、新零售等众多领域。

除了苹果 face ID ,国内最近几年也兴起了 3D 人脸识别功能 ,包括手机解锁、支付宝刷脸支付等等。目前国内支持 3D 人脸识别的机型有：华为 nova3、华为 mate20、小米 8 透明探索版、OPPO Find X。

（奥比中光 Astra 3D 传感摄像头）

2.1 2D 图像传感

这是日常中接触比较多的 2D 成像技术，比如数码照相机。

普通的 2D 成像是用平面传感器接收被拍摄物体反射或者发出的可见光，从而形二维图像。

传统图像传感和获取设备会将 **3D** 场景中的视觉信息转化为 **2D** 数组，将现实世界原本 3D 空间中的点作为离散的 2D 映射在图像平面上（像素），比如，一部手机拍出来的照片分辨率为 1792 x 828，即像素为 1200 万。

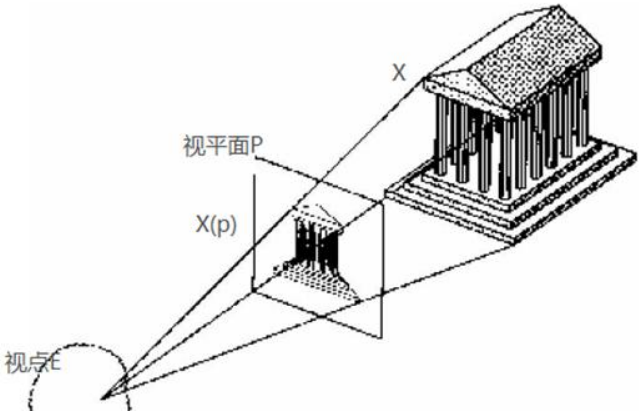
然而，由于现实世界是三维世界，2D 成像便存在物体特征损失(无法获取物体深度信息)的情况，这也就意味着，2D 成像并不支持与物体三维信息的测量，例如 3D 人脸识别、三维建模等 AI 功能，2D 图像技术都无法支持。

其中，3D 视觉信息生成 2D 视觉信息的过程，利用了透视投影技术的齐次矩阵形式，如下：

$$[X'] = [C][X]$$

[X]代表 3D 世界中的点，[C]代表摄像头转化矩阵（涉及平移、缩放变换等等一系列矩阵），[X']代表转化为 2D 图像上的点。当然这里有两点细节：

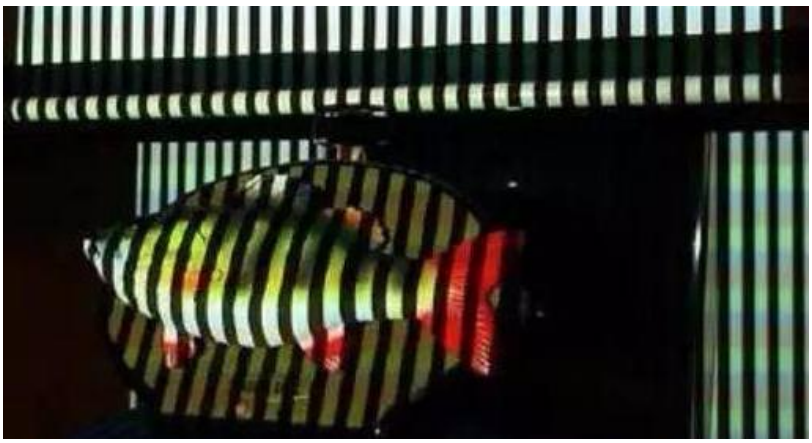
- 用齐次矩阵形式：因为在摄影成像的过程中，视平面的原点与视点原点并不相同。因此在计算过程中涉及到坐标系的变化即点的平移，转化为齐次矩阵形式之后方便计算。
- 透视投影技术，如下：



(透视图成像原理)

2.2 3D：投影式结构光

基本的结构光方案所基于的原理是光学三角法。研究发现，当把一些特殊形式的光投到有不同深度的物体上时，光的纹路会发生变化，而我们可以通过采集这些纹理变化，来计算位置和深度，进而复原整个三维空间。



（图源：tianya.cn）

例如光带打到物体上，遇到突出的物体条纹就会改变原有的状态，或断开，或弯曲，其程度取决于物体各部分的深浅。直接打到平面，条纹则不发生改变。

结构光技术成本较低，技术发展成熟，已广泛应用于计算机视觉、机器视觉和光学测量等领域，在日常生活中，也多用于电子产品前置摄像头。

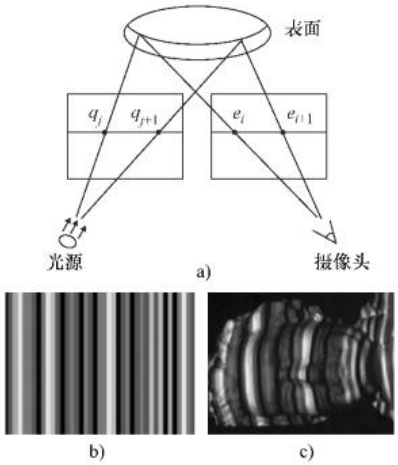
同为 3D 视觉传感，**立体视觉技术**根据人眼原理，利用 2 个相机，从不同视角捕获两个 2D 图像，

通过三角测量来恢复景深信息。但是如果物体表面**没有强烈的纹理变化**，立体视觉技术难以达到高精度；此外，也无法在**两个均匀白色平板表面**获得任何景深信息。因此，在某种意义上来说，DFP 技术是有意义的。

数字条纹投射技术（Digital Fringe Projection、DFP）这个技术已开发多年，是目前被证明具有压倒性优势的技术，并得到广泛应用。其结构化模型呈现正弦并由激光干涉仪产生。

这种方法的基本原理在于利用相位信息建立对应关系，因此对物体的纹理变化具备高的鲁棒性。

DFP 技术将一个相机替换成投影设备 ,投影设备打出带有 patterns 的调制光便可进行相应的三维重构。这一技术的基本原理如下：



a) 照射图案投射在场景上，摄像头捕捉反射图像。由图案与影像之间的相对变形计算得到某个点的景深；

b) 投射条状图案示例。在实际应用中，通常会使用红外光，且图案会更为复杂。

c) 条状图案经 3D 物体反射后所得的捕捉图像。来源： Zhang， Curless and Seitz， 2002

结构光技术大体可以分为四类:

1，随机或伪随机编码类

2，二值结构编码类

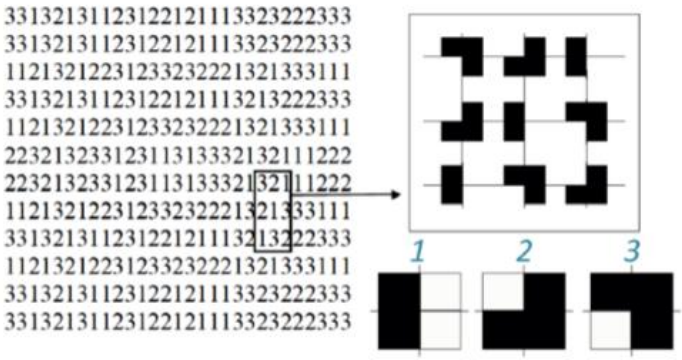
3，N-ary 编码类

4，利用相位信息建立关系的 DPF

1. 随机或伪随机编码类

这种方法的优点就是**简单易操作**，也容易理解：打出独特的 Patterns,直接通过硬匹配 patterns 的方式获取对应点。

该方法的 patterns 在 x,y 方向上变化，商业上也受到广泛应用，比如 Kinect 为伪随机条纹的引用案例（3D points can be reconstructed by triangulation using the captured patterns）

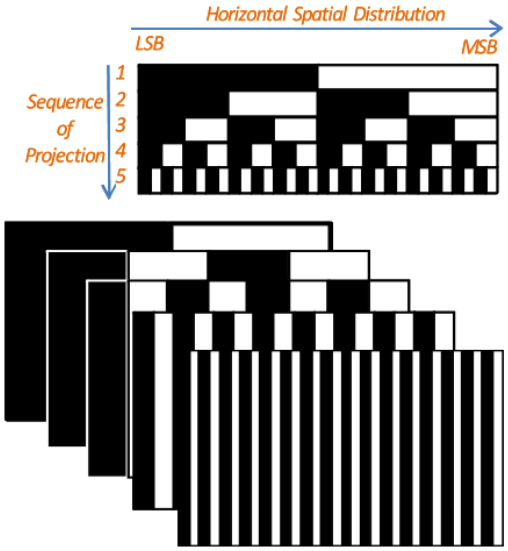


缺点： 由于是二维变化的 patterns，所以 patterns 的大小首先需要大于投影设备的 pixel size, 其次还要大于相机的 pixel size。因此达不到像素级的**分辨率**。

2. 二值结构编码类，即平方二进制散焦技术。

只需要 1 为二进制结构化模型，而不是 8 位的灰度模型，这也使物体表面特性变化稳定，具有处理系统中噪声的能力（**稳健性**）。技术也大大降低了数据传输率，从而使得大于 120Hz 的 3D 形状测试速度成为可能。同时，这个技术的编码、解码算法简单也决定了它的**简单性**。这两个性质让二进制散焦技术得到广泛应用。

缺点：在于**空间分辨率限制**，和随机及伪随机条纹是一个道理：必须大于像素点的大小才能进行重建。此外，以及大量需要汇编的图案，因此**速度慢**：为了提高分辨率，往往要对 patterns 进行 N 次细分，而 N 次细分的代价是大量的 patterns->low measurement speed。



3. N-ary 编码类

N 点编码，即**多进制汇编。**不同于使用两个灰度值（0/255）来为每一个像素值创建独特码字，而是利用这些值之间的一个子集。最极端的情况下会使用所有的灰度值。用书中的话就是：utilizes a subset of the range(using all of the gray-scale values)。

因此，可以显著降低投射所需时间（**高速**），也可以获得**相对高的分辨率。**

当然，测量方法对噪声敏感，易受到相机及投影仪离焦的影响，依然受限于分辨率。

4. 利用相位信息建立关系的 DFP，即连续正弦相位汇编技术。

这项技术也是所用最多的。下图为使用三步相移测量法实现的 3D 传感的实例：

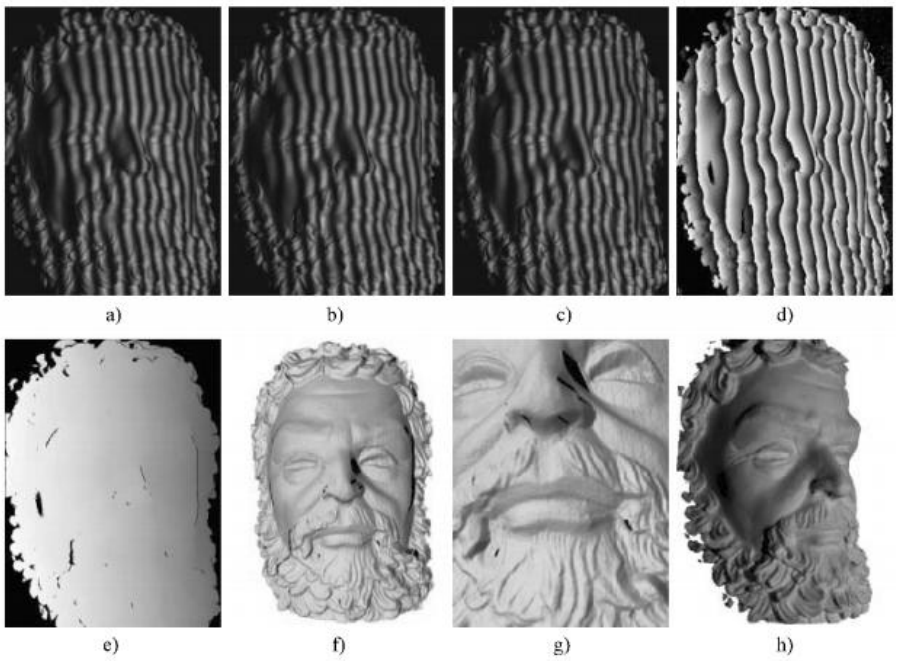
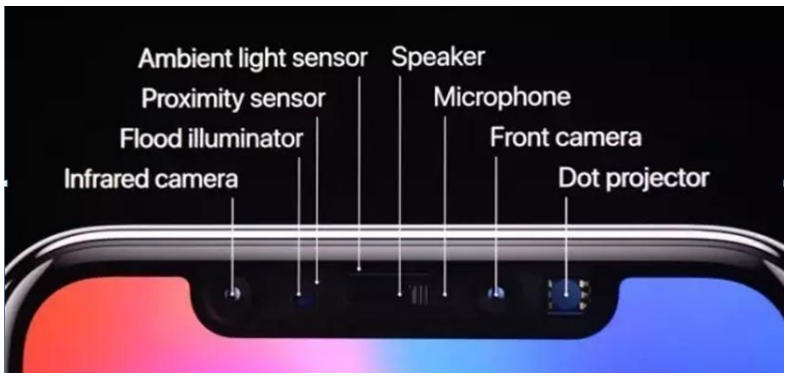


图 5.7　采用三步相移法实现 3D 传感的实例。a）$I_1(-2\pi/3)$；b）$I_2(0)$；c）$I_3(2\pi/3)$；d）包裹相位图；e）去包裹相位图；f）阴影模型绘制的 3D 形状；g）镜头拉近后的图像；h）纹理映射绘制 3D 形状。来源：Zhang S and Huang PS 2006b。转载获取 SPIE 许可

iPhone X 也正是采用的结构光来实现 Face ID 的。

苹果的 Face ID 组成比较复杂，用的是 **3D 结构光**的解决方案，未雨绸缪的苹果早在 2013 年就收购了 PrimeSense 这家专门做 3D 传感器的公司。PrimeSense 这家公司或许大家知道的不多，但是提起微软的 Kinect，相信大家都知道这款体感游戏的鼻祖，早期的 Kinect 正是使用了 PrimeSense 的 3D 传感器。

还记得那屏幕上端抢眼的“刘海”吗？置于其中的是一套称作“原深感摄像头”的系统，除了常规器件，还加入了红外摄像头、点阵投射器、泛光感应元件，也就是 Face ID 的核心元件。其通过投射超过 3 万个光信息识别点，由摄像头收集信息并通过算法分析，实现了它的人脸识别。



(iPhone Face ID、iPhone X 新特性：从 3D 人脸识别到 A11Bionic 神经引擎)

2.3 3D：立体 3D 成像法

和我们生活最贴近的当然是各大品牌发布会时所强调的双目摄像技术。

经过过去几年猛烈的增长和普及后，现面临着销量增速下滑的问题，正力求创新突破窘境。而双摄作为近年为数不多的创新，各大终端厂家均表现出了强烈的兴趣，相继发布带双摄的器件机型。苹果公司在 2016 年发布的 iPhone7Plus 系列中加入双摄配置， 此举具有行业标杆意义，各手机厂家相继跟进，现几乎已成各大品牌旗舰机标配。

手机型号	上市时间	CMOS 图像传感器	双摄类型
HTCEVO3D	2011	未知	不同像素立体摄像头
荣耀 6Plus	2014	后置主副一致：OV8865（800W）	同像素平行双摄像头
华为 P10	2016	后置主副：1200W	同像素黑白双摄像头
360 奇酷旗舰版	2016	后置副：1200W 后置副：索 IMXMONO（1300W）	同像素黑白双摄像头
华为 P9	2016	后置主：1200W 后置副：1200W	同像素黑白双摄像头
LGG5	2016	后置主：索尼 IMX234（1600W）	广角+长焦摄像头
iPhone7 P	2017	后置主：1300W 后置副：1300W	广角+长焦摄像头
华为 mate 9	2017	后置主：1200W 后置副：1200W	同像素黑白双摄像头

(主流广商旗舰机型双摄方案选择)

已经上市的华为 nova3 采用的就是双目 3D 人脸识别方案 ,而且是 IFAA(互联网金融身份认证联盟) 提供的标准，这也是其达到支付级的主要原因。

传统的**单目测距**原理：先通过图像匹配进行目标识别（各种车型、行人、物体等），再通过目标在图像中的大小去估算目标距离。这就要求在估算距离之前首先对目标进行准确识别 ,是汽车还是行人，

是货车、SUV 还是小轿车。准确识别是准确估算距离的第一步。要做到这一点，就需要建立并不断维护一个庞大的样本特征数据库，保证这个数据库包含待识别目标的全部特征数据。比如在一些特殊地区，为了专门检测大型动物，必须先行建立大型动物的数据库；而对于另外某些区域存在一些非常规车型，也要先将这些车型的特征数据加入到数据库中。如果缺乏待识别目标的特征数据，就会导致系统无法对这些车型、物体、障碍物进行识别，从而也就无法准确估算这些目标的距离。

很明显,单目系统的优势在于成本较低，对计算资源的要求不高，系统结构相对简单；缺点是：（1）需要不断更新和维护一个庞大的样本数据库，以达到较高识别率；（2）无法对非标准障碍物进行判断；（3）距离并非真正意义上的测量，准确度较低。

双目测距技术应时而生。

立体 3D 成像法即**双目成像技术（Stereo System）**，顾名思义，根据人眼的原理而开发的技术。空间的两点观察同一事物形成的像差能够为感知被观察事物的景深提供足够信息。这一现象据说首先是有 Charles Wheatstone 爵士在两个世纪前发现的，他声称“…大脑是通过投射到两个视网膜上产生的两张不同图像来感知事物的三个维度的…”

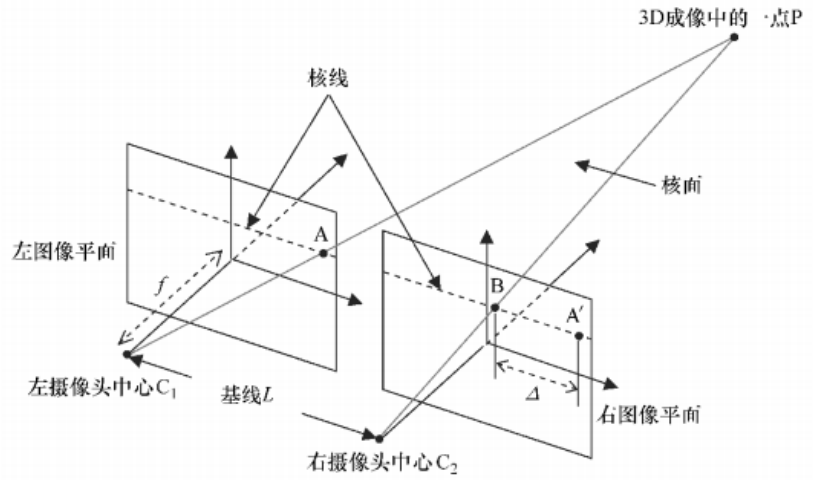
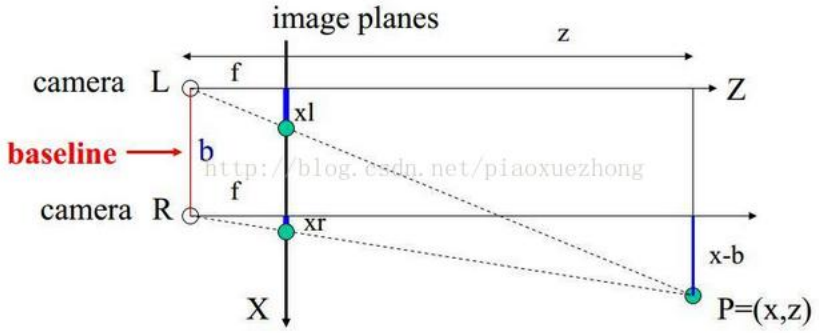
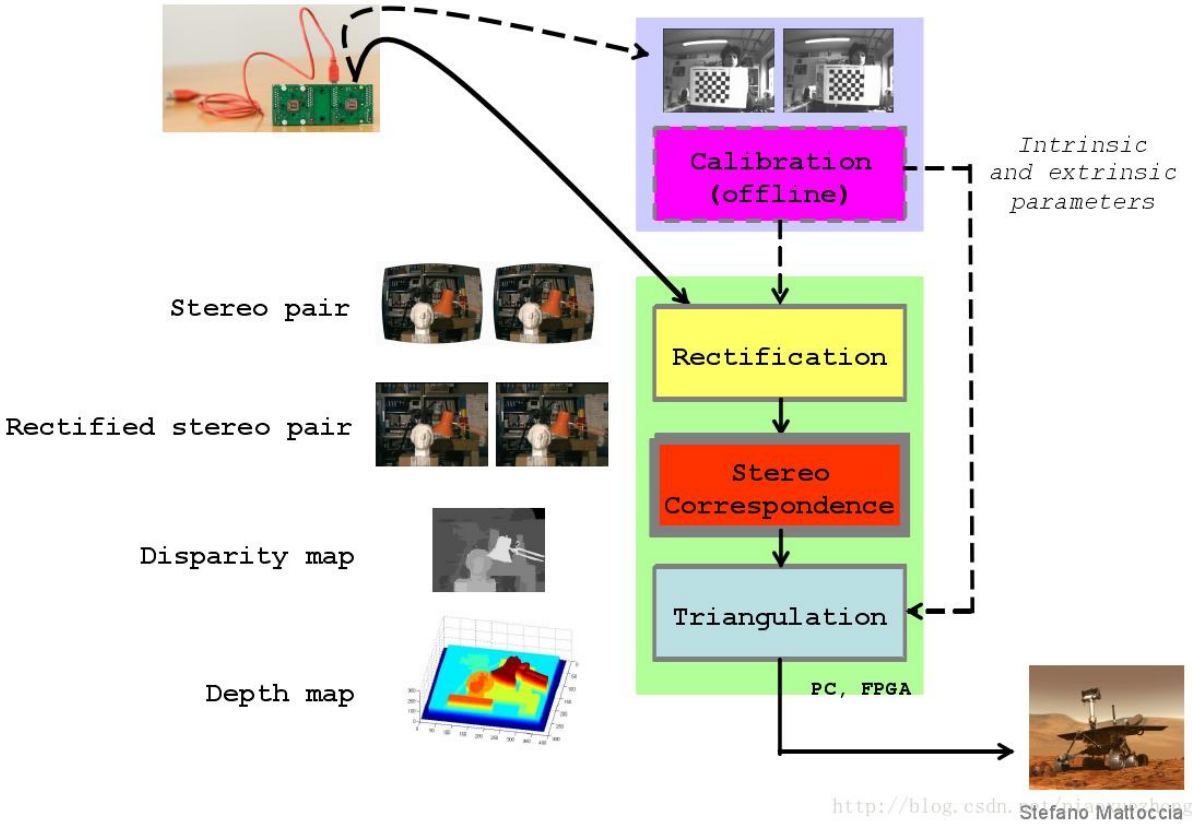


图 4.3　立体 3D 成像方法基本原理。以两台对齐且经过校准的摄像头的简单情况为例，两个摄像头的光学中点分别为 C_1 和 C_2，两者之间基线距离为 L。3D 世界中的点 P 经左右两个摄像头成像分别得点 A 和 B。右图像平面上的点 A′对应左图像平面上的点 A。B 和 A′核线之间的距离称为双眼视差 Δ，此值可知与点 P 到基线之间的距离（或景深）成反比



(来源：双目视觉测距原理，数学推导及三维重建资源)

根据三角形相似定律，我们最终可以得到空间点 P 离相机的距离（深度）z。



双目系统**优势**：（1）成本比单目系统高，但尚可接受，与激光雷达等方案相比成本较低；（2）没有识别率的限制，直接对所有障碍物直接进行测量，无需先进行识别再测算；当然也不用维护样本数据库；（3）直接利用视差计算距离，精度比单目高。

难点：（1）计算量非常大，对计算单元的性能要求非常高，这使得双目系统的产品化、小型化的难度较大。所以在芯片或 FPGA 上解决双目的计算问题难度比较大。国际上使用双目的研究机构或厂商，

绝大多数是使用服务器进行图像处理与计算，也有部分将算法进行简化后，使用 FPGA 进行处理。（2）双目的配准效果，直接影响到测距的准确性。（3）对环境光照非常敏感。（4）不适用于单调缺乏纹理的场景。（5）计算复杂度高。（6）相机基线限制了测量范围。

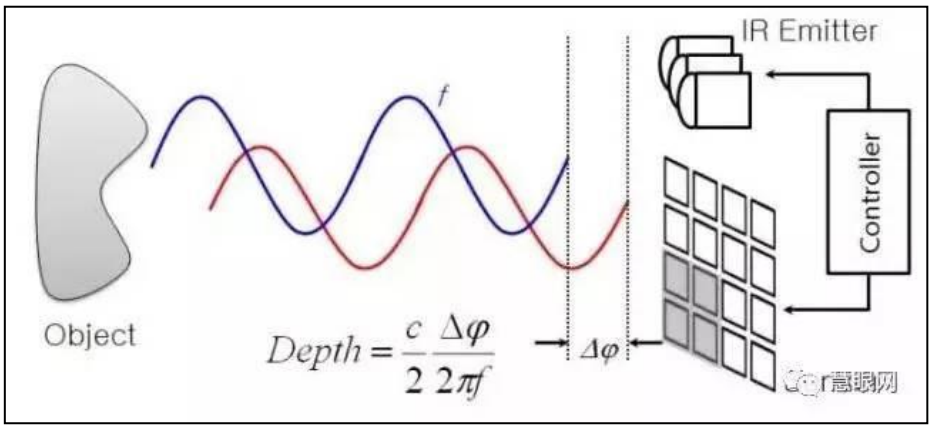
2.4 3D：ToF 飞行时间法

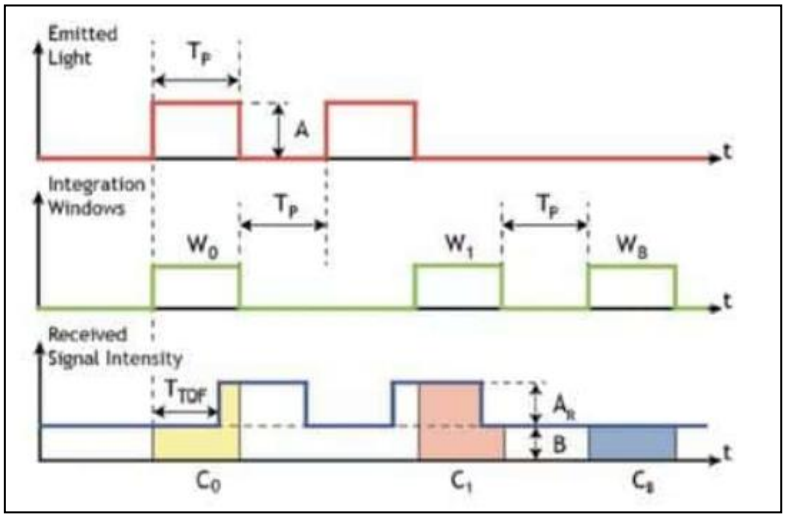
飞行时间法 3D 成像方法利用调制红外光来照射物体和场景， 计算光从成像设备出发后经物体或场景反射后回到光源的往返时间 （常采用相移测量技术[9]）， 测出物体各点的距离， 由此获得景深映射。 这套系统通常具备全场范围成像能力，包括已调幅的照射源和图像传感器阵列。

军事上和无人驾驶汽车上用的工业级激光雷达（LiDAR）也采用到了 ToF 技术，利用激光束来探测目标的位置、速度等特征量 结合了激光、全球定位系统 GPS 和惯性测量装置(Inertial Measurement Unit，IMU）三者的作用，进行逐点扫描来获取整个探测物体的深度信息。

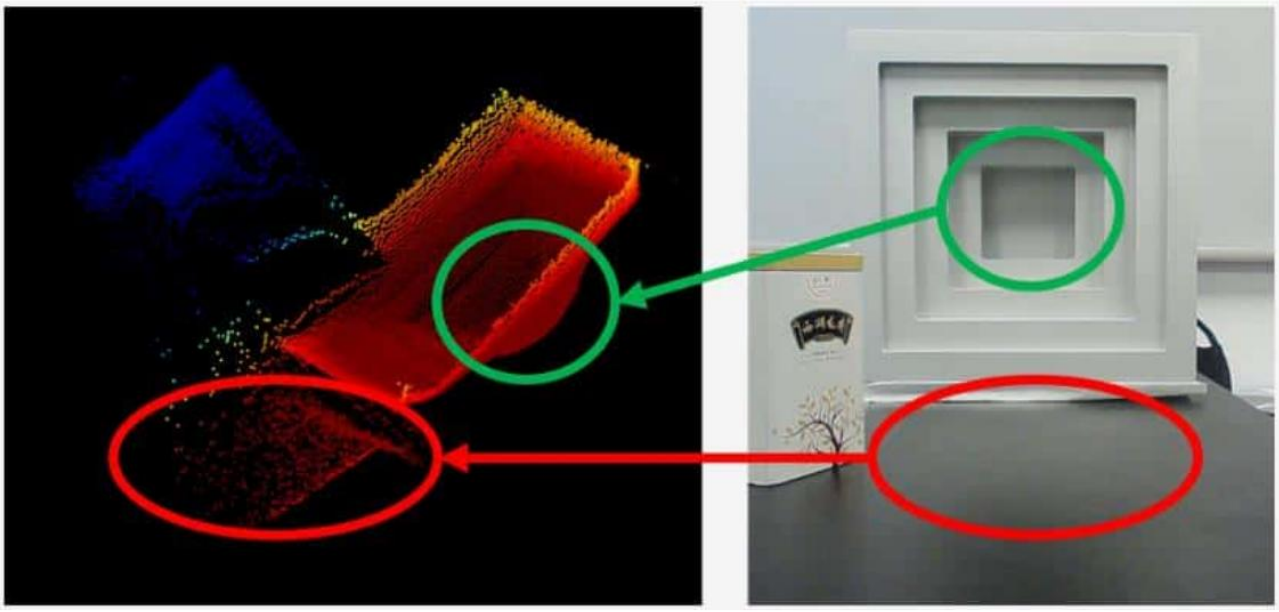
- **i-ToF**，即 indirect ToF，用测量信号的相位来间接地获得光的来回飞行时间。
- **d-ToF** (direct time-of-flight) 技术直接测量光脉冲的发射和接收的时间差。

i-ToF 可分为两种：连续波调制（CW-iToF）和脉冲调制（PL-iToF）,分别发射连续的正弦信号和重复的脉冲信号；前者是通过解析正弦信号相位解析深度，而后者是解析脉冲信号相位来解析深度。两种方法的原理图分别如下：(来源)





i-ToF 主要挑战：**多径干扰（Multi-Path Interference, MPI)**：真实场景中存在复杂的漫反射甚至镜面反射，MPI 在原理上会使得测量值变大，严重影响三维重建的效果。如下：



多径干扰示意图：多径干扰导致标准件测量点云形状扭曲（绿色），以及桌面错误地测量成标准件镜像（红色）.来源

MPI 是困扰 i-ToF 多年的重要问题，一直是 i-ToF 广泛应用的最大障碍。在过去的十年中，微软，MIT，Waikato 大学等诸多研究机构在解决 MPI 问题上做出了大量算法和系统层面的尝试[13]，但仍无法根除该问题。

TOF 技术优势：

- 体积小，误差小

TOF 相机要求接收端与发射端尽可能的接近，越接近，由于发射、接收路径不同而带来的误差就越小，从体积紧凑角度来讲有着天然的优势；

- 直接输出深度信息

  TOF 可以直接输出深度信息，不需要类似双目立体视觉或者结构光等需要通过算法计算来获得深度信息。

- 抗干扰强

  TOF 不受表面灰度和特征影响，太阳光由于没有经过调制，所以 TOF 抗强光能力也较好。TOF 的精度不随着距离的变化而变化，基本可以稳定在 cm 级。

TOF 技术劣势

- 分辨率偏低，功耗大
- 功耗部分有待提高。
- 解决方案不够成熟

三种技术介绍到此介绍完毕。最后总结一下 3 种 3D 成像技术方案的优劣：

	双目	结构光	i-ToF	d-ToF
适用场景	近距离	近距离	中远距离	中远距离
基本原理	三角测距	三角测距	相位测距	时间测距
Sensor	RGB/IR CMOS Sensor	IR CMOS Sensor	i-ToF CIS	SPAD Array
工艺难度	容易	容易	中等	难
传感器信号	模拟	模拟	模拟	数字
发射光脉冲	无	低频率	中高频率	高频率
测量精度	近距离高，随测量距离平方下降	近距离高，随测量距离平方下降	距离呈线性关系	在工作范围内相对固定
功耗	低	中	高	中
多路径串扰	容易解决	难度适中	较难解决	容易解决
量产标定	简单	中等	难	中等

3 交互方式的优劣分析

3.1 多种交互方式

首先先介绍几款交互方式的基本情况，然后表格汇总。

视觉交互：基本情况在上文已经介绍了很多，不再赘述。

语音交互：语音界面的有趣之处在于它允许自然地与机器交互。

听觉感知传递的信息仅次于视觉，生活中的例子如下：

- 智能家居 - 通过简单的语音命令，可以轻松管理时间，警报，温度，控制灯的色彩，开关时间，摄像头。

- 智能车辆 - 驾驶汽车时，语音界面非常方便，免提信息，保持驾驶员安全，让人专注于道路。
- 基于电话的应用程序 - 在进行电话银行业务或致电客户服务代理时，使用语音识别，无需密码或验证，这对许多银行客户来说都是一种解脱。例如，花旗银行能使客户轻松地在电话中进行验证。

触觉交互：触觉在交互中的作用是不可低估的，尤其对有能力缺陷的人，如盲人，是至关重要的。

触觉反馈已经被许多界面用来传达重要的系统状态信息，特别是用于传递一个给定的操作已经完成的信号。例如，如果没有明显的触觉或听觉的感觉，开关按钮的操作容易令人困惑。如果没有明显的视觉或听觉的开关操作指示，那么触摸则是提供反馈最可能的途径。键盘便被设计成

了提供这种类型的反馈。

在触觉界面方面，研究人员已经做了很多尝试，包括通过一系列的电动马达提供触觉反馈的电脑鼠标（G 鳃 el et al. 1995），能让用户感觉到屏幕上纹理（例如编织物）的电脑（Dillon et al. 2001），由乳胶制成的、能让人同时使用触觉和语音进行交流的手机（Park et al. 2013）等。

3.2 优劣分析

项目	视觉交互	语音交互	触觉交互
优势	（1）准确，可长时保留信息，可展示复杂的内容。	（1）语音使用自然语言并将其转换为命令免去了动手的麻烦；	（1）信息较隐蔽，受环境因素影响较小
劣势	（1）需要花时间看，受距离影响较大，输出设备占用体积大。 （2）摄像头的隐私问题、漏洞问题。	（1）隐私性不好：语音界面通常不利于隐私和嘈杂的环境。文本，触摸和图形用户界面可以是私密的，尤其是手机，大型平板电脑或笔记本电脑屏幕上的屏幕过滤器； （2）速度慢：人类在获取知识到时候，远不如视觉上接收信息速度快；	（1）不适合输出略微复杂的信息，距离极近。

因此，我认为对于视觉交互这种输入离不开摄像头、输出离不开显示器的技术，优缺点在于：

优点：

1. 直观易理解。人获取信息主要靠眼，无论是显示器中的动画视频，还是纸张中的问题图像，又或者

全息投影，视觉输出可以让人快速直观理解到信息。

2. 便捷生活。近年来的人脸识别用处很大，手机解锁免去输入密码、密码被盗的麻烦。对比与以前的指纹解锁，后者会受到潮湿度等环境因素的影响。支付宝支付也节省了人们输入密码的时间，能够快速购物，减少排队时间。
3. 配合动作使交互更加便捷。比如谷歌 motion sense 隔空切歌，只需要在手机上方做一些手势，摄像头会捕捉信息，就可以播放下一首等功能。

缺点：

1. 多模交互更加强大。仅仅靠视觉是不够的，人类感知虽说除了视觉，但仍然有听觉、触觉等，更何况这并不能让盲人受益。多模交互能够视人机交互更加优秀。
2. 计算机性能问题。从输入设备来看，离不开摄像头。镜头工业使得相机像素越来越高，那么计算机在处理图像数据的时候，若保证不失真，那么对于计算机的性能以及算法的优化均有一定的挑战。
3. 隐私问题。这一点和语音交互相同。亚马逊的语音交互产品 Alexa 在一次事故中将一对美国夫妇的秘密录音发送给欧洲的陌生人。那么视觉交互在人们的生活中是否也可能存漏洞问题？
4. 大范围环境建模问题。对于科幻电影中的大范围环境建模来说，工业上实现仍然存在技术难点，结构光技术和双目技术对于超出一定范围之后的景物就失去了精度，ToF 也存在多路径问题。

4 参考文献和资料

《实感交互：人工智能下的人机交互技术》
锥状体、柱状体
为什么一到晚上眼里就失去色彩-zhihu.com
人眼的视觉特性
为什么打印和印刷时要使用 CMY 颜色模型，而不是 RGB-zhihu.com
为什么要引入齐次坐标
【数字条纹投影技术基础 2】非接触光学三维测量技术综述
Structured-light 3D surface imaging: a tutorial
3D 传感与 2D 成像的区别，及 3D 传感在 AI 领域的应用优势解析
从双摄到 3D 成像，中国光学崛起之路
双摄像头测距、环境建模-zhihu.com
《人工智能之人机交互》报告：AI+人机交互的现状与未来
三分钟深度解读虹膜识别手机
双目视觉测距原理，数学推导及三维重建资源
清华创业团队：ToF 技术白皮书
交互方式的选择