机器人视觉感知

视觉是人类，也是多数动物最重要的环境感知方式。从寒武纪大爆发开始，动物多次演化出不同结构形式的眼睛，以图在演化竞争中获得优势，充分说明视觉在环境感知中的重要性。对于机器人而言，视觉同样赋予其感知环境，从而具备自主或半自主工作的能力。本章将对机器人的视觉感知加以详细介绍。

1. 视觉成像基本原理

人类和动物通过不同结构类型的视觉器官形成视觉，这些视觉器官通常称为眼睛。我们要问，为什么不能直接成像？比如，在一块白色幕布前放置一根点燃的蜡烛，为什么幕布上看不到蜡烛的图像，而只能看到一团闪烁的亮光？图1a显示了在幕布前放置一根蜡烛的示意图，可以看到，由于光线的散射，幕布上的每一个点都会反射蜡烛各个部位发出的光线，导致光线之间互相混淆，因此人们在幕布上看到的仅仅是一团模糊的光影，看不到清晰的蜡烛成像。

如果在蜡烛和幕布之间放一块不透明的挡板，挡板上钻一个小孔，仅允许光线透过小孔到达幕布，如图1b所示。可以看出，幕布上每一个点仅会反射蜡烛一个部位发出的光线，避免了光线的互相混淆，从而形成一个清晰的蜡烛倒像。这一原理，即所谓的小孔成像。

|  |  |
| --- | --- |
|  |  |

图1 小孔成像原理

进一步观察还会发现，无论物体、小孔和幕布之间的距离是多少，总是可以在幕布上形成清晰的倒像。物体与小孔越近，则成像越大，反之越小。这一特点称为小孔成像具有无穷远景深。



图2. 小孔成像景深

小孔成像原理简单，且具有无限景深的优点，但同时也存在多种缺点。由小孔成像原理可以看到，小孔直径越小，则产生的像越清晰。但是，当小孔直径减小时，能够透过小孔到达幕布的光线总量也在显著减少，从而导致成像昏暗不清。此外，当小孔直径减小到一定程度时，光线会发生衍射现象，影响成像的清晰程度。因此，小孔成像很少在实际的光学成像系统中得到使用。



图3. 小孔成像的局限(左图：大直径小孔导致图像发散；右图：小直径小孔由于衍射同样导致图像发散)

凸透镜能够将同一个点发出的散射光线折射到另一个点，从而形成图像，如图3所示。由于所有能够到达凸透镜的光线都会被折射到幕布上，因此凸透镜所形成的图像远比小孔成像更为明亮和清晰。由于凸透镜具有成像清晰明亮的优势，人类和脊椎动物，以及章鱼等软体动物均进化出了具有凸透镜形状的晶状体，作为眼睛的成像器官。

|  |  |
| --- | --- |
|  | 人眼球结构中，分别相当于照相机镜头和光圈的是（ ）A．巩膜和虹膜B．玻璃体和瞳孔C．角膜和视网膜D．晶状体和瞳孔..._易学啦 |

图4. 凸透镜成像(这张图画的不太好)

1. 镜头

单个凸透镜的成像质量较差，存在色散等多种缺陷。因此，实际的光学系统通常会组合使用多个凸透镜和凹透镜，形成透镜组，以提高成像质量。这些透镜组通过外壳组合在一起，构成了镜头。镜头最前面的透镜通常还会镀膜以提高透光量，从而进一步提高成像质量。由于本书不涉及到详细的镜头设计，为了便于理解，以下仍以单个凸透镜为基础，简要介绍与视觉感知相关的镜头参数。需要注意的是，

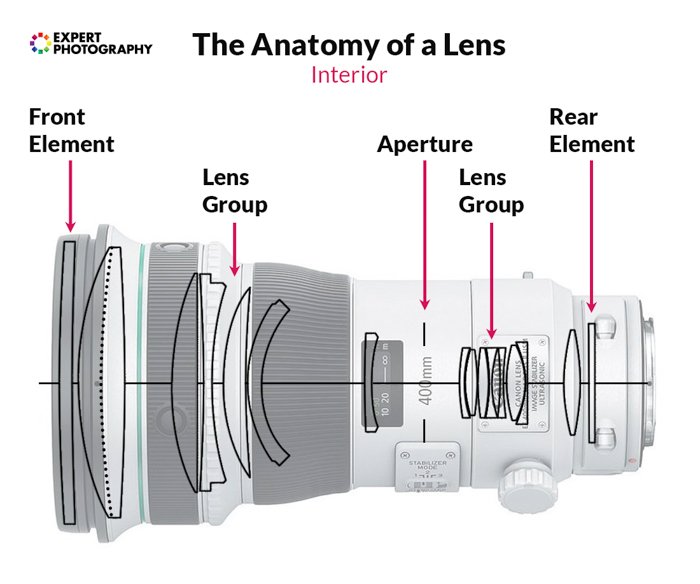


图5. 镜头结构示意图

镜头在其一端收集物体的光线，将光线在另一端汇聚为实像，并投影到成像平面上。汇集光线的点称为焦点，镜头中心到焦点的距离称为焦点距离。当镜头为凸镜时，焦点距离将根据镜头的厚度（膨胀）程度不同而各不相同，膨胀程度越大焦点距离越短。当焦点距离近时，镜头能够将大范围物体形成较小的像，称为广角镜头。当焦点距离较远时，镜头能够将远处的物体形成较大的像，称为长焦镜头。

图表, 图示, 雷达图

描述已自动生成

1. 光圈和快门

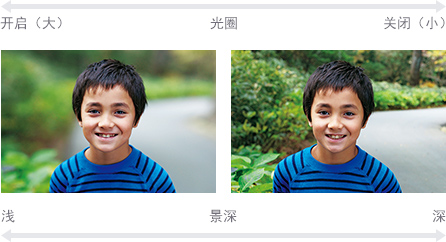
当凸透镜与幕布的距离一定时，只有当物体恰好位于其焦点时，才会在幕布上形成清晰图像，远离或靠近凸透镜时成像均变得模糊。这一现象称为凸透镜具有有限且较浅的景深。此外，当物体的亮度不一致时，凸透镜所形成的像同样具有不一致的亮度，不利于胶片或感光元件形成稳定成像(当亮度较大时，会出现过曝；亮度较低时，出现欠曝)。



图4. 凸透镜景深

为解决以上问题，通常在透镜与幕布之间，增加一个直径可变的小孔，称为光圈。光圈仅允许部分透镜折射光线到达幕布，通过小孔成像原理增强了图像的清晰程度。当物体偏离透镜的焦点时，光圈越小，则幕布上的成像越清晰。也就是说，减小光圈能够增大景深，使更大范围内的物体在幕布上形成清晰图像。反过来，当光圈增大时，景深减小，仅允许透镜焦点附近小范围内的物体在幕布上形成清晰图像。

光圈的这一特点在摄影和视觉感知中具有重要价值。在有些情况下，例如一台巡游机器人，希望能够看清大范围内的环境物体，应当令摄像机的光圈较小。另外一些情况下，例如在流水线上检测物体是否存在缺陷，希望仅对特定位置的物体清晰成像，而背景能够被模糊化，则应当令摄像机的光圈较大。



光圈的大小同样会影响到幕布上图像的明亮程度。光圈较小时，能够到达幕布的光线较少，图像较暗。光圈较大时，能够到达幕布的光线较多，图像较亮。在实际应用中，当环境光照较强，成像有可能过曝时，可以减小光圈；当环境光照较弱，成像可能欠曝时，则增大光圈。通过调节光圈大小，确保在不同光照条件下，幕布上的成像具有大致一样的亮度。

光圈通常用“F/数字”表示，比如F/1.4、F/2.8、F/4、F/5.6、F/8、F/10等。为了方便，很多时候忽略掉/，直接写为F1.4、F2.8、F4、F5.6、F8、F10。数值越大，则光圈直径越小。每增加一档数值，则进光量减小一半(即光圈面积减小一半)。



光圈大小同时控制了成像的景深和亮度，很多时候这两者存在矛盾。例如，当环境光强烈时，希望拍摄浅景深的图像。此时光圈较大，入射光量较多，很可能导致图像过曝。为了进一步控制

1. 感光传感器

早期相机采用以卤化银为基础的化学底片作为感光元件。当光线通过镜头照射到底片上时，光线使卤化银分解形成影像，经过显影、固影和清洗后成为照片。1969年贝尔实验室发明了以感光耦合元件([charge-coupled device](https://en.wikipedia.org/wiki/Charge-coupled_device)，CCD)为基础的半导体感光传感器，1985年奥林巴斯公司发明了NMOS感光传感器，在此基础上美国航空航天局的喷气动力实验室(NASA JPL)发展出了CMOS(complementary metal oxide on silicon)感光传感器。将CCD传感器或CMOS传感器安装在镜头之后，能够将图像直接转换为电信号，经过高速采样和处理后，得到数字图像。

CMOS传感器的成像质量一般低于CCD传感器，因此在对成像要求较高的领域，如摄影、传媒等，通常采用CCD传感器。但是，CMOS传感器结构更简单，生产成本较低，其功耗更是远小于CCD传感器(通常仅有CCD传感器的百分之一)。对于多数移动机器人，特别是飞行机器人而言，低功耗是一个显著的优势。因此，在工业自动化、机器人控制等领域，CMOS传感器得到更为广泛的使用。

CCD传感器和CMOS传感器仅仅对光照强度敏感，每个感光点上的光强越大，则形成的数字量数值越大。因此，传感器仅能形成不同程度的黑白图像，无法反应出颜色信息，称为灰度图。有多种方法对传感器加以改进从而产生彩色图像，其中最常见的一种方法称为拜尔滤镜(Bayer filter)。传感器芯片上的感光像素点被组成2\*2的单元，每个单元包含红色、绿色和蓝色滤镜，从而使每个像素点仅能感受到红色、绿色或蓝色的光照强度。通常情况下，每个单元中有两个像素点用于感知绿色光强，另外两个像素点感知红色和蓝色光强。由于每个像素点只能感知一种颜色的深浅，需要通过内插算法，获得每个像素点的完整色彩，形成RGB图像(色彩的表示也需要说明)。图XX显示了基于拜尔滤镜和内插算法产生RGB彩色图像的原理示例。

|  |  |
| --- | --- |
|  |  |

图X 拜尔滤镜和内插算法产生彩色图像([Bayer filter - Wikipedia](https://en.wikipedia.org/wiki/Bayer_filter))

尽管RGB图像能够准确反映出感知对象的颜色信息，但是很多时候，机器视觉算法并不关心颜色，而是通过物体的形状特征检测感知。由拜尔滤镜原理可知，采用这种方式拍摄图像时，滤镜会造成原始图像出现马赛克，需要采用内插算法恢复，从而导致图像清晰度下降。同时，RGB图像每个像素的位数较多(通常为24位)，数据存储和传输时需要更大的存储空间和传输带宽。因此，当图像颜色对于视觉检测没有影响时，可以考虑选择不包含色彩滤镜的灰度传感器，提高图像清晰度，降低数据存储和传输压力。

CCD和CMOS传感器的技术参数有很多种，受到篇幅限制在这里不做详细解释，仅对最常用的几种参数加以说明。在选择图像传感器时，考虑最多的因素是传感器的像素数量。

1. 工业相机和针孔相机

机器人领域通常使用工业相机或针孔相机实现视觉感知。工业相机相比民用相机而言，其结构更紧凑，机械强度更高，具有更高的图像稳定性，更高传输速率和更高抗干扰能力。当机器人应用于相对恶劣环境，或对于可靠性要求较高时，通常首先工业相机。在选择工业相机时，主要考虑以下参数。

1. 光学接口

工业相机通常由机身和镜头两部分组成。机身中安装CCD或CMOS传感器、控制电路、通信接口等硬件设备。根据需要，在机身上安装不同焦距的镜头，从而对一定范围内的场景加以感知。为了结构简单可靠，工业相机通常不使用变焦镜头。工业相机与镜头之间通过光学接口安装固定，光学接口常用的标准包括C口、CS口和F口，选择使用镜头时要注意匹配。



1. 分辨率

分辨率由工业相机所采用的成像传感器芯片分辨率决定，是芯片靶面排列的像素数量。工业相机的分辨率通常用水平和垂直分辨率两个数字表示，如：1920（H）x 1080(V)，前面的数字表示每行的像素数量，即共有1920个像素，后面的数字表示像素的行数，即1080 行。现在相机的分辨率通常表示多少K,如1K（1024)，2K(2048)，4K(4096)等。在采集图像时，相机的分辨率对图像质量有很大的影响。在对同样大的视场（景物范围）成像时，分辨率越高，对细节的展示越明显。

1. 帧频

帧频表示相机采集图像的频率，单位fps（Frame Per second），即一秒钟可以拍摄到的图像帧数。例如，30fps表示相机再1 秒钟内最多能采集30帧图像。需要注意的是，帧频是相机在理想情况下的最高图像采集速率，在实际使用中，它还受到环境光照、光圈大小等影响。例如，当环境光线较暗时，为了保证图像曝光充足，曝光时间会被延长，从而导致每秒钟实际采集的图像数量小于帧频。

1. 像元深度

数字相机输出的数字信号，即像元灰度值，具有特殊的比特位数，称为像元深度。对于黑白相机这个值的方位通常是8-16bit。像元深度定义了灰度由暗道亮的灰阶数。例如，对于8bit的相机0代表全暗而255代表全亮。介于0和255之间的数字代表一定的亮度指标。10bit数据就有1024个灰阶而12bit有4096个灰阶。每一个应用我们都要仔细考虑是否需要非常细腻的灰度等级。从8bit上升到10bit或者12bit的确可以增强测量的精度，但是也同时降低了系统的速度，并且提高了系统集成的难度（线缆增加，尺寸变大），因此我们也要慎重选择。(这段话要再整理下)

1. 数据总线

工业相机通常通过数据总线将图像传输至用于接收和处理图像的计算机中。常用的数据总线类型包括：

以太网：

USB：

1394总线：

工业相机可靠性较高，成像质量较好，但是其价格成本也相应较高，且体积重量相对较大。在设计开发低成本或小型机器人时，针孔相机是另一种常用的相机类型。

1. 相机数学模型