第十一章 成像模型

## 引言

视觉系统在现代科技和工程领域中扮演着至关重要的角色。它是一种模仿人类视觉系统的技术，能够通过感知环境中可见的光信号，进行图像捕捉、处理和理解。视觉系统的搭建涉及到多个关键技术，其中包括视觉成像模型和视觉系统标定。

视觉成像模型是描述光学成像过程的数学模型，它通过建立相机和场景之间的关系，将三维世界中的物体映射为二维图像。这些模型基于光的物理性质和相机的内部参数，能够帮助我们理解图像形成的原理和规律。通过对视觉成像模型的研究和分析，我们能够准确地推测出在特定条件下图像的质量、畸变、透视和尺寸等特征。

然而，要使视觉系统能够准确地捕捉和处理图像，还需要进行视觉系统的标定。视觉系统标定是指确定相机的内外参数，以及其他相关传感器的校准过程。通过标定，可以获取到相机的畸变参数、相机和世界坐标系之间的转换关系，从而使得视觉系统能够通过图像数据来还原三维空间的形状和位置。准确的系统标定对于视觉导航、目标跟踪、立体视觉等应用都至关重要。【最后再写】

## 11.1 成像模型

成像模型是构成计算机视觉系统的基本依据，它既解释了视觉系统是如何捕获周围环境信息并形成图像的，同时也是视觉系统硬件选型、视觉系统参数标定，乃至立体视觉系统构建、三维场景恢复的主要数学依据。

根据相机成像原理的不同，其对应的成像模型自然也有所区分，但是在实际应用中使用最多的依然是透视投影模型。一则是因为它的数学形式比较简单，在齐次坐标下，能够形成线性的表达形式，二则是因为它描述了可见光条件下，透镜组合的成像过程，是日常生活中使用的照相机、录像机等设备所普遍符合的成像模型，第三则是因为透视投影模型的演变历史悠久，在中国可以追溯到春秋战国时期。

因此，本章的内容从透视投影模型开始，介绍它的数学表达形式。在此基础上，针对其中涉及的各个参数介绍如何进行相机标定。最后，将理论与实际相结合，介绍如何根据透视投影模型对实际的视觉系统进行硬件选型。

### 11.1.1 针孔模型

透视投影模型产生可以从针孔模型讲起，而针孔模型产生来源于这样一个思考。假设在一张成像幕布的前方放置好用于成像的物体，如图11-1所示，那么该物体的像是否能够在幕布上呈现出来呢？答案显然是否的。那么为什么得不到物体的像呢？细究其中的原因就会发现，之所以没有得到物体的像，并不是因为没有来自该物体的光线投射到幕布上，而是因为，投射到幕布上的光线太多了。细看图11-1就会发现，这时来自物体上不同点的光线会投射到幕布上的同一点，从而导致在该点上分不清对应的光线到底来自何处，于是自然也就不会有物体的影像出现了。

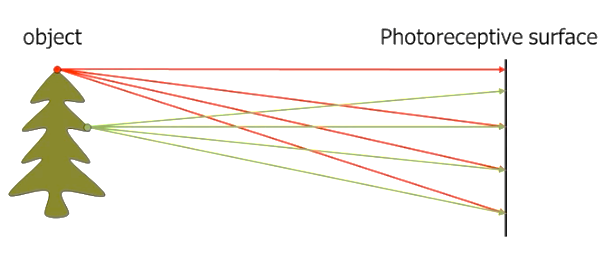


图11-1针孔相机初级模型

有了上面的认识，自然也就能够想到，如果幕布不能成像是因为投射到它上面的光线太多了，那么是不是可以在物体和幕布之间设置一个障碍，阻挡住大部分光线，这样是否就会有物体像出现在幕布上呢？于是，小孔成像模(Pinhole Model)型就这样出现了，如图11-2所示。显然，障碍物中间的小孔是必须的，否则光线的传播就会不彻底阻挡住，使得幕布上不能获得任何光线，自然也不会形成物体而像。这个至关重要的小孔就被称为针孔或者光心。从图11-2中不难发现，在小孔成像模型中幕布上得到像的方向与物体的方向是相反的，即得到是物体的倒像。然而，小孔成像模型困扰人们的远不止“倒像”这一点。

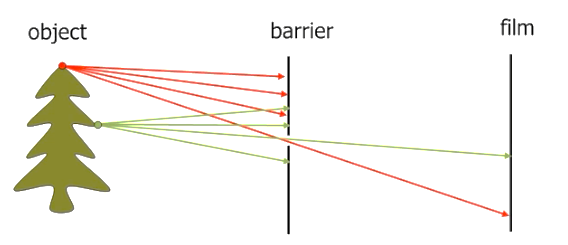


图11-2针孔相机简易模型

小孔成像模型的原理早在两千四五百年前，就已经被发现了。在《墨经》中就有这样的记载：“景到，在午有端，与景长。说在端。”“景。光之人，煦若射，下者之人也高；高者之人也下。足蔽下光，故成景于上；首蔽上光，故成景于下。在远近有端，与于光，故景库内也。”在后一段话中不仅解释了小孔成像的原理，而且还被看做是光沿直线传播的最早记录。在西方的自然研究著作中最早提及小孔成像的，被认为出现在公元5世纪到6世纪由亚里士多德撰写的《问题集》(Problemata)中。小孔成像模型还被著名的莱昂纳多·达芬奇用于绘画过程中。而谈到绘画，就不得不说起另外一个人的名字，叫做雷尼尔·盖马·弗里休斯(Reinerius Gemma Frisius)。这是一位来自荷兰的数学家，在一次观测日蚀的过程中，他绘制了第一幅描述小孔成像的插画，相应的也催生了“暗箱”这种光学设备的产生。

“暗箱”这中光学设备，反映了小孔成像模型带给人们的苦恼。一方面，为了得到清晰的像，人们不得不使得小孔的孔径尽可能小，以使得图11-2中的障碍物能够阻挡住尽可能多的光线，从而避免来自不同物体点的光线在幕布上相互干扰。另一方面，由于到达幕布的光线太少，使得所得到的像尽管逐渐变清晰，但是其强度却在不断下降，从而导致一定要在暗箱中才能够看到物体的像。可是，有趣的是，由于光的波粒二象性，当孔径非常小时会导致光的衍射现象，于是得到像不仅强度低了，而且也不会变清晰，反而会模糊起来。这一过程被图11-3形象的描绘了出来。

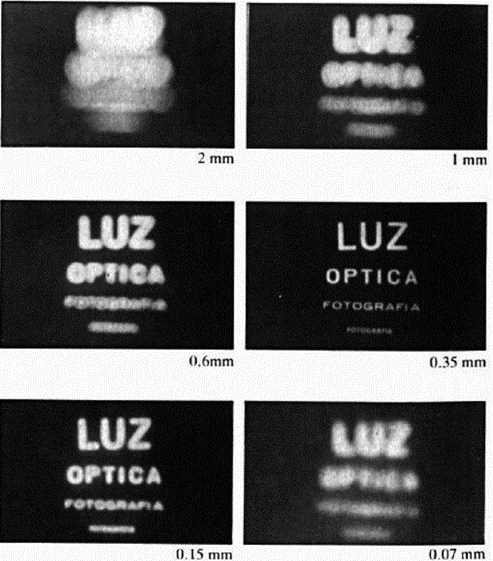


图11-3 针孔大小对图像的影响

11.1.2 薄透镜模型

与针孔相比，透镜可以更好地控制光线的折射和聚焦，因此可以在成像过程中解决清晰度和亮度之间的冲突。相机中的透镜系统根据需要使用不同的镜片来控制光线，以使其在胶片上形成清晰、明亮的图像。

当使用合适大小的透镜来取代针孔进行成像时，光线从物体顶端发出时会被透镜折射并聚焦到胶片上的一个点上。这是因为透镜能够根据其形状和折射率的不同，使光线的传播路径发生弯曲和折射。

透镜的大小和形状会影响成像的效果。对于理想的透镜模型来说，光线从物体不同的部分发出时，会经过透镜的不同区域，并根据透镜的形状和物体的位置进行折射。这一过程使得来自物体各个部分的光线都能在胶片上聚焦到一点上，形成清晰的图像。如图11-4所示。

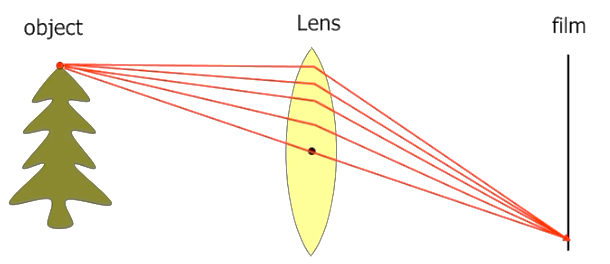


图11-4 透镜折射模型

在透镜模型中，平行于光轴的光线会被透镜聚焦在焦点处。因此，焦点是指距离透镜最近的一个点，它能够将平行于光轴的光线聚集在一起。同时，如果光线垂直于光轴，那么它将会直接通过透镜，不会被折射或偏转。

在透镜的物理模型中，我们通常会用到两个关键概念：焦距*f*和成像公式。焦距是指透镜将光线聚焦到该透镜的一点处所需要的距离。成像公式则是描述物体距离透镜、像距离透镜和透镜的焦距之间的关系。当光线传播方向穿过光学中心，它们的传播路径可以近似为一条直线，这意味着只需要在光线传播路线上确定两个点，即可使用成像公式计算出物体的位置和大小。

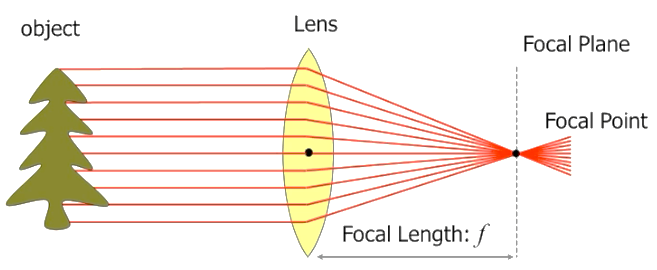


图11-5 透镜折射光线演示

薄透镜公式是用来描述薄透镜的成像性质的一个基本公式。它是基于几何光学的近似，适用于薄透镜、光线近轴传播以及物体和图像距离透镜远大于焦距的情况。薄透镜公式可以表示为：

wuju1

其中，*f*表示透镜的焦距，*e*表示图像距离透镜的距离，*z*表示物体距离透镜的距离。这个公式表明了物体和图像距离透镜的关系。当我们知道其中两个参数时，可以使用薄透镜公式来计算第三个参数。具体来说，如果物体距离透镜很远（*z*趋近于无穷大），这意味着焦距等于图像的距离。这种情况下，焦点在透镜的另一侧，图像也会在这一侧。薄透镜公式可以简化为：

wuju3

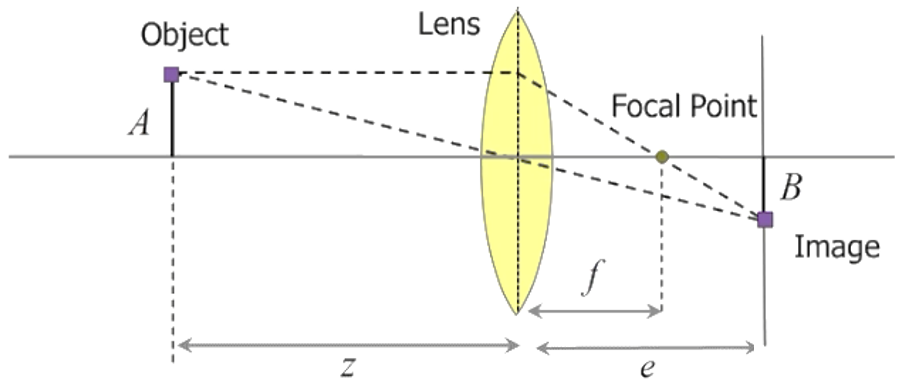


图11-6 薄透镜

薄透镜成像模型是一种简化的光学系统模型，它提供了一种基于几何光学近似的方便计算薄透镜成像的方法。通过掌握薄透镜成像模型，可以更好地理解现代相机、显微镜和其他光学系统的成像原理，以及对调节成像效果的参数进行优化。

11.1.3 透视投影模型

在接下来的部分，我们将详细讨论从3D世界到已知的数字图像投影建模所涉及的参数细节。虽然我们将使用针孔模型进行结果导出，但同样适用于近轴折射模型。如前所述，3D空间中的点*P*可以投影到图像平面Π'上的2D点*P'*，这个映射被称为投影变换。然而，3D点在图像平面上的投影并不直接对应其在实际3D空间中的位置。这是因为数字图像中的点通常处于与图像平面中的点不同的坐标系中。此外，数字图像被分为离散的像素，而图像平面中的点是连续的。还有一些非线性因素，如失真，被图像传感器引入。为了解释这些差别，我们将引入一些变换，以便将3D世界中的任意点映射到像素坐标平面上。

小孔成像模型是相机成像中最常用的模型之一。在这个模型中，物体的空间坐标和图像坐标之间存在线性关系，因此相机参数的求解可以简化为解线性方程组的问题。图11-7展示了世界坐标系、相机坐标系、图像坐标系和像素坐标系之间的关系。

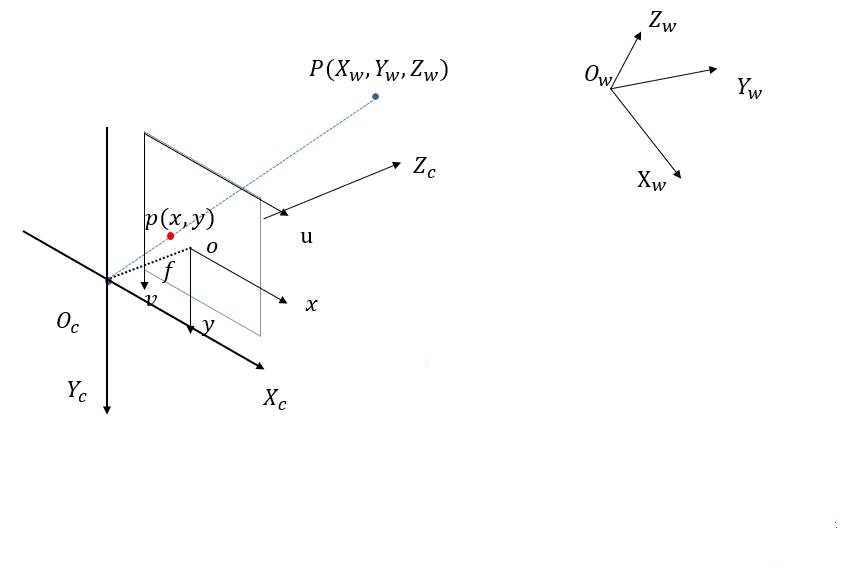


图11-7 不同坐标系对应位置联系

图像处理中涉及到以下四个坐标系：

*Ow-XwYwZw*：世界坐标系，是用来描述相机位置的坐标系，单位为米。它是客观三维世界的绝对坐标系，也被称为客观坐标系。由于数码相机位于三维空间中，所以需要世界坐标系作为参考，来描述数码相机的位置，以及描述其他物体在这个三维环境中的位置。用(*Xw, Yw, Zw*)来表示它们的坐标值。

*Oc-Xc Yc Zc*：相机坐标系，以相机光心为原点(在针孔模型中也即针孔为光心)，*z*轴与光轴重合，指向相机的前方(与成像平面垂直)，*x*轴与*y*轴的正方向与物体坐标系平行。其中，图中的*f*表示摄像机的焦距。

*o-xy*：图像坐标系，以相机光轴与像素坐标系的交点位置为坐标原点。图像坐标系表示为*o-x y*，并且单位为毫米。

*uv*：像素坐标系，以图像左上角为原点，单位为像素。*u*轴和*v*轴分别与图像坐标系的*x*轴和*y*轴平行，并用(*u, v*)表示其坐标值。相机采集到的图像一开始是以标准电信号的形式存在，然后通过模数转换转变为数字图像。每幅图像以一个*M×N*的数组的形式进行存储，其中*M*行*N*列的图像中的每个元素代表该图像点的灰度值。这样的每个元素称为像素，像素坐标系则是基于像素单位的图像坐标系。

在图中，*P*代表世界坐标系中的一个点，即真实3D世界中的点；*p*代表点*P*在图像中的成像点，在图像坐标系中的坐标表示为(x, y)，在像素坐标系中的坐标表示为(*u, v*)；*f*代表相机焦距，即*o*和*Oc*之间的距离。

世界坐标系到相机坐标系的转化：

从世界坐标系变化到相机坐标系属于刚体变换，即物体不会发生形变，只需要进行旋转和平移。如图11-8所示，*R*是一个3×3的旋转矩阵，*T*是3×1的平移向量。

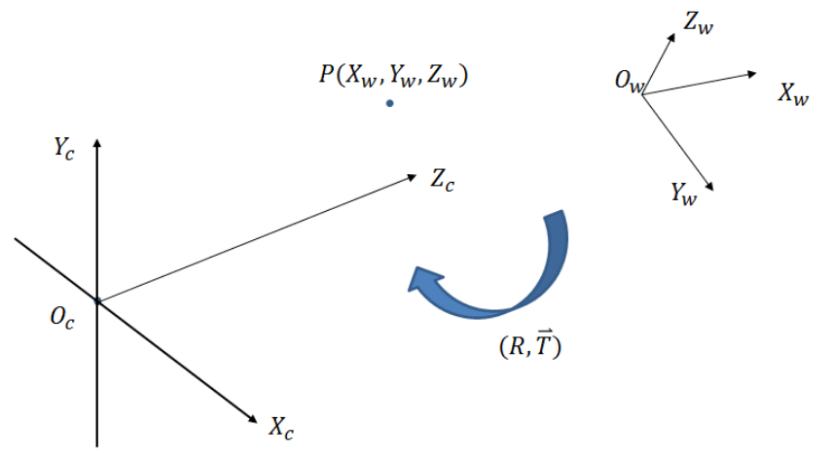


图11-8 世界坐标系到相机坐标系转化

由图可以得到*P*点在相机坐标系中的坐标为：





相机坐标系是指相机内部的坐标系，它是三维的，可以表示相机内的位置、方向和姿态等信息。图像坐标系是指相机传感器上的坐标系，它是二维的，表示了相机所拍摄图像的位置和分辨率等信息。为了实现从三维的相机坐标系到二维的图像坐标系的转化，需要进行坐标系变换。相机坐标系到图像坐标系的转化：两坐标系属于投射投影关系，从3D转化成2D，如图11-9所示。

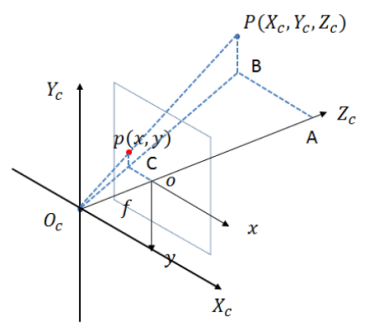


图11-9 相机坐标系与图像坐标系的转化

由于与可得：得出以下比例关系：



由上式可得：





图像坐标系与像素坐标系的转化：在进行两者关系确定之前，我们假设每个像素在*u*轴和*v*轴方向上的物理尺寸分别为*dx*和*dy*。值得注意的是，像素坐标系和图像坐标系都位于成像平面上，它们的原点和度量单位不同。图像坐标系的原点通常位于相机光轴与成像平面的交点，一般情况下该点位于成像平面的中心位置。图像坐标系的度量单位是毫米，属于物理尺寸单位，而像素坐标系的度量单位是像素，我们通常描述一个像素点的位置是通过行数和列数来表示的。

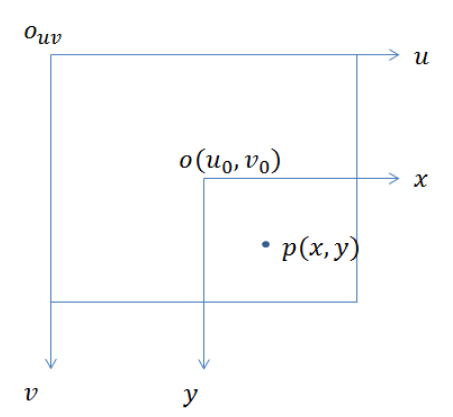


图11-10 图像坐标系与像素坐标系的转化

由图可求得所得到的公式，然后进行形式变换得到公式。





同时可以用另一种矩阵表示方式：



其中(*u0, v0*)是图像坐标系原点在像素坐标系中的坐标，*dx*和*dy*分别是每个像素在图像平面*x*和*y*方向上的物理尺寸。通过以上四个坐标系的转换可以得到一个点从世界坐标系转换到像素坐标系的公式，如下式所示。





其中，上述*K*表示相机的内参数，*M1*表示相机的外参数。相机的内参数是与相机自身特性相关的参数，例如焦距、像素大小等。而相机的外参数是在世界坐标系中描述相机位置和旋转方向等参数的方式。

相机畸变：是指相机成像时产生的系统性失真，包括径向畸变和切向畸变。径向畸变是由透镜非理想形状或透镜光学性能不均匀引起的一种畸变。在径向畸变中，光线在透镜表面附近的区域会更加弯曲，而远离透镜中心的区域相对较直。这会导致图像中心和边缘的比例变化，出现桶形畸变或枕形畸变。桶形畸变使图像中心部分相对于边缘部分显得缩小，而枕形畸变则使图像中心部分相对于边缘部分显得放大。

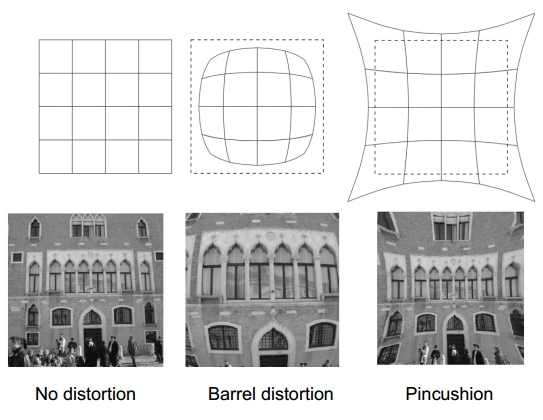


图11-11 径向畸变

切向畸变是由相机镜头装配时不平行于图像平面引起的一种畸变。当相机镜头与图像平面不完全平行时，会导致图像中的直线出现弯曲或倾斜的现象。

畸变可以用数学形式进行描述。在平面上，任意点*p*可以用笛卡尔坐标表示为[*x,y*]T，也可以用极坐标表示为[*r,*]T，其中*r*表示点*p*离坐标系原点的距离，表示与水平轴的夹角。径向畸变可以看作是坐标点沿着长度方向发生的变化，也就是其距离原点的长度发生变化。切向畸变可以看作是坐标点沿着切线方向发生的变化，也就是水平夹角发生变化。

对于径向畸变，不论是桶形畸变还是枕形畸变，它们都随着离中心的距离增加而增加。我们可以使用多项式函数来描述畸变前后的坐标变化。这类畸变可以通过与距中心距离相关的二次及更高次多项式函数进行校正：[*x,y*]T是未纠正的点，[*x1,y2*]T是纠正后的点。

555666777

在式描述的径向畸变纠正模型中，畸变程度较小的图像中心区域，主要由畸变系数k1来进行纠正。而对于畸变程度较大的边缘区域，主要由畸变系数k2来进行纠正。需要说明的是，这种简单的纠正方法对于一些复杂的畸变情况可能效果不佳，需要使用更高级别的畸变校正算法或更多的校正系数来纠正畸变。

切向畸变中使用m1和m2来进行纠正

555666777=============

因此，将式(11-1-13）和式（11-1-14)合并，对于相机坐标系中的一点P(X,Y,Z)，我们能够通过五个畸变系数找到这个点在像素平面上的正确位置：

(1)将三维空间点投影到归一化图像平面。设它的归一化坐标[x,y]T。

(2)对归一化平面上的点进行径向畸变和切向畸变纠正。

555hhhh

(3)将纠正后的点通过内参数矩阵投影到像素平面，得到该点在图像上的正确位置。

ee

11.2视觉系统搭建

机[器视觉](https://link.zhihu.com/?target=https://so.csdn.net/so/search?q=%E6%9C%BA%E5%99%A8%E8%A7%86%E8%A7%89&spm=1001.2101.3001.7020" \t "https://zhuanlan.zhihu.com/p/_blank)系统是综合现代计算机、光学、电子技术的高科技系统。机器视觉技术通过计算机对系统摄取的图像进行处理，分析其中的信息，并做出相应的判断，进而发出对设备的控制指令。机器视觉系统的具体应用需求千差万别，视觉系统本身也可能有多种不同的形式，但都包括以下流程:

(1)图像采集利用光源照射被观察的物体或环境，通过光学成像系统采集图像，通过相机和图像传感器将光学图像转换为[数字图像](https://link.zhihu.com/?target=https://so.csdn.net/so/search?q=%E6%95%B0%E5%AD%97%E5%9B%BE%E5%83%8F&spm=1001.2101.3001.7020" \t "https://zhuanlan.zhihu.com/p/_blank)，这是机器视觉系统的前端和信息来源。

(2)图像处理和分析计算机通过图像处理软件对图像进行处理，分析获取其中的有用信息。如PCB板的图像中是否存在线路断路、纺织品的图像中是否存在疵点、文档图像中存在哪些文字等。这是整个机器视觉系统的核心。

(3)判断和控制图像处理获得的信息最终用于对对象（被测物体、环境）的判断，并形成对应的控制指令，发送给相应的机构。如摄取的零件图像中，计算零件的尺寸是否与标准一致，不一致则发出报警，做出标记或进行剔除。

在整个过程中，被测对象的信息反映为图像信息，进而经过分析，从中得到特征描述信息，最后根据获得的特征进行判断和动作。最典型的机器视觉系统一般包括: 光源、镜头、光学成像系统、相机、图像采集卡、机器视觉软件、通信协议。

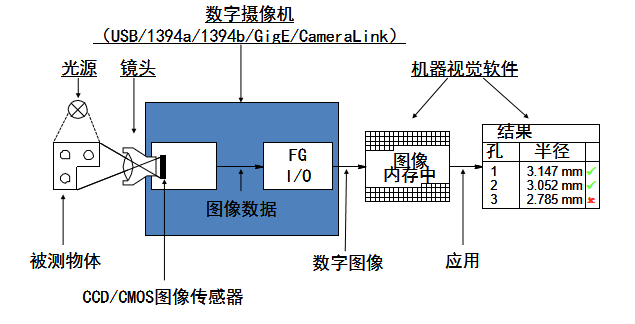


图11-12 视觉系统组成图

11.2.1光源

视觉系统中光源是决定成像系统成像质量的最主要因素之一，它对于图像质量的影响是直接的。光源提供的光照强度、光照方式、光照角度等因素都会影响获得的图像的清晰程度、对比度和亮度等视觉特性，进而影响系统的性能及实时性。



图11-13 各种光源

不同类比照明光源有其各自的优点和缺点，以下是它们的对比：

(1)白炽灯：白炽灯是一种常见的光源，它的价格低廉，可以提供温暖舒适的光照。但它的寿命短，能效低，也容易产生热量和紫外线。

(2)荧光灯：荧光灯比白炽灯更节能，且寿命更长。但它的颜色较为单调，不能很好地还原物体的真实颜色，也存在闪烁的问题。

(3)LED灯：LED灯具有高能效、长寿命、多样化的颜色和渐变效果等优点，可制成各种形状、尺寸及各种照射角度；可根据需要制成各种颜色,并可以随时调节亮度；通过散热装置，散热效果更好，光亮度更稳定;·使用寿命长（约3万小时，间断使用寿命更长）；反应快捷，可在10us或更短的时间内达到最大亮度。同时还可以方便地集成和控制。与其它光源相比，它的成本和初始投资较高，但随着技术的发展和普及，其成本也在逐渐降低。

在具体的应用中进行光源选择，需要针对不同的应用场景和需求，选择合适的照明光源至关重要。在实际应用中，通常需要根据实际情况综合考虑其成本、能效、寿命、颜色还原效果和光污染等因素来选择适合的照明光源。

光源的光照方式也会影响图像的质量。例如，在图像采集中，均匀光照可有效减少图像中的噪点和阴影，从而提高图像的清晰度和质量。此外，根据具体应用场景的不同。因此，在选择合适的光源时需要考虑到应用场景，选择适合的光照方式。当涉及到照明技术时，有几种常见的光照方式，包括：

前向光漫射照明：前向光漫射照明是将光源聚焦在被观察物体的前方，并以一个特定的角度照射到物体表面上。部分入射光被物体表面散射或反射，从而使物体的表面特征，如凹凸、纹理等更加清晰可见。这种方式常用于强调物体的表面形状和纹理。

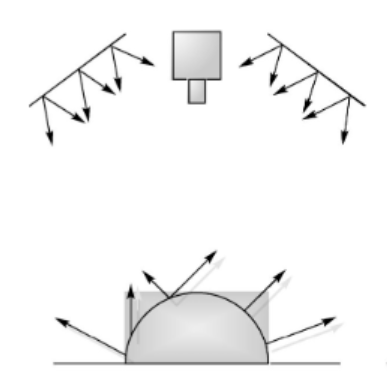


图11-14 前向光漫射照明

背光照明：背光照明是将光源放置在被观察物体的背后，以透过物体并照亮物体。它在物体与背景之间产生对比度，显示出物体的边缘和轮廓。背光照明常用于检测物体的不透明度、轮廓和透光性等特征。

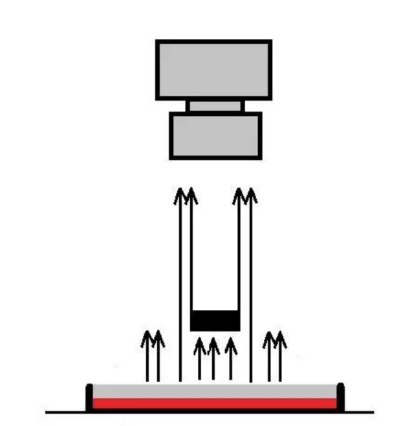


图11-15 背光照明

同轴光照明：同轴光照明是将光源放置在镜头的正后方，以与被观察物体的视线方向一致。这种方式可以消除或减轻物体表面的反射，并提供均匀的光照，以显示物体的表面细节。同轴光照明常用于检测光滑表面或需要消除反射的应用。

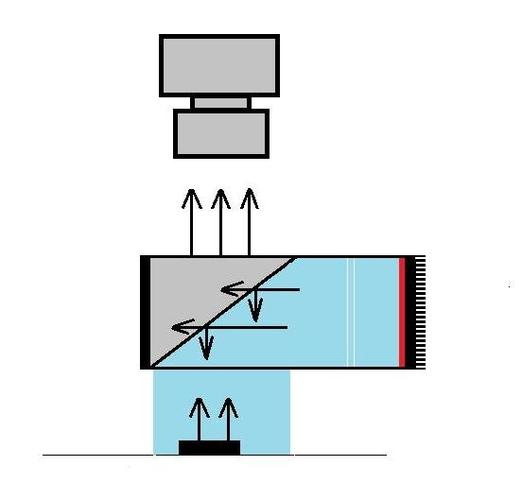


图11-16 同轴光照明

11.2.2镜头

如果将机器视觉系统与人类视觉系统进行类比，那么相机的传感器芯片就如同人的视网膜，而镜头则相当于眼睛内的晶状体。各种现实世界中的图像都通过这个“晶状体”对光线进行变换（汇聚）后，投射在“视网膜”上。在机器视觉系统中，镜头常和相机作为一个整体出现，它的质量和技术指标直接影响成像子系统的性能，合理地选择和安装镜头是决定机器视觉成像子系统成败的关键。

机器视觉成像系统使用的镜头通常由凸透镜和凹透镜结合设计而成。单个凸透镜或凹透镜是进行光束变换的基本单元。凸透镜可对光线进行汇聚，也称为会聚透镜或正透镜。凹透镜对光线具有发散作用，也称为发散透镜或负透镜。两种透镜成像均遵循高斯成像公式，通过把它们结合使用，在校正各种像差和失真后，设计出具有不同结构和技术指标的复合镜头系统。与镜头相关的主要技术参数有镜头分辨率、焦距、最小工作距离、最大像面、视场／视场角、景深、光圈和相对孔径及其安装接口类型等。

在机器视觉系统中，镜头的主要作用是将目标成像在图像传感器的光敏面上。镜头的质量直接影响到机器视觉系统的整体性能，合理地选择和安装镜头，是机器视觉系统设计的重要环节。

镜头分辨率：是指镜头在空间上能够分辨的最小细节。常用的方法是使用正弦光栅来测试。从信号处理的角度来看，任何非周期图像信号都可以被看作周期图像的叠加，而周期图像则可以被分解为亮度按正弦变化的图形的叠加。因此，通过研究镜头对亮度按正弦变化图形的反应，可以研究镜头的性能和分辨力。正弦光栅是一种亮度按正弦变化的图像，其中黑白相间的栅格被定义为一个线对（*lp*），它所占据的长度是正弦光栅的空间周期，单位是毫米。正弦光栅的空间周期倒数就是空间频率，表示每毫米内的线对数，单位是线对/毫米(*lp/mm*)。通过拍摄正弦光栅并研究每毫米内可分辨的线对数，可以了解镜头的分辨率。镜头的分辨率越高，它每毫米内能够分辨的线对数就越多。机器视觉系统设计时，只需查询镜头参数表即可获得其分辨率信息。



图11-17 正弦光栅

焦距是指无限远处目标在镜头像方成像位置到像方主面的距离。焦距决定了不同物距上目标的成像位置和成像大小。常见的镜头焦距包括6mm、8mm、12.5mm、25mm和50mm等。对机器视觉成像系统来说，工作距离就是物距。视觉系统模型假设工作距离相对于焦距为无限远。因此，镜头的产品参数中通常会说明最小工作距离。当相机在小于最小工作距离的环境下运作时，会出现图像失真，影响机器视觉系统的可靠性。

最大像面、视场／视场角：都是用来衡量镜头成像范围的关键参数。最大像面是指镜头能支持的最大清晰成像范围（常用可观测范围的直径表示），超出这个范围所成的像对比度会降低而且会变得模糊不清。最大像面是由镜头本身的特性决定的，它的大小也限定了镜头可支持的视场的大小。

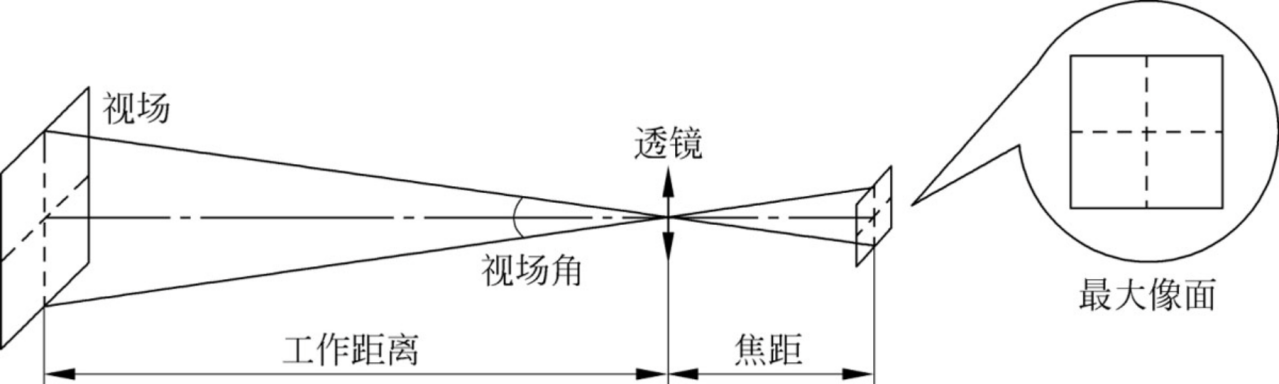


图11-18 机器视觉系统中镜头的视场和最大像面

景深：也是一个与镜头和成像系统关系十分密切的参数，它是指在镜头前沿着光轴所测定的能够清晰成像的范围。在成像系统的焦点前后，物点光线呈锥状开始聚集和扩散，点的影像沿光轴在焦点前后逐渐变得模糊，形成一个扩大的圆，这个圆称为弥散圆。若这个圆形影像的直径足够小（离焦点较近），成像会足够清晰，如果圆形再大些（远离焦点），成像就会显得模糊。当在某个临界位置所成的像不能被辨认时，则该圆就被称为容许弥散圆。焦点前后两个容许弥散圆之间的距离称为焦深。在目标物一侧，焦深对应的范围就是景深。

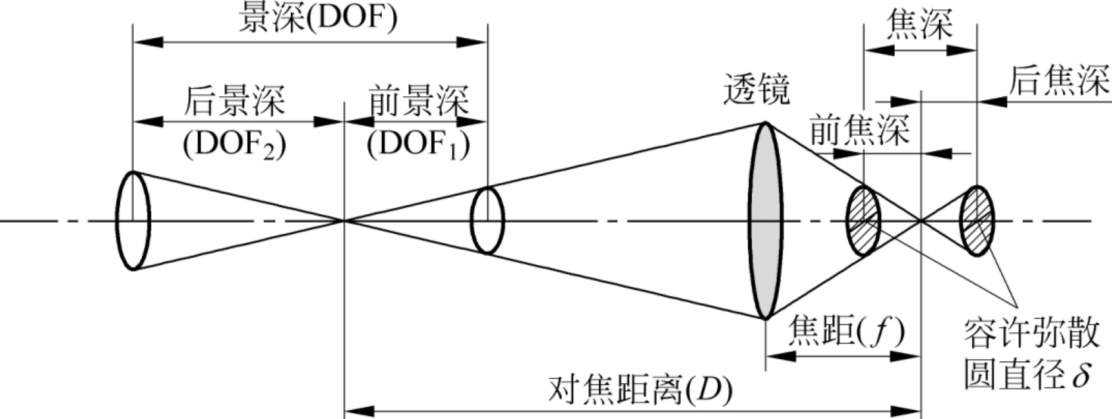


图11-19 景深示意图

如果把观测物平面和光轴看作是三维坐标系（*x，y，z*），则视场在（*x，y*）平面上限定了观测范围，而景深则从*z*方向上确定了可清晰观测的范围。与景深相关的计算公式如下：







其中，δ为容许弥散圆的直径，*f*为镜头焦距，*D*为对焦距离，*F*为镜头的拍摄光圈值。光圈值*F*常用镜头焦距和镜头入瞳的有效直径*Din*的比值来表示，它是镜头相对孔径*Dr*的倒数，即：





从景深公式可以看出，后景深要大于前景深，而且景深一般随着镜头的焦距、光圈值、对焦距离（可近似于拍摄距离）的变化而变化。在其他条件不变时：

1）光圈越大（光圈值*F*越小），景深越小；光圈越小（光圈值*F*越大），景深越大。

2）镜头焦距越长，景深越小；焦距越短，景深越大。

3）距离越远，景深越大；距离越近，景深越小。

接口：镜头与相机之间的物理接口必须匹配才能安装在一起搭配使用。常见的接口标准有*C*接口（C-mount）、*CS*接口（CS-mount）和*F*接口（F-mount）。在机器视觉领域，目前*C*和*CS*接口的镜头及相机占主导地位，它们的唯一区别是背焦距不同，如图11-20所示。*F*接口常用于高像素数的线扫描相机（2048像素以上），获取比*C*和*CS*接口镜头更大的图像。*C*接口镜头的背焦距是17.526mm，*CS*接口镜头背焦距则为12.5mm，因此，只要为*C*接口镜头配备一个5mm的扩展管（转换器），就可以得到*CS*接口的镜头，但*CS*镜头却不能与*C*接口的相机搭配使用。*C*接口是镜头的国际标准，因此有很多*C*接口的镜头可供选择。关于*C*接口镜头和相机，有一个重要例外，就是*C*接口的3CCD相机不能和*C*接口镜头协同工作，因此需要查阅相机供应商提供的镜头兼容性列表。

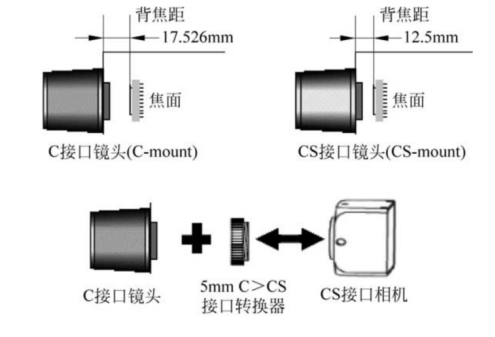


图11-20 C与CS接口

11.2.3成像芯片

视觉系统的成像芯片通常指的是图像传感器，它是摄像机、智能手机、监控摄像头等电子设备中的核心部件之一。常用的主要有CCD（Charge-Coupled Device）芯片和CMOS（Complementary Metal-Oxide-Semiconductor）芯片都是数字图像传感器（Image Sensor）的类型，用于捕捉光学图像并转换为电信号。

CCD芯片是较早的技术，它使用一系列光电二极管和电荷耦合器件来收集和传输光信号。CCD图像传感器的工作原理是将光子转化为电子，并将电荷耦合器件中的电荷顺序传送到芯片的边缘，然后进行转换和放大以形成数字图像。CCD实际上可以被看作由多个MOS（Metal Oxide Semiconductor）电容组成。在*P*型单晶硅的衬底上通过氧化形成一层厚度约为100~150nm的SiO2绝缘层，再在SiO2表面按一定层次蒸镀一层金属或多晶硅层作为电极，最后在衬底和电极间加上一个偏置电压（栅极电压），即可形成一个MOS电容器，如图11-21所示。

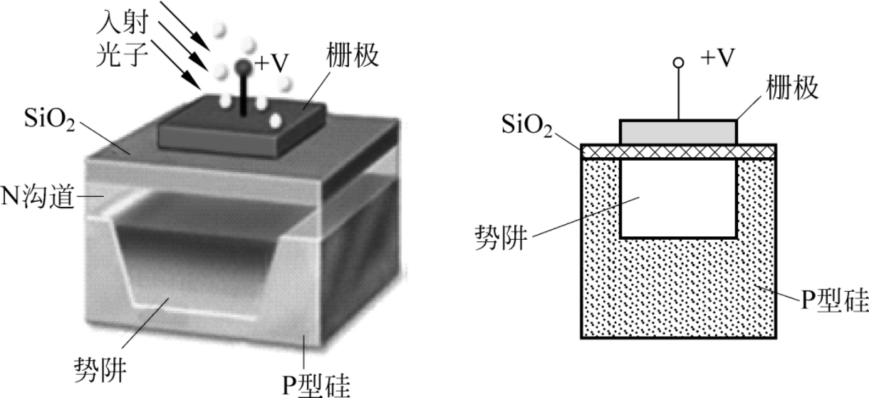


图11-21 C与CS接口

当光线投射到MOS电容上时，光子穿过多晶硅电极及SiO2层，进入*P*型硅衬底，光子的能量被半导体吸收，产生电子空穴对，产生的电子立即被吸引并储存在势阱中。由于势阱中电子的数量随入射光子多少（即光强度或亮度）增加而增加，而且即使停止光照，势阱中的电子当栅极电压未产生变化时在一定时间内也不会损失，这就实现了光电转换和对光照的记忆存储功能。由此，在*P*型轨衬底上生成多个MOS电容，来制作以其为基本单元的图像传感器。例如，可以将多个MOS电容排列成一行或点阵来扫描或抓取外部图像。考虑到集成芯片尺寸较小，因此可以结合透镜成像的特点，对场景所成的实像进行采集，这就是CCD相机的雏形，如图11-22所示。

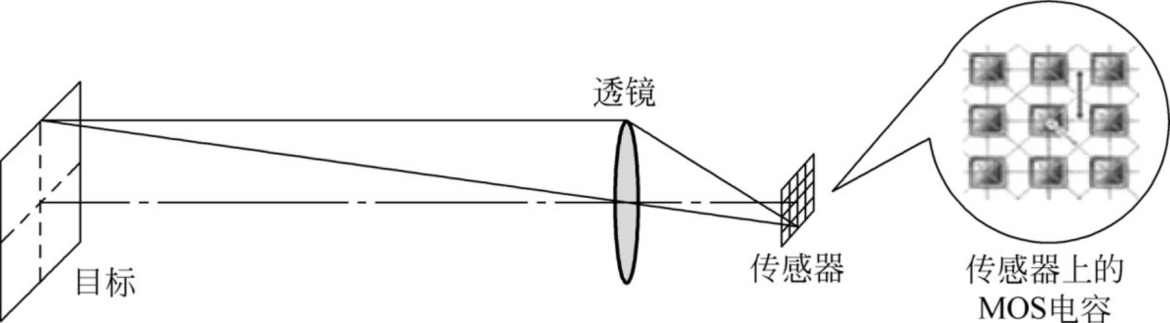


图11-22 CCD相机雏形

CMOS（Complementary Metal Oxide Semiconductor）的开发最早出现在20世纪70年代初。20世纪90年代初期，随着超大规模集成电路（VLSI）制造工艺技术的发展，CMOS得到迅速发展。CMOS的光电转换原理与CCD相同，二者的主要差异在于电荷的转移方式上。CCD中的电荷会被逐行转移到水平移位寄存器，经放大器放大后输出。由于电荷是从寄存器中逐位连续输出的，因此放大后输出的信号为模拟信号。在CMOS传感器中，每个光敏元的电荷都会立即被与之邻接的一个放大器放大，再以类似内存寻址的方式输出（图11-23），因此CMOS芯片输出的是离散的数字信号。之所以采用两种不同的电荷传递方式，是因为CCD是在半导体单晶硅材料上集成的，而CMOS则是在金属氧化物半导体材料上集成的，工艺上的不同使得CCD能保证电荷在转移时不会失真，而CMOS则会使电荷在传送距离较长时产生噪声，因此使用CMOS时，必须先对信号放大再整合输出。CCD传感器芯片将电荷转换为模拟信号，再经放大、A/D转换后才以数字信号形式输出。

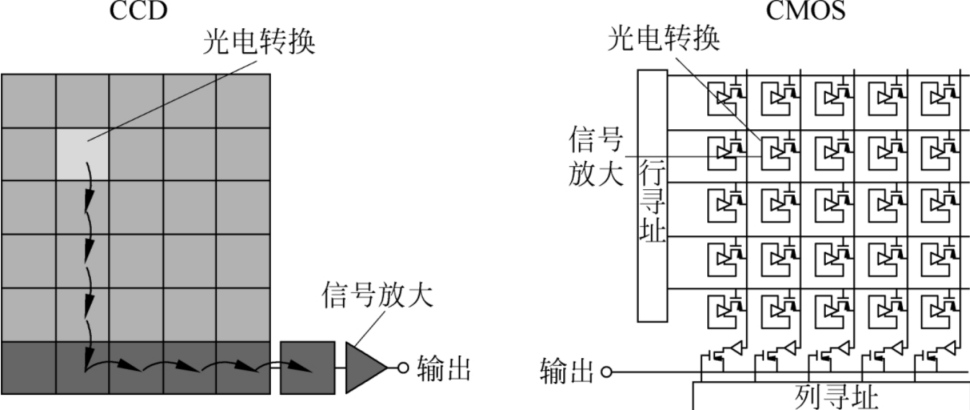


图11-23 CCD与CMOS传感器的信号转移方式

三个关键的要素决定了成像芯片的选择：动态范围、速度和响应度。动态范围决定系统能够抓取的图像的质量，也被称作对细节的体现能力。芯片的速度指的是每秒钟芯片能够产生多少张图像和系统能够接收到的图像的输出量。响应度指的是芯片将光子转换为电子的效率，它决定系统需要抓取有用的图像的亮度水平。芯片的技术和设计共同决定上述特征，因此系统开发人员在选择传感器时必须有自己的衡量标准，详细的研究这些特征，将有助于做出正确的判断。

(1)CMOS传感器芯片直接将每个电荷放大后转换为数字信号输出，往往成像一致性差。

(2)CCD在电荷转移过程中不会失真，且信号统一放大后才输出，因此成像质量和一致性高。

(3)CCD传感器有更大的填充因子和更高的信噪比，对光更加敏感，更适应低对比度的场合。

(4)CMOS传感器可以获得比CCD传感器高很多的图像传输速度，更适用于高速场合。

(5)CMOS传感器的信号经过放大后才进行转移，所以它的功耗要比CCD低，更适用于应于便携设备。

表11-2 CCD和CMOS的比较

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 特 点 | CCD | CMOS | 性 能 | CCD | CMOS |
| 像素输出信号 | 电荷 | 电压 | 制造工艺 | 复杂 | 简单 |
| 芯片输出信号 | 模拟 | 数字 | 动态范围 | 高 | 中 |
| 填充因子 | 高 | 中 | 像素一致性 | 高 | 低 |
| 噪声 | 低 | 中高 | 光灵敏度 | 高 | 中 |
| 分辨率 | 低 | 高 | 低对比度适应性 | 强 | 弱 |
| 功耗 | 中高 | 低 | 传输速度 | 中高 | 高 |
| 成本 | 中 | 中低 |  |  |  |

11.2.4通讯协议

在机器视觉中，常用的通信协议包括以下几种：

(1)RS-422：RS-422属于低压差分信号（Low Voltage Differential Signalin, LVDS）它们用两根电压相反的线同时传输信号，在接收端通过两根线上的信号强度差来得到最终的数据。信道中同时出现在两根线上的噪声因相减后被去除，因此信号的电压要求低，抗噪声能力强，传输距离远。例如RS-422的最远传输距离可达1200m。

(2)RS-644（常称为LVDS）信号电压为350mV，比RS-422更低（RS-422为2.4V）。它进一步降低了功耗（仅为RS-422的1/8）和噪声，传输速率最高可达400Mb/s。RS-422及RS-644通常使用68针或100针的高密度物理接口，但各相机厂家的引脚定义不尽相同（对应采集卡引脚定义也不完全一样），如果要更换相机，必须确认接口各引脚信号的定义与更换之前的定义相同，必要时还要做信号转换板。

(3) Camera Link：它提供了高带宽和低延迟的数据传输，适用于高速图像采集和处理的应用。Channel Link是Camera Link的前身，它在LVDS基础上对并行数据传输进行了串行化，降低了传输电缆的数量。虽然Channel Link接口的引脚较LVDS少了很多，但是由于它没有对接头形式做统一约定，因此各厂家的产品之间依然有差异。使用户可以从图像卡发送控制指令，以编程的方式控制摄像机。

(4) USB 2.0：USB 2.0是一种基于USB接口的通信协议，用于连接相机和计算机。它提供了简单、便捷的连接方式，并支持高速数据传输和图像采集。USB总线连接的系统是一种用于点对点通信的主从系统，其目的是作为一种通用标准来取代现有的各种串行或并行的计算机I/O协议。USB接口可热拔插，连接方便，可以连接设备数多达127个，两个设备之间最长通信距离为5m。USB 2.0向下兼容USB 1.1、USB 1.0，数据的传输率达到120Mb/s~480Mb/s，可完全满足工业图像传输的速度需求。

这些通信协议在机器视觉领域中广泛应用，可以满足不同应用场景下的需求，实现图像采集、传输和处理的高效率和稳定性。

表11-3 通信协议的比较

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 接口标准 | RS-422 | LVDS(RS-644) | Camera-Link | USB 2.0 |
| 连接器 | 68/100pin | 68/100pin | 26pin | USB |
| 传输速率 | 10Mb/s | 400Mb/s | 最高3.6Gb/s | 480Mb/s |
| 传输距离 | 1200m | 最大35m | 7~10m | 3~7m |
| 设备数 | 10 | 1 | 1 | 最多127 |
| 拓扑结构 | 到采集卡 | 到采集卡 | 点对点 | 主/从 |

11.2.5 视觉系统的选型过程

图像分辨率、系统分辨率和像素分辨率是机器视觉系统设计时较常见的参数，它们通常与客户对机器视觉系统的需求关系最为密切，是选择相机和镜头的重要依据。

图像分辨率指图像中存储的信息量，是每英寸图像内有多少个像素点，分辨率的单位为PPI，通常叫做像素每英寸。图像分辨率一般被用于PS中，用来改变图像的清晰度。

系统分辨率指成像系统可以识别出监测目标的最小细节或最小特征。诸如“要求系统能检测0.1mm的目标"、"要求系统测量精度达到0.01mm"之类的要求一般都和系统分辨率相关。

像素分辨率指为了表示检测目标所需要的像素数。一般情况下，可以根据客户对检测目标中最小特征的要求来确定最小像素分辨率。如果将整个图像看作周期为最小特征大小的周期信号，则根据奈奎斯特采样定律，必须对信号每个周期采样2个点以上，才能完整恢复该信号。因此如果客户没有特别要求，常用至少两个像素来代表检测目标中的最小特征，这可被看作是图像传感器的奈奎斯特定律。

图像传感器应具备的最小像素分辨率常通过下面的公式计算：

rrerewr

式中：*Rmin*为最小像素分辨率。*Lmax*为检测目标的最大长度。*lmin*为检测目标的最小特征长度（视觉系统的分辨率)。*pmin*为表示最小特征的像素数。在无特别要求时，*pmin*=2，如果客户要求使用多于2像素来表示最小特征，则最小分辨率将适当增加。

视场（Field of View，FOV）指成像系统中图像传感器可以监测到的最大区域。在机器视觉系统设计时，考虑到一般都会使被检测目标尽量填满整视场，因此常用视场大小代替目标的最大长度Lmax来计算视觉系统的像素分辨率。

如果横纵方向上视场大小为[FOVh，FOVv]，检测目标的最小特征的大小为[*Ih*，*lv*]，则图像传感器应具有的最小像素分辨率为：

wwwww

成像系统视场的大小可以通过研究其成像规律得知。目前，机器视觉系统常用使用配备各种镜头系统的工业CCD/CMOS相机作为成像系统，镜头系统一般使用透镜系统，其成像遵循高斯成像公式：

wwwww4242

通常将像距与物距的比值定义为透镜的放大率M：

5656

数字图像是对成像系统输出的信号进行数字化后的结果，成像系统反映真实场景的性能和质量直接决定整个机器视觉系统的性能。

影响机器视觉成像系统成像质量的因素主要包括：光源（Illumination）、系统分辨率（System Resolution）、像素分辨率（Pixel Resolution）、对比度（Contrast） 、景深（Depth of Field，DOF）、投影误差（Perspective Error）和镜头畸变（Lens Distortion）。而这些因素（参数）却直接或间接地由硬件选型和安装方式决定。

工业或研究领域的成像系统多种多样，常见的有工业CCD/CMOS相机、工业显微镜、生物显微镜、X射线成像仪、红外成像仪、热成像仪等。无论这些成像系统的原理有多复杂，都可被抽象为下图所示的简化模型。如下图11-24。

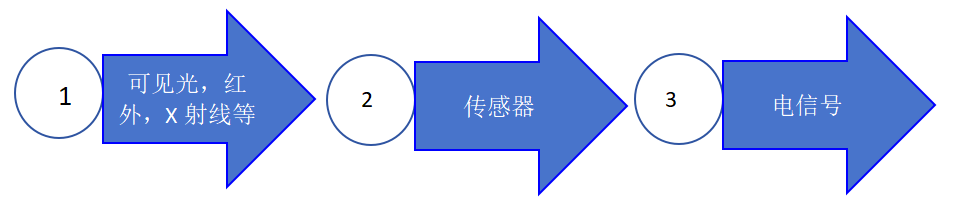


图11-24 简化模型

千差万别的成像系统对现实世界中的可见光、红外、X射线、热量等实施某种转换T (*x，y*)，将物理量转换为电信号，再经图像采集设备采样、量化后生成数字图像。

镜头系统简化模型，由于对机器视觉成像系统来说，相机镜头到所检测目标的距离（称为工作距离，相当于物距）相对于相机焦距可近似认为是无穷远。若将其带入高斯成像公式，可得出此时相机像距近似等于其焦距，也就是说相机成像在焦平面上。据此，可以将镜头系统抽象为类似小孔成像的简化模型，如下图所示：

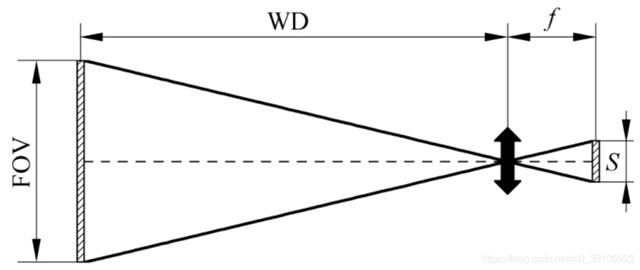


图11-25 小孔成像的简化模型

根据该简化模型，可以得出机器视觉系统图像传感器尺寸S（传感器平面某个方向上的长度)、视场FOV、工作距离WD及镜头焦距f之间的约束关系：

8989898

此时，透镜的放大率M则可以等效为：

545

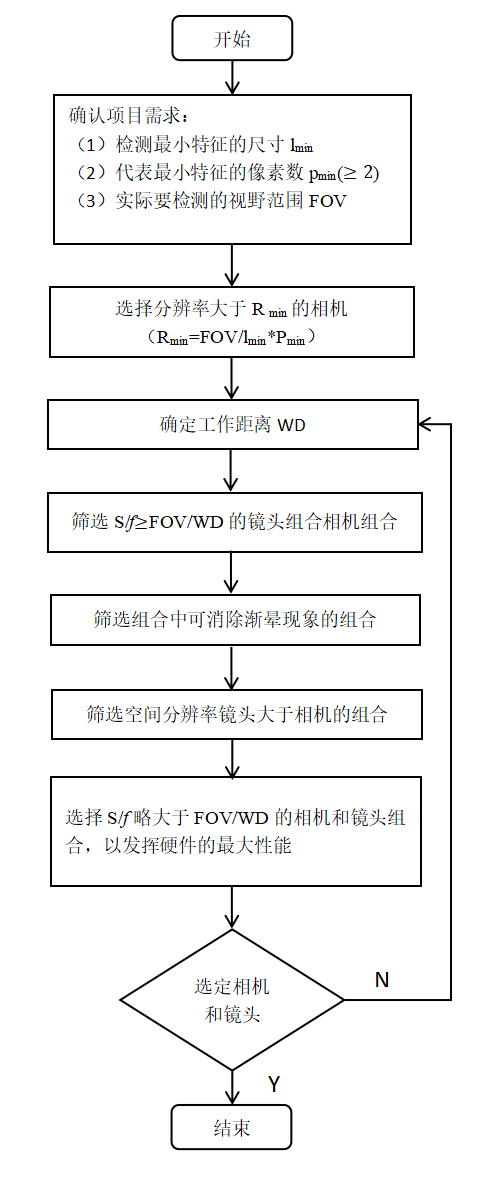
如果进一步将前述最小像素分辨率的计算公式与该约束关系结合（用视场FOV代替目标的最大长度*Lmax*)，则可以得出以下成像系统简化模型的参数约束关系：

545aadsda

该公式所显示的参数间的相互约束关系是机器视觉系统设计和搭建部署时系统设计和设备选型的基础。

实际中传感器尺寸S可以通过查询相机的技术规范获知，焦距*f*、工作距离WD直接由所选择的镜头决定。在已知这些参数时，可以很容易地计算出视场FOV。相机的像素分辨率由其有效像素区域（即传感器尺寸）决定，通常用横向和纵向有效像素数来表示（如768×576）。为机器视觉系统所选择的相机像素分辨率，必须大于或等于按照项目需求（包括对最小特征尺寸*lmin*和用于表示最小特征的像素数*pmin*的要求）计算出的最小像素分辨率*Rmin*。下图进一步显示了公式中涉及的参数含义：

工业机器视觉系统中，镜头与相机的选型非常重要。如果事先既未确定相机又未确定镜头，则需要先了解项目工作环境对相机安装（工作距离）、要检测的最大范围(视场)、最小特征的尺寸和代表它的像素数的要求，然后根据这些条件来计算应使用何种镜头或相机。下图为机器视觉项目选择镜头和相机的简化流程：



成像系统简化模型的参数约束关系：

wrg

*Lmin*：检测最小的特征尺寸

*Pmin*：代表最小特征的像素数

FOV：实际要检测的视野范围

FOVmin：最大视场

*Rmin*：相机的最小分辨率

WD：工作距离

*S*：传感器尺寸

*f*：镜头焦距

图11-26 选择镜头和相机的简化流程

11.3 视觉系统标定

视觉系统搭建完毕后，视觉系统标定是确保视觉系统能够准确测量和识别对象的关键步骤。相机标定指建立相机图像像素位置与场景点位置之间的关系，根据相机成像模型，由特征点在图像中坐标与世界坐标的对应关系，求解相机模型的参数。相机需要标定的模型参数包括内部参数和外部参数。相机标定方法有：传统相机标定法、主动视觉相机标定法、相机自标定法，如图11-24所示。

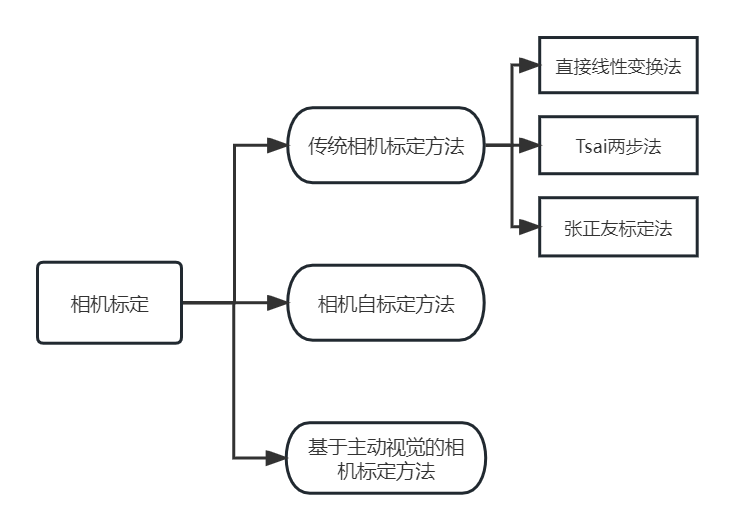


图11-27 相机标定方法

11.3.1 常用的标定方法

相机标定方法有：传统相机标定法、相机自标定法、主动视觉相机标定方法。

(1)传统相机标定法

该法需要使用尺寸已知的标定物，通过建立标定物上坐标已知的点与其图像点之间的对应，利用一定的算法获得相机模型的内外参数。根据标定物的不同可分为三维标定物和平面型标定物。三维标定物可由单幅图像进行标定，标定精度较高，但高精密三维标定物的加工和维护较困难。平面型定标物比三维标定物制作简单，精度易保证，但定标时必须采用2幅或2幅以上的图像。Tsai于1987年首先提出利用共面点的两步定标算法，为平面型标定物的应用奠定了基础。张正友教授也对基于网格的平面型标定物的定标方法进行了研究。张正友教授假设平面型标定物在世界坐标系中 ZW=0，通过线性模型分析计算得出相机参数的优化解，然后用基于最大似然法进行非线性求精。传统相机标定法在标定过程中始终需要标定物，且标定物的制作精度会影响标定结果。同时有些应用场合不适合放置标定物也限制了传统相机标定法的应用。

(2)相机的自标定法

目前出现的自标定算法中主要是利用场景约束或者利用相机运动约束。相机的运动约束条件太强，因此使得其在实际中并不实用。利用场景约束主要是利用场景中的一些平行或者正交的信息。其中空间平行线在相机图像平面上的交点被称为消失点，它是射影几何中一个非常重要的特征，所以很多学者研究了基于消失点的相机自标定方法。自标定方法灵活性强，可对相机进行在线定标。但由于它是基于绝对二次曲线或曲面的方法，其算法鲁棒性差。

(3)基于主动视觉的相机标定法

指已知相机的某些运动信息对相机进行标定。该方法不需要标定物，但需要控制相机做某些特殊运动，利用这种运动的特殊性可以计算出相机的内部参数。如 Hartley提出了一种使相机作纯旋转运动来对相机进行标定的方法。基于主动视觉的相机标定法的优点是算法简单，往往能够获得线性解，故鲁棒性较高，缺点是系统的成本高，实验设备昂贵，实验条件要求高，而且不适合于运动参数未知或无法控制的场合。

11.3.2 单目视觉系统标定

张正友相机标定法是一种常用的单目视觉标定方法。为了克服传统方法的局限性，张正友相机标定法是张正友博士提出的一种基于2D平面靶标的相机标定方法。该方法使用单平面棋盘格作为标定板，在拍摄多张标定板的图片后，通过将实际世界中的点(世界坐标)和图像上的点(像素坐标)进行一一对应，可以求解出世界坐标和像素坐标之间的对应关系。传统的相机标定方法通常使用三维标定板，需要非常精确的制作且难度较大。而张正友相机标定方法则介于传统标定法和自标定法之间，克服了传统标定法对高精度标定物的要求。它只需要使用一个打印出来的棋盘格作为标定板，而且该棋盘格的角点坐标是已知的，如图11-28所示。相比于自标定法，张正友相机标定法提高了精度并且更易操作。因此，张正友相机标定法被广泛应用于计算机视觉领域。

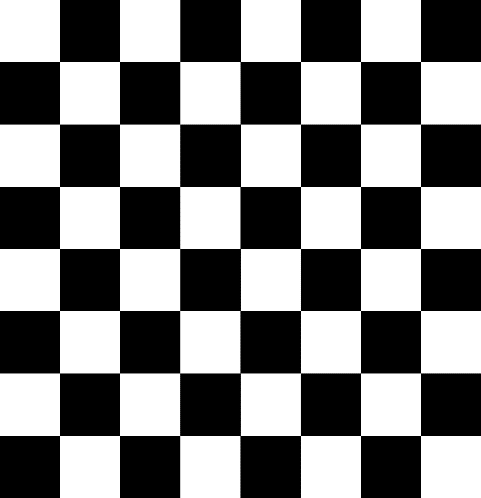


图11-28 标定所需棋盘格

张正友相机标定法中确定相机内参矩阵和外参矩阵的思路如下：先求解内参矩阵与外参矩阵的乘积。再求解内参矩阵，最后求外参矩阵。求解内参矩阵与外参矩阵的积，将世界坐标系固定在棋盘格上，则棋盘格上任一点的世界坐标中*Z=0*，因此，原单点无畸变的成像模型可以化为下式。其中，*R*1，*R*2为旋转矩阵的前两列。为了简便，将内参矩阵记为*A*。



对于不同的图片，内参矩阵*A*为定值；对于同一张图片，内参矩阵*A*，外参矩阵[*R1, R2, T*]为定值；对于同一张图片上的单点，内参矩阵*A*，外参矩阵[*R1, R2, T*]，尺度因子*Z*为定值。将A[*R1, R2, T*]记为矩阵*H*，*H*即为内参矩阵和外参矩阵的积，记矩阵*H*的三列分别为[*H1, H2, H3*]，则有：



利用上式，消去尺度因子*Z*，可得到：



此时，由于尺度因子*Z*已经被消除，因此上述方程适用于同一张图像上的所有标定板角点。(*u, v*)表示像素坐标系下的标定板角点坐标，(*Xw, Yw*)表示世界坐标系下标定板角点的坐标。通过图像识别算法，我们可以获取标定板角点的像素坐标(*u, v*)。由于标定板的世界坐标系是已知的，标定板上每个格子的大小也已知，我们可以计算出世界坐标系下的(*Xw, Yw*)。通过观察方程，我们可以看出矩阵H是一个齐次矩阵，具有8个独立未知元素。每个标定板角点可以提供两个约束方程，即(*u, Xw, Yw*)之间的对应关系提供了两个约束方程。因此，当一张图像中的标定板角点数量达到4个时，就可以计算出该图像对应的矩阵H。

求解内参矩阵，此时已知矩阵*H*= *A*[*R1, R2, T*]，接下来将要求解相机的内参矩阵*A*。此处利用*R1，R2*作为旋转矩阵*R*的两列，很明显它们存在单位正交的关系，其实这个不难理解，因为*R1，R2*是分别绕*x*轴和*y*轴得到的，而*x*轴和*y*轴均垂直*z*轴，即：



且由*H*和*R1，R2*的关系可得：



代入上式可得两个约束条件为：



很容易发现，上述两个约束方程中均存在矩阵*A-TA-1*。因此，记*B= A-TA-1*。接下来将先求解出矩阵*B*，再通过矩阵*B*求解相机的内参矩阵*A*。同样地，为了简便我们将相机内参矩阵*A*记做：



则*A*的逆矩阵为：



则用矩阵*A*表示矩阵*B*得：



由上式很容易发现*B*是一个对称阵，所以*B*的有效元素就剩下6个。同时，可以使用*B= A-TA-1*将前面通过*R1，R2*单位正交得到的约束方程化为：



因此，为了求解矩阵*B*，必须计算。则：



上述方程看起来有些复杂，所以记做：





则上述方程化为：

此时，通过*R*1,*R*2单位正交得到的约束方程可以化为：



其中。

已知矩阵*H*，而且矩阵*v*是由矩阵*H*的元素组成的，因此矩阵*v*是已知的。在这种情况下，我们只需要求解向量*b*，就可以得到矩阵*B*。每张标定板图片提供了一个*vb=0*的约束关系，该约束关系包含两个约束方程。然而，向量*b*包含*6*个未知元素。因此，单张图片提供的两个约束方程不足以求解出向量*b*。为了解决这个问题，我们需要使用3张标定板照片，以得到3个*vb=0*的约束关系，即6个约束方程，从而可以求解向量*b*。



根据矩阵*B*的元素和相机内参的对应关系（如上式），可得到：











即可求得相机的内参矩阵：



求解外参矩阵，对于同一相机而言，相机的内参矩阵是由相机的内部参数决定的，与标定板和相机的位置关系无关。这就是为什么在第二阶段的"求解内参矩阵"中，我们利用不同图片(标定板和相机位置之间存在变化)得到的矩阵H来共同求解相机的内参矩阵A。然而，外参矩阵反映了标定板和相机之间的位置关系。对于不同的图片，标定板和相机的位置关系是不同的，因此每张图片对应的外参矩阵都是不同的。

在等式*A*(*R1, R2, T*) = *H*中，我们已经求解得到矩阵*H*(对于同一张图片是相同的，而对于不同的图片是不同的)和矩阵*A*(对于不同的图片是相同的)。通过公式：(*R1, R2, T*)=*A-1H*，即可求得每一张图片对应的外参矩阵(*R1, R2, T*)。

在这里需要注意，完整的外参矩阵应该为。然而，由于张正友标定板将世界坐标系的原点选在棋盘格上，因此棋盘格上任意点的物理坐标*Z=0*。这意味着旋转矩阵*R*的第三列*R3*可以被消除，因此在坐标转换中*R3*并不起作用。然而，为了使*R*满足旋转矩阵的性质(即列与列之间单位正交)，因此可以通过向量*R1，R2*的叉乘，即*R3=R1×R2*，通过这种方式，我们可以获得*R3*。此时，我们可以得到相机的内参矩阵和外参矩阵。

综上所述，我们将张正友标定法具体归纳为如下步骤：

(1)打印一张棋盘格，把它贴在一个平面上，作为标定物，通过调整标定物或摄像机的方向，为标定物拍摄一些不同方向的照片。

(2)从照片中提取棋盘格角点。

(3)估算理想无畸变的情况下，五个内参和六个外参。

(4)应用最小二乘法估算实际存在径向畸变下的畸变系数。

(5)极大似然法，优化估计，提升估计精度。

思考题与本章习题

1. 已知相机坐标系中一个三维点的坐标为Pc=(1, 0.5, 7)，并知道相机的焦距f=12.5mm，像元为正方形，并且尺寸为dx×dy=4μm×4μm，图像分辨率为1920pixel×1080pixel，图像主点为图像的中心点，请计算Pc在数字图像中的像素位置
2. 已知视觉测量系统中被测物体大小为150×150mm，测量精度要求为0.3mm/pixel，相机距离被测物体约500mm，图像传感器芯片的大小为11mm×8mm，试确定相机的分辨率及镜头的焦距各是多少。
3. 机器视觉系统一般由哪几部分组成？系统应用的核心目标是什么？
4. 简述张正友相机标定法的原理。

