

第 5 章 智能机器人的视觉

机器人的视觉功能在于识别环境、理解人的意图并完成工作任务。机器人的视觉技术包括给定图像的检测与跟踪、多目视觉与距离测量、时序图像检测运动并跟踪、主动视觉等。智能机器人通常利用立体视觉恢复周围环境的三维信息、识别道路、判断障碍物,实现路径规划、自主导航等。

5.1 机器视觉基础理论

机器视觉是光学成像问题的逆问题,它通过获取图像、创建或恢复现实世界模型,实现对现实客观世界的观察、分析、判断与决策。机器视觉技术正广泛地应用于工业检测等各方面,在一些危险场景感知、不可见物体感知等场合,机器视觉更凸显其优越性。

5.1.1 理论体系

1982 年,马尔首次从信息处理的角度综合了图像处理、心理物理学、神经生理学及临床精神病学的研究成果,提出了一个较为完善的视觉系统框架。他认为对视觉系统的研究应分为 3 个层次,即计算理论层、表达与算法层和硬件实现层,如图 5.1 所示。

计算理论层	计算的目的	计算的合理性	执行计算的策略
表达与算法层	信息的编码	数据结构和符号	对应功能的算法
硬件实现层	I/O设备	计算机配置	计算机体系结构

图 5.1 马尔视觉理论的 3 个层次及其所对应的内容

马尔理论是机器视觉研究领域的划时代成就,多年来对图像理解和机器视觉的研究发展起到了重要的作用。视觉系统的 3 个层次如下所示。

- (1) 计算理论层是视觉信息处理的最高层次,是抽象的计算理论层次,它回答系统各部分的计算目的和计算策略。
- (2) 表达与算法层是要进一步回答如何表达视觉系统各部分的输入、输出和内部的信息,以及实现计算理论所规定目标的算法。



(3) 硬件实现层要回答的是“如何用硬件实现各种算法”。

机器视觉研究可以分为如下五大研究内容。

1. 低层视觉

低层视觉的主要研究任务是采用大量的图像处理技术和算法,对输入的原始图像进行处理。

(1) 利用图像滤波、图像增强、边缘检测等技术,抽取图像中诸如角点、边缘、线条、边界及色彩等关于场景的基本特征。

(2) 各种图像变换(如校正)、图像纹理检测、图像运动检测。

2. 中层视觉

中层视觉的主要研究任务是恢复场景的深度、表面法线方向、轮廓等有关场景的 2.5 维信息。

(1) 系统标定、测距成像系统、立体视觉等。

(2) 明暗特征、纹理特征、运动估计等。

3. 高层视觉

高层视觉的主要研究任务是在以物体为中心的坐标系中,在原始输入图像、图像基本特征、2.5 维图的基础上,恢复物体的完整三维图,建立物体的三维描述,识别三维物体并确定物体的位置和方向。另外,主动视觉(active vision)涵盖了上述各个层次的研究内容。

4. 输入设备

输入设备通过光学摄像机或红外、激光、超声、X 射线对周围场景或物体进行探测成像,得到关于场景或物体的二维或三维数字化图像。

5. 体系结构

研究机器视觉从设计到实现中涉及的信息流结构、拓扑结构等一系列相关的问题。

5.1.2 关键问题

机器视觉系统的主要困难体现在以下几方面。

1. 图像多义性

三维场景被投影为二维图像,深度和不可见部分的信息被丢失。不同形状的三维物体投影在图像平面上可能产生相同图像,如图 5.2 所示。不同角度获取的同一物体图像可能存在很大差异。

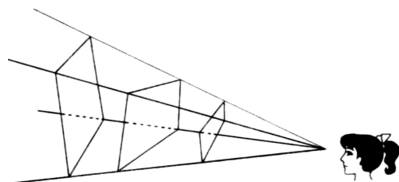


图 5.2 不同形状的三维物体投影在图像平面上产生相同图像

2. 环境因素影响

照明、物体形状、表面颜色、摄像机及空间关系变化都会对获取的图像有影响,几个立方体构成的多义性图像如图 5.3 所示。

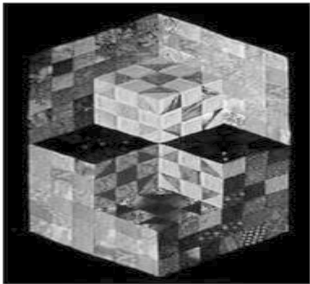


图 5.3 几个立方体构成的多义性图像

3. 知识导引

在不同的知识导引下,同样的图像将会产生不同的识别结果。不同的知识导引也可能产生不同的空间关系。

4. 大数据

灰度图像、彩色图像、高清图像、深度图像、图像序列的信息量会非常大,需要很大的存储空间和计算处理能力。

5.2 成像几何基础

成像系统即是将三维场景变换成二维灰度或彩色图像。这种变换可以用一个从三维空间到二维空间的映射来表示。

$$\begin{aligned} f: \mathbf{R}^3 &\rightarrow \mathbf{R}^2 \\ (x, y, z) &\rightarrow (x', y') \end{aligned} \tag{5.1}$$

如果考虑时变三维场景,则上述变换是四维空间到三维空间的变换。如果再考虑某一波段或某几个波段的光谱,则上式的维数将增加到五维或更高维。

5.2.1 基本术语

简单的三维图像获取过程如图 5.4 所示。



图 5.4 三维图像获取过程

1. 投影

一般地,将 n 维的点变换成小于 n 维的点称为投影,平面几何投影变换的分类如图 5.5 所示。三维场景投影将三维空间的点变换成二维图像中的点。

2. 投影中心

如图 5.6(a)所示,投影线回聚于投影中心(COP)。对于视觉系统,投影中心也称为视点或观察点。

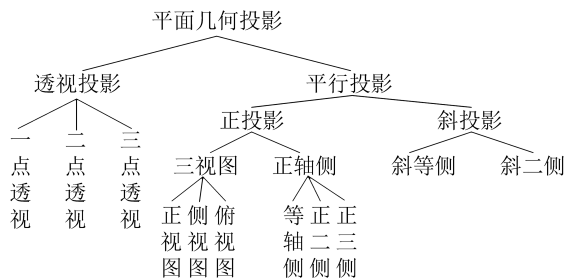


图 5.5 平面几何投影分类

3. 投影线与投影面

从投影中心向物体上各点发出的射线称为投影线,投影面是物体投影所在的假想面。如图 5.6(b)所示,投影线可以是直线或曲线。投影面通常是平面,但有的场合也应用曲面作为投影面。

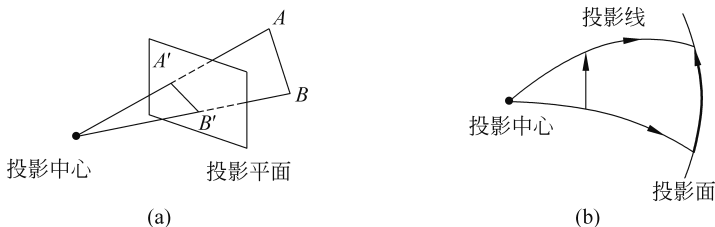


图 5.6 投影过程中的投影线与投影面

4. 投影变换

投影变换是将一种投影点的坐标变换为另一种投影点的坐标的过程。三维空间到二维空间的两种常用投影变换分别是透视投影变换和平行投影变换。

(1) 透视投影变换。如图 5.7(a)所示,透视投影变换的投影中心与投影平面之间的距离为有限远。

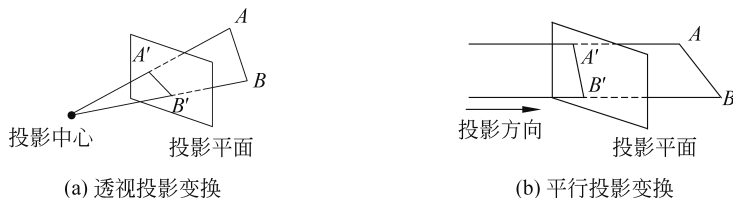


图 5.7 透视投影变换和平行投影变换

(2) 平行投影变换。如图 5.7(b)所示,投影中心与投影平面之间的距离为无限远。可见,平行投影变换是透视投影变换的极限状态。

5.2.2 透视投影

1. 透视现象

由于观察距离及方位引起视觉的不同反应,就是透视现象。利用透视规律,可以正确表现出物体之间的远近层次关系,使观察者获得立体的空间感觉,图 5.8 所示现象正体现了这



一点。



图 5.8 透视现象

在文艺复兴时期,人们发现要把一个事物画在一块画布上,就好比是用自己的眼睛当作投影中心,把实物的影子影射到画布上去。数学家对图形在中心投影下的性质进行研究,逐渐产生了射影几何这门学科。

17 世纪初期,开普勒最早引进了无限远点的概念。“无限远点”在有限的图像上形成的像却是看得见的。与画面成一角度的平行线簇经透视变换后都汇流成一个远方的点,这个点称为灭点或消失点。图 5.9 分别给出了一点透视、两点透视与三点透视的效果图。近大远小的视觉效果,使得图形深度感强,看起来更加真实。

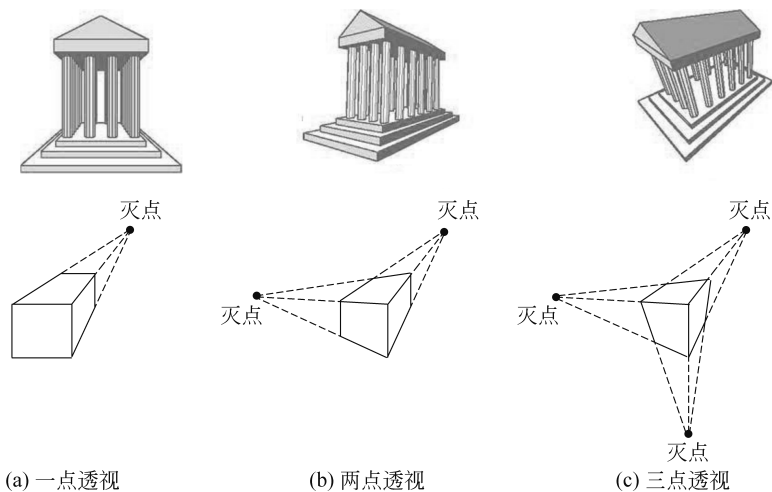


图 5.9 透视效果图



2. 透视投影成像模型

如图 5.10 所示,透视投影可以用针孔成像模型来模拟,其特点是所有来自场景的光线均通过一个投影中心(针孔中心)。透视投影倒立成像的几何示意图如图 5.11 所示,经过投影中心且垂直于图像平面(成像平面)的直线称为投影轴或光轴。

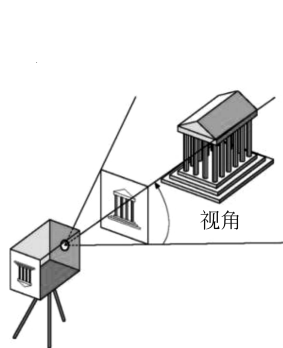


图 5.10 针孔成像模型

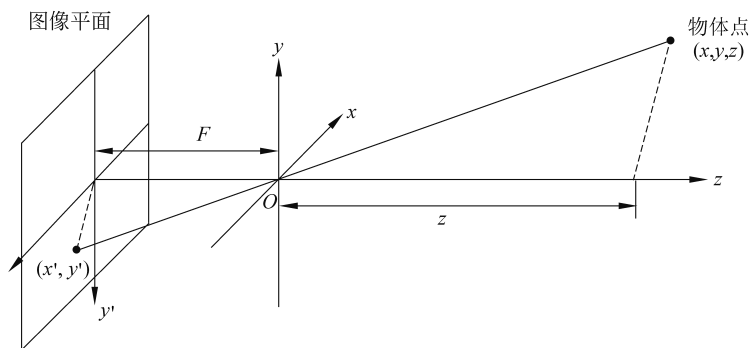


图 5.11 透视投影倒立成像的几何示意图

5.2.3 平行投影

平行投影也称为正交投影,是指用平行于光轴的光将场景投射到图像平面上。如图 5.12 所示,正交投影是透视投影的一个特例,当透视投影模型的焦距 f 很大且物体距投影中心很远时,透视投影就可以用正交投影来近似。

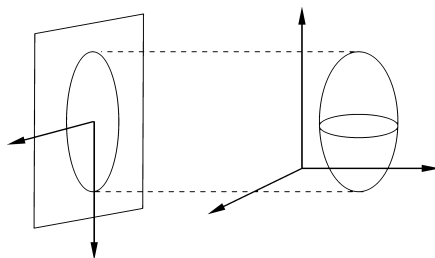


图 5.12 正交投影几何示意图

5.2.4 视觉系统坐标变换

1. 坐标系

在几何学中,为了用数字描述空间物体的大小、形状和位置,必须引进笛卡儿坐标系。用户总是习惯于在自己熟悉的坐标系中描述客体或绘制图形,这个用户定义客体的坐标系称为用户坐标系,或称为客体坐标系。

(1) 用户坐标系有直角坐标系、极坐标系、对数坐标系、球形坐标系等。在图形系统中,一般只用到直角坐标系。直角坐标系又称为宇宙坐标系,可以分为二维直角坐标系和三维直角坐标系。

(2) 设备坐标系一般是二维坐标系,图形的输出在设备坐标系中进行。设备坐标系包括绘图仪坐标系和显示屏幕坐标系。



(3) 规格化坐标系是与设备无关的坐标系,用来构造与设备无关的图形系统。通常取无量纲的单位长度作为在规格化坐标系中图形输入输出的有效空间, X 和 Y 方向的取值范围为 $[0,1]$ 。

用户坐标系、规格化坐标系和设备坐标系三者之间的关系如图 5.13 所示。

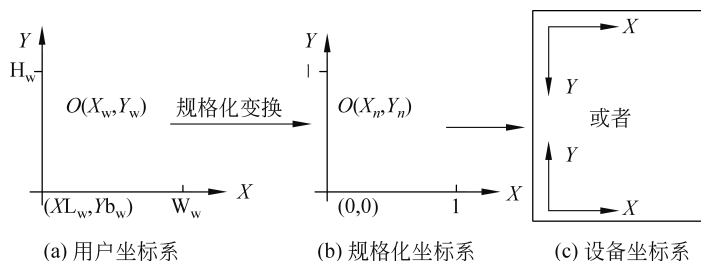


图 5.13 三种坐标系之间的关系

机器视觉系统中通常涉及以下几种坐标系。

- (1) 像素坐标: 表示图像阵列中图像像素的位置。
- (2) 图像平面坐标: 表示场景点在图像平面上的投影。
- (3) 摄像机坐标: 即以观察者为中心的坐标,将场景点表示成以观察者为中心的数据形式。

- (4) 场景坐标: 也称作绝对坐标(或世界坐标),用于表示场景点的绝对坐标。

2. 齐次坐标

齐次坐标表示法就是用 $n+1$ 维向量表示一个 n 维向量。这使得二维的几何变换可以用一种统一的矩阵方式来表示。

在 n 维空间中,点的位置矢量具有 n 个坐标分量 $(P_1, P_2, \dots, P_n)$,它是唯一的。若用齐次坐标表示时,此向量有 $n+1$ 个坐标分量 $(hP_1, hP_2, \dots, hP_n, h)$,它是不唯一的。

考虑对笛卡儿空间内点 P 分别进行旋转、平移、放大、缩小,对应的射影空间内 $P[p] \rightarrow P'[p']$ 的变换操作可用 4×4 的矩阵 T_i 来作为 P 的齐次坐标的线性变换:

$$p' = p \cdot T_i \quad (5.2)$$

式中, $P'[p']$ 表示 P 点变换后对应射影空间内的点。

1) 旋转变换

空间内物体绕 x 、 y 、 z 轴旋转角度 θ ,对应的变换矩阵 T_i 可表示如下:

$$\begin{aligned} T_x &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & T_y &= \begin{bmatrix} \cos\theta & 0 & -\sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ T_z &= \begin{bmatrix} \cos\theta & \sin\theta & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (5.3)$$



2) 平移变换

空间内物体在 x 、 y 、 z 方向平移 (h, k, l) , 对应的变换矩阵 T_t 可表示如下:

$$T_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ h & k & l & 1 \end{bmatrix} \quad (5.4)$$

3) 扩大、缩小变换

空间内物体以原点为中心, 在 x 、 y 、 z 轴方向扩大或者缩小 m_x 、 m_y 、 m_z 倍, 或者全体的 $1/m_w$ 倍, 则对应的变换矩阵 T_m 可表示如下:

$$T_m = \begin{bmatrix} m_x & 0 & 0 & 0 \\ 0 & m_y & 0 & 0 \\ 0 & 0 & m_z & 0 \\ 0 & 0 & 0 & m_w \end{bmatrix} \quad (5.5)$$

表现移动的上述 4×4 矩阵 T_i 的各元素的效果图如图 5.14 所示。

5.2.5 射影变换

在三维空间中, 以某一个视点为中心往二维平面上投影的过程称为透视变换。如图 5.15 所示, 这种将平面 Π 上的图形投影到另一图像平面 μ 上的过程称作“配景映射”。这种“配景映射”可以多次进行, 称为射影变换。显示这个图像时, 显示器有固定的坐标系。用它的坐标直接表现是必要的, 这个变换称为表示变换。

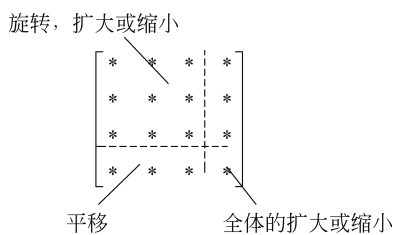


图 5.14 矩阵 T_i 各元素的效果图

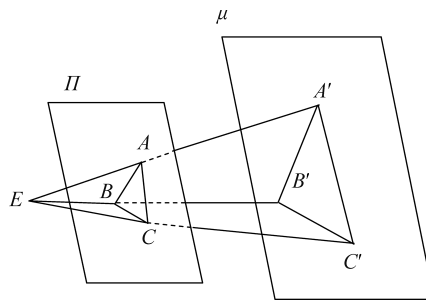


图 5.15 配景映射

三维空间的坐标系规定为现实世界坐标, 称为实坐标或世界坐标。在三维空间中, 三维物体的投影和图像化过程如图 5.16 所示。

- (1) 三维空间内物体的坐标变换, 从作为笛卡儿坐标的世界坐标变换到齐次坐标。
- (2) 必要的话, 可以旋转、扩大、缩小。这由 4×4 且秩为 4 的矩阵 T 来表达。
- (3) 在射影空间内的平面上进行透视变换。这由 4×4 且秩为 3 的矩阵 M_1 来表达。
- (4) 从射影空间内的平面到射影平面的变换。这由 4×3 且秩为 3 的矩阵 M_2 来表达。
- (5) 从射影平面出发到表示射影平面的变换, 用 4×3 且秩为 3 的矩阵 M_3 来表达。

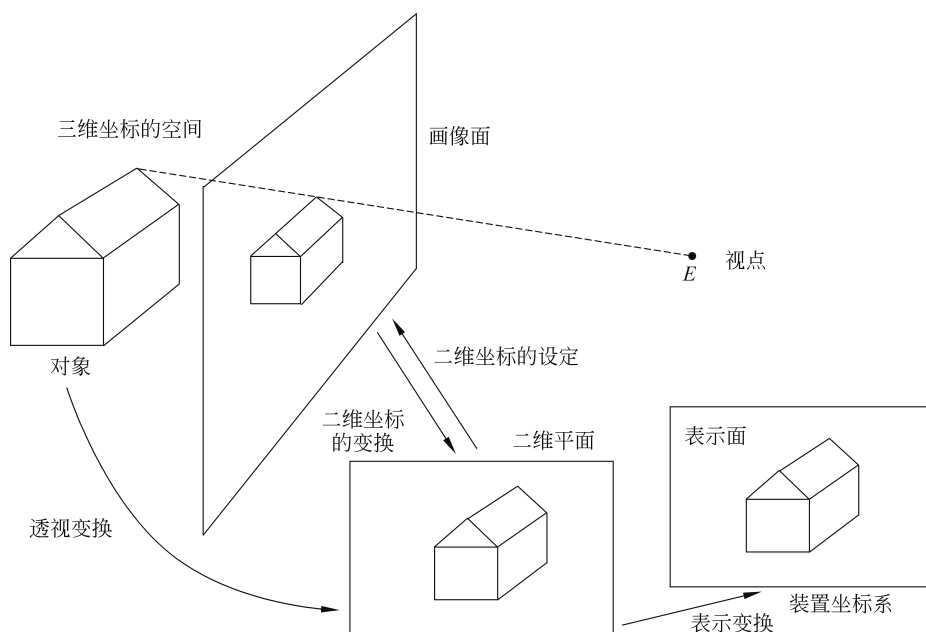


图 5.16 三维空间内物体图像的形成过程

(6) 射影平面上的点齐次坐标进行笛卡儿坐标变换。

步骤(2)到步骤(5)的变换全部为线性变换,可以用矩阵形式表达,即

$$\mathbf{M} = \mathbf{T} \cdot \mathbf{M}_1 \cdot \mathbf{M}_2 \cdot \mathbf{M}_3 \quad (5.6)$$

对于步骤(1)中的对象点 $P(x, y, z)$, 齐次坐标变换为 $(x, y, z) \rightarrow [x, y, z, 1]$ 。射影变换的结果即是右乘矩阵 $\mathbf{M}$:

$$[s, t, u] = [x, y, z, 1] \mathbf{M} \quad (5.7)$$

步骤(6)是为了得到成像平面的坐标进行的笛卡儿坐标变换,可表达为

$$[s, t, u] \rightarrow [s/u, t/u] \quad (5.8)$$

5.3 图像的获取和处理

5.3.1 成像模型

成像系统的建模是建立摄像机成像面坐标与客观三维场景的对应关系。

1. 成像坐标变换

成像变换涉及不同坐标系之间的转换,从三维场景到数字图像的获得所经历成像的变换如图 5.17 所示。

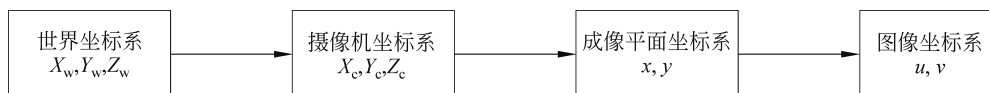


图 5.17 坐标系转换关系图



1) 图像坐标系

摄像机采集的图像是以 $M \times N$ 的二维数组存储的。如图 5.18 所示,在图像上定义的直角坐标系 $u-v$ 中,坐标系原点位于图像的左上角,图像坐标系的坐标 (u, v) 是以像素为单位的坐标。

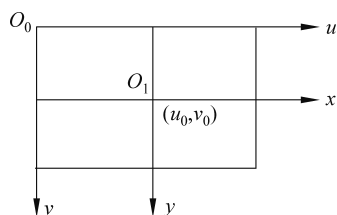


图 5.18 图像坐标系 $u-v$

2) 成像平面坐标系

图像坐标系中的坐标 (u, v) 只表示像素位于数组中的列数与行数,并没有用物理单位表示出该像素在图像中的位置,因此需要再建立以物理单位(如 mm)表示的成像平面坐标系 $x-y$ 。

若原点 O 在 $u-v$ 坐标系中的坐标为 (u_0, v_0) ,每一个像素在 x 轴与 y 轴方向上的物理尺寸为 dx, dy ,则图像中任意一个像素在两个坐标系下的坐标关系如下:

$$u = \frac{x}{dx} + u_0 \quad (5.9)$$

$$v = \frac{y}{dy} + v_0 \quad (5.10)$$

用齐次坐标与矩阵将上式表示为

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (5.11)$$

3) 摄像机坐标系

摄像机坐标系是以摄像机为中心制定的坐标系。摄像机成像的几何关系如图 5.19 所示。

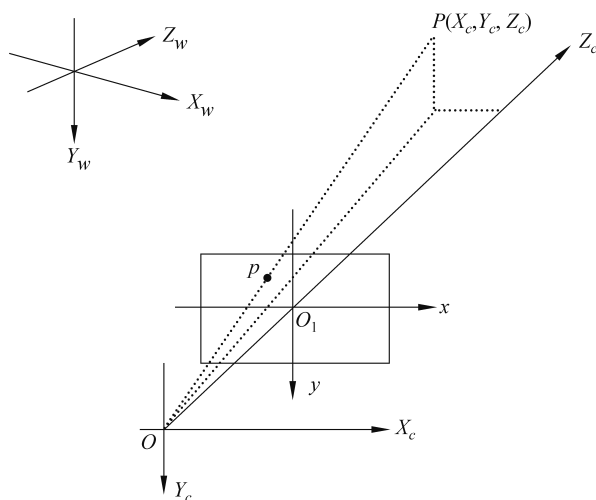


图 5.19 摄像机成像与摄像机为中心制定的坐标系的几何关系