第十二章 立体视觉与三维重建

引言

视觉系统标定与三维重建是计算机视觉领域中的重要研究方向，其在许多领域中具有广泛的应用和重要的意义。这些技术已经开始逐步得到广泛应用，包括在智能制造、智能物流和智能交通等领域。

立体视觉系统是一种模仿人眼视觉系统的技术，使用多个摄像头获取多视角的图像数据，重建物体的三维模型。通过收集多个视角图像，计算机可以处理这些图像并重建整个场景，从而获取场景的深度信息，得到更为精确的3D模型。三维重建的任务是通过从多个二维图像中推导出场景的三维结构，以获取物体的准确位置、形状和运动信息。还可以非常有效地解决物体测量、建筑构造、文物保护和电影特效等任务，同时也是增强现实技术的重要组成部分。

点云处理技术是从点云数据中提取有用信息的关键步骤，可以通过优化点云配准、去噪、分割和形状重建等方法，为后续的数据处理提供更为精确的数据基础。点云技术的核心思想是通过以云点为单位来描述物体的空间几何形状和拓扑关系，从而在实现三维数据分析和可视化方面发挥了重要作用。

立体视觉系统与三维重建在计算机视觉领域扮演着重要的角色。研究和应用视觉系统与三维重建技术有助于推动计算机视觉技术的发展，并为各个应用领域提供准确、高效的视觉测量和重建手段。

12.1 立体视觉系统

立体视觉系统是通过两个(或多个)摄像头在不同位置分别捕捉同一场景的图像，它通过分析两个图像中物体的位置差异和几何关系，从而实现对物体的三维感知。立体视觉系统的实现通常需要使用算法和技术来进行图像匹配、深度恢复和三维重建等操作。总而言之，立体视觉系统可以通过模拟人眼的立体视觉原理，帮助机器实现对物体的三维感知和理解。

12.1.1 平视双目视觉系统

双目视觉系统是立体视觉系统的一个特例，它使用相似的技术和方法来获取深度信息，但只需要两个摄像头，相对立体视觉系统来说更为简单和便宜。双目视觉系统已广泛应用于自动驾驶、机器人、安防、医疗等领域中，并且随着技术的进步，其性能和应用前景不断提高和拓展。

双目视觉系统类似于动物的眼睛，不仅可以获取场景的平面图像信息，还能计算被测目标的距离和相对深度信息。如图11-12所示，双目视觉系统由左右两个相机组成，它们分别位于世界坐标系中被测目标表面上的任意点*Pw*(*Xw, Yw, Zw*)的左相机坐标系中的投影点*p*(*uL, vL*)和右相机坐标系中的投影点*p*(*uR, vR*)。当使用单目视觉系统观察时，我们只能知道点*Pw*(*Xw, Yw, Zw*)可能位于左相机光心*OL*与点*p*(*uL, vL*)构成的射线上，或者位于右相机光心*OR*与点*p*(*uR, vR*)构成的射线上。但如果同时使用双目视觉系统观察，并且如果

我们能够确定像点(*uL, vL*)和(*uR, vR*)对应于同一空间特征点*Pw*(*Xw, Yw, Zw*)，那么就可以通过计算两个像点与相机光心所构成的两条直线的交点来确定点*Pw*(*Xw, Yw, Zw*)的精确位置。需要注意的是，由于相机倒立成像的原理，图12-1中将正立的成像平面放置在相机前方的焦距处，以方便讨论。

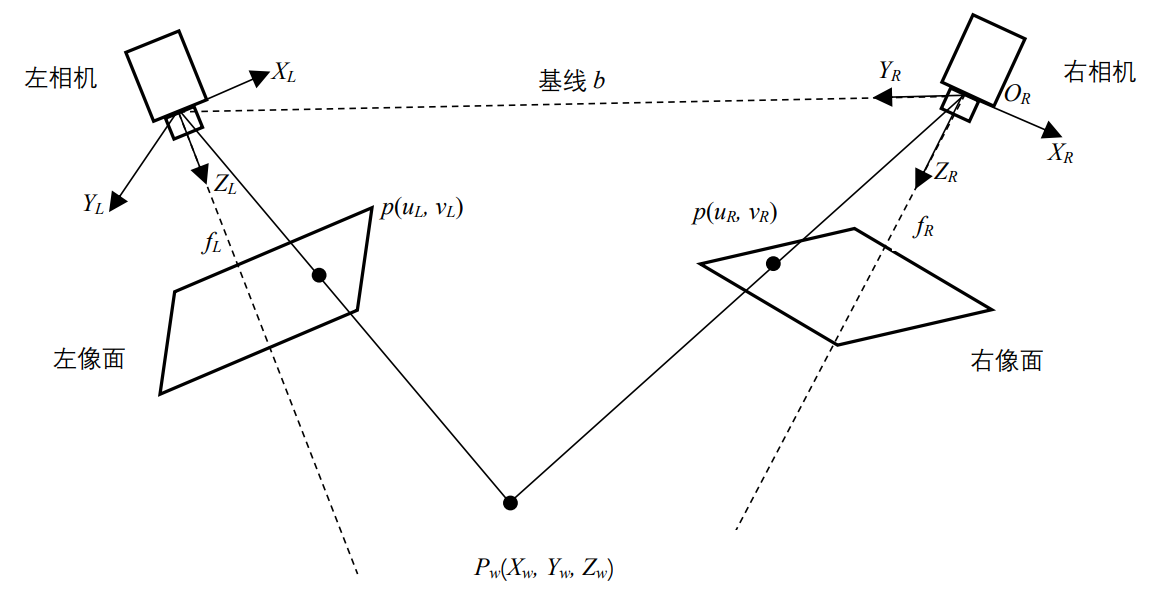


图12-1 双目视觉示意图

双目视觉系统可以利用两个相机采集的图像中的两个同源点(Homologous Points)之间的距离来重建空间中的三维信息，这是通过三角几何法实现的。同源点指的是空间中的某个点在不同图像中的成像点，也称为共轭点(Conjugate Points)。它们在图像中的距离差异被称为视差(Disparity)。

研究双目系统时一般先要明确研究过程所涉及的以下几种类型的坐标系：

(1)世界坐标系。预先在环境中选择的基准三维坐标系，用来描述相机和环境中物体的坐标位置。在双目视觉系统中，一般将左相机坐标系定义为世界坐标系。

(2)相机坐标系。以相机光心为原点，光轴为*Z*轴的三维坐标系。在双目系统中，左右两个相机坐标系可通过旋转和平移进行转换。

(3)图像坐标系。相机所成图像的二维坐标系，可以用像素数量描述，也可以用物理尺寸描述。对于相机坐标系中的某个三维空间点，可通过三角投影变换计算得到其在图像坐标系中像点的位置。

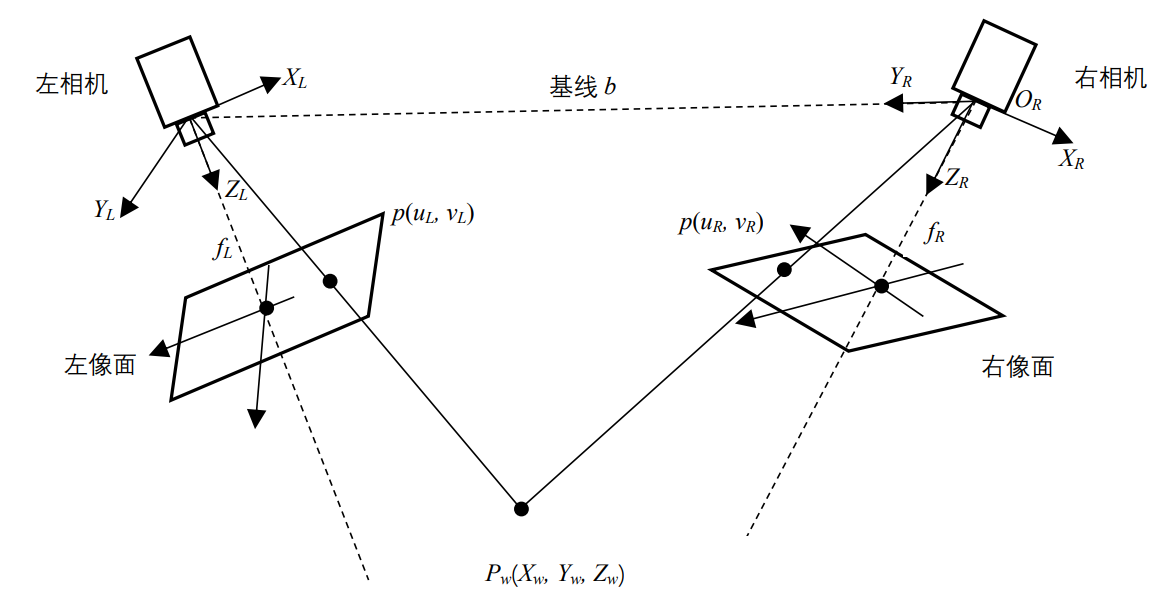


图12-2 双目系统中的坐标系

根据前述讨论，若定义左相机坐标系就是世界坐标系，那么世界坐标系中的点*Pw*(*Xw, Yw, Zw*)到左相机的图像坐标系中的位置可直接通过三角投影关系计算得到：



下标*L*表示与左相机相关的值。该点在右相机图像坐标系中的位置同样可以通过投影变换计算得到。然而，由于投影变换必须在相应的相机坐标系中进行，因此必须先将空间点的世界坐标*Pw*(*Xw, Yw, Zw*)通过旋转平移变换为右相机坐标系中的坐标*PR*(*XR, YR, ZR*)，然后才能正确计算。表示如下：



因此，如果我们已经知道左右相机的坐标系、焦距以及空间中某一点在两个相机图像坐标中的位置，就可以通过两个相机的投影变换关系得到四个方程。根据这四个方程，我们可以求解出空间点的世界坐标*Pw*(*Xw, Yw, Zw*)。在实际应用中，常常使用简化问题的平视双目系统模型。图12-3展示了平视双目系统的结构和特点。通常情况下，系统中的两个相机类型相同(至少焦距相同)并且平行安装。左相机坐标系被定义为世界坐标系，而左右相机坐标系的x轴是共线的，并且与基线平行。平视双目系统也可以通过移动单个相机来实现，但在这种情况下，相机的运动轨迹必须是直线。

平视双目视觉系统特点：

(1)两个相机类型相同，或至少焦距相同

(2)两相机平行安装。若为单个相机运动系统，则运动轨道为直线。

(3)两相机*X*轴共线，且与基线平行。

(4)定义左相机坐标系为世界坐标系

假定图12-3中平行安装的两个相机中心连线的距离为*B*(称为基线)，空间点*Pw*(*Xw, Yw, Zw*)在左右两个相机所成图像的坐标系中的位置分别为*p*(*uL, vL*)和*p*(*uR, vR*)。

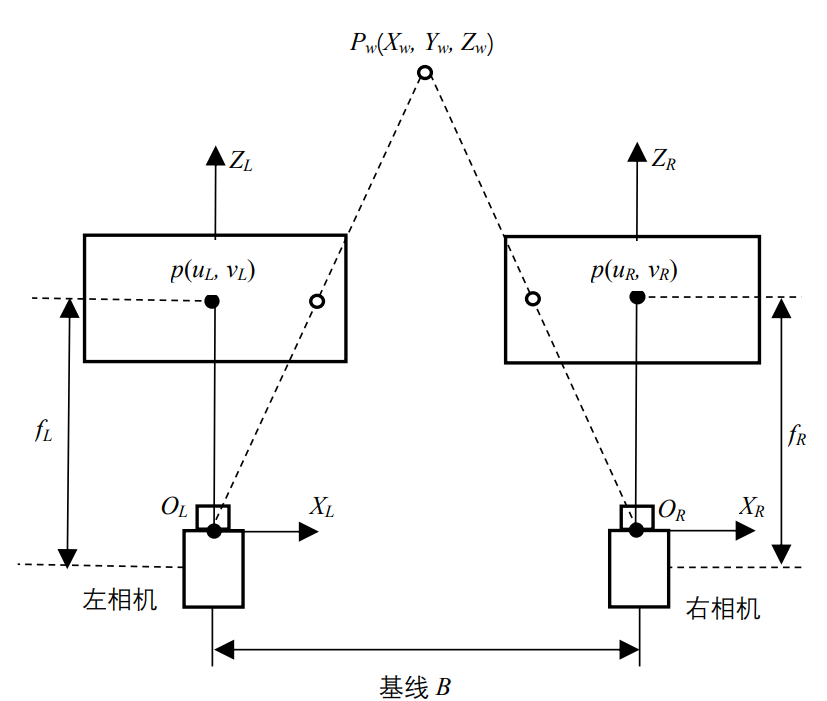


图12-3 平视双目系统及其特点

由于两个相机类型相同，因此所成图像在同一个平面上，从而空间点在两个图像坐标系中*y*方向的坐标相同，即*vL=vR*。进一步根据几何关系可得到：



*f*表示相机焦距(左右相机焦距相同)。需要注意的是，在计算像点在右相机图像上的位置时，需要先通过坐标平移将空间点*Pw*(*Xw, Yw, Zw*)的变换到右相机坐标系中，然后再进行计算。如果我们定义视差*d*为*p*(*uL, vL*)和*p*(*uR, vR*)两点在图像中的距离差异，则有以下关系式：





因此，只要左相机像面上的任意一点能够在右相机像面上找到对应的匹配点，就可以根据视差确定该点的三维坐标。更进一步地，只要被测目标像面上的任意点可以在另一相机中找到对应的匹配点，就可以确定被测目标的三维坐标。

12.1.2一般立体视觉系统

一般立体视觉主要分为三种：基于视差的立体匹配模型、基于结构光的立体视觉模型和基于多视角的立体视觉模型。

基于视差的立体匹配模型：这种模型通过计算两个视角的图像之间的像素差异，来获取场景中每个像素的深度信息。例如，当我们将一个物体从左眼的视野中移动到右眼的视野中，我们会观察到物体在两个眼睛之间的视差变化。通过计算这个视差，我们可以估计物体的深度信息。基于视差的立体匹配模型通过计算每个像素的视差来实现这一目标。 常用的匹配算法包括基于块匹配的算法(如均值差异和绝对差异)和基于图优化的算法。

基于结构光的立体视觉模型：这种模型使用结构光投射器和相机来获取场景的深度信息。结构光投射器会通过投射编码的光图案来改变场景的纹理，相机则用于捕捉投射后的图像。通过分析图像中的纹理变化，可以反推场景中物体的深度。 基于结构光的立体视觉模型通常需要使用高质量、高精度的结构光投射器和相机来实现比较准确的深度测量。

基于多视角的立体视觉模型：这种模型利用多个相机同时捕捉不同角度的图像，然后通过三角测量等技术来恢复场景的三维结构。多视角的信息可以提供更多的深度信息，从而提高立体视觉的准确性和稳定性。多视角的立体视觉模型通常需要对系统进行校准，以确保多个相机之间的几何关系和图像对齐。在实际应用中，基于多视角的立体视觉模型可以通过多个相机组成的阵列来实现，例如在3D扫描和人体姿态识别等应用中有广泛应用。

12.1.3 立体视觉系统标定

通过标定过程，立体视觉系统可以得到准确的摄像头参数和图像校正参数，从而提高系统的匹配精度和深度估计准确性。立体视觉系统标定是一个复杂且精细的过程，需要使用专门的标定工具和方法，并且要确保标定过程中的环境和条件与实际使用场景尽可能地相近，以获得更好的结果。

为了使用立体视觉系统进行视差计算和获取三维深度信息，首先需要获取以下信息：

相机的内部参数：包括每个相机的焦距、光心以及畸变模型等。

相机的外部参数：包括每个相机坐标系与世界坐标系之间的旋转和平移关系。

相机间的相对位置关系：包括不同相机坐标系之间的旋转和平移关系，以及多个图像中同一空间点对应的同名像素点的坐标。

在实际工程中，获取相机内外参数和相机间的相对位置关系被称为对立体视觉系统的标定过程。为了完成立体视觉系统的标定，通常会让系统的不同相机对同一个标定点阵进行图像采集，然后根据点阵的信息来计算各种参数。图12-4展示了双目视觉系统的标定过程。

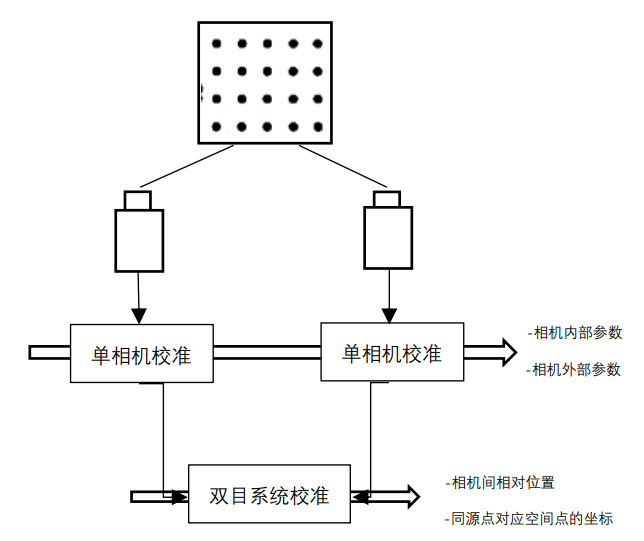


图12-4 双目视觉系统的标定过程

双目立体视觉系统的标定过程可以分为相机标定和立体视觉系统标定两个阶段。在相机标定阶段，我们会计算出系统中每个相机的焦距、光心位置、相机的畸变模型，并确定每个相机坐标系与世界坐标系之间的关系。

如前述单目视觉标定，表示三维空间点与其图像点变换关系的投影矩阵*M*可通过6个以上的标定点计算得到。而基于它和相机内外参数矩阵的关系，能计算得到相机的内外参数。此后，就可以继续对整个立体视觉系统进行标定，来确定系统中各个相机的坐标系与其他相机坐标系之间的旋转和平移关系，以及多幅图像中同源像点所对应空间点的坐标等。

设双目立体视觉系统中左右相机坐标系相对世界坐标系的旋转和平移矩阵分为(*RL, BL*)和(*RR, BR*)，则三维世界坐标系中的点*P*(*Xw, Yw, Zw*)在两个相机坐标系中的坐标*PL*(*XL, YL, ZL*)和*PR*(*XR, YR, ZR*)可通过坐标旋转和平移得到。用矩阵表示为：



消去P可得到以下PL和PR之间的关系式：







由上式可知，若对双目立体视觉系统中的左右相机分别完成标定得到(*RL, BL*)和(*RR, BR*)，则左相机坐标系到右相机坐标系的转换关系就能通过*R*=(*RLRR-1*)和*B*=(*BL-RLRR-1BR*)来确定。然而，在实际应用中，通常并不会通过空间点和各成像点之间的关系来间接计算立体视觉系统中相机之间的相对位置关系。相反，我们会基于极线几何理论直接获取两个相机图像中对应点之间的极线约束，然后才完成后续的立体视觉图像调整、图像对应点匹配和目标的深度信息计算等过程。

12.2 三维重建

12.2.1 极线几何理论

极线几何理论可用于研究立体视觉系统中各个相机图像同源像点之间的关系，它不仅在双目立体视觉图像的对应点匹配中有着重要作用，而且在三维重建和运动分析中也具有广泛应用。

如图12-5所示，这是极线几何的典型设置。在这个设置中，灰色区域表示对极平面，橙色线表示基线，而蓝色线表示对极线。在多视图几何中，多个相机、三维点以及它们在每个相机图像平面上的投影之间存在着紧密的关联。这种几何结构，将相机、三维点和相应的观测值联系在一起，被称为对极几何结构。标准对极几何设置两个相机，观察同一3D点*P*，其在每个图像平面中的投影分别位于*p*和*p’*。相机中心位于*O1*和*O2*，它们之间的线称为基线。由两个相机中心和*P*定义的平面称为对极平面。基线与两个图像平面的交点称为对极*e*和*e’*。最后，由对极平面和两个图像平面的交点定义的线称为对极线。

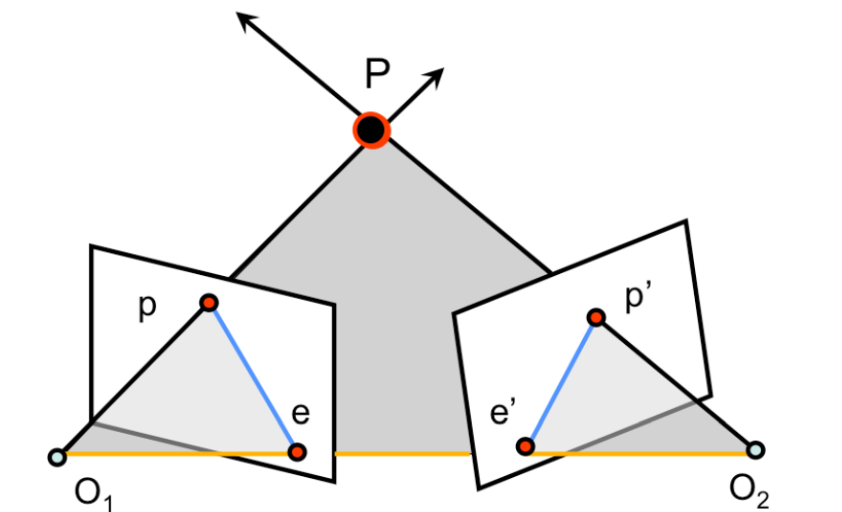


图12-5 极线几何设置

一个特殊的对极几何的例子如图12-6所示，这时图像平面彼此平行。当图像平面彼此平行时，由于连接中心*O1*、*O2*的基线平行于图像平面，因此外极*e*和*e’*，将位于无穷远处。请注意，此时对极线平行于每个图像平面的*u*轴。

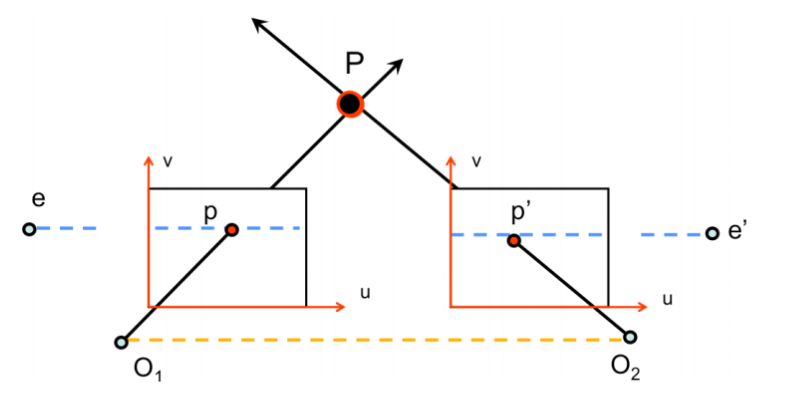


图12-6 两图像平面平行时极线设置

然而，在现实世界中，一般没有3D点*P*的确切位置，但可以确定其在图像平面上的投影点*p*。还能够确定相机的位置、方向和相机矩阵。已知这些条件，根据相机位置*O1*、*O2*和图像点*p*，同样可以定义对极平面。利用此对极平面，可以确定对极线。根据定义，点*P*在第二个图像平面上的投影*p*,必须位于第二张图像平面的对极线上。因此，通过对对极几何的基本了解，可以让我们在不了解场景的3D点的情况下在图像对之间创建强约束。所以，在不了解3D场景结构的情况下，可通过对极几何的基本知识对图像创建强约束。

同时定义*M*和*M′*为将3D点映射到各自的2D图像平面位置的相机投影矩阵。假设世界坐标系与第一相机相关联，再旋转*R*，然后平移*T*，得到第二相机，如图所示。那么两个相机的投影矩阵分别为：*M*=*K*[*I, 0*]和*M′*=*K′*[*R, T*]。

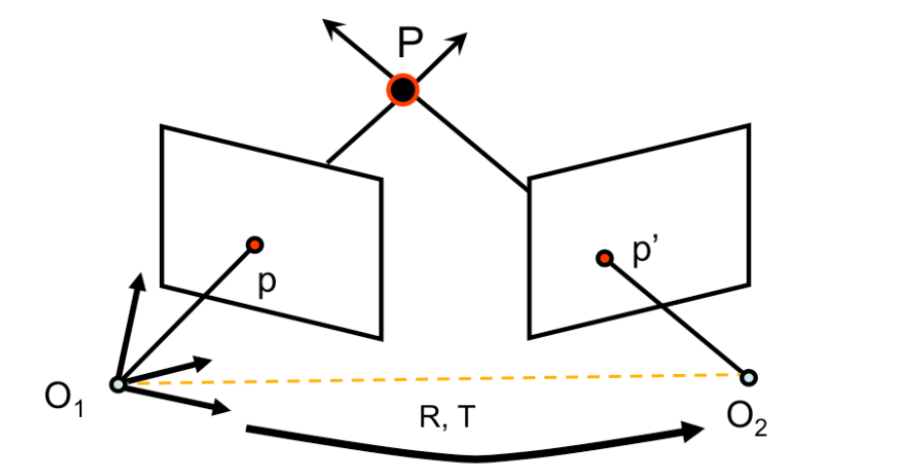
-

图12-7 相机的旋转和平移

在最简单的情况下，假设此处是标准相机，即*K*＝*K′*＝*I*。则*M*和*M′*可以化简为：



这表示，点*p’*在第一相机参考系中的位置是*RTp’-RTT*。因为向量*RTp’-RTT*和*RTT*位于对极平面，所以可以得到：*RTT*×(*RTp’-RTT*)= *RTT*×*RTp’*=*RT*(*T×p’*)，这样就可以获得一个垂直于对极平面的向量。同时也证明，位于对极平面的*p*垂直于*RT*(*T×p’*)，综上所述列出点积为零的约束条件：





利用线性代数的知识，可以为叉积引入一个不同的矩阵表达式：将任意两个向量*a*和*b*之间的叉积表示为矩阵向量乘法：



将上述矩阵和约束条件结合，就可以将叉积项转换为矩阵向量乘法，得到如下等式：







将矩阵*E*=[*T×*]*R*称作本质矩阵，通过此矩阵为极线约束创建了更为精炼的表达式：



上述本质矩阵*E*是包含5个自由度的3×3矩阵，此矩阵的秩为2。本质矩阵有助于计算与*p*和*p′*相关的对极线。例如，*l’=Ep*给出了相机2的图像平面中的对极线。同样，*l’=ETp*给出了相机1的图像平面上的对极线。本质矩阵一个特殊的性质是，它与对极线的点积等于零，即*Ee=ETe’=0*。因为对于相机1图像平面中的任何点*x*(*e*除外)，相机2图像平面中的对应对极线*l’=Ex*包含对极线*e’*。因此，对于所有点*x*，*e’*满足等式*e’T(Ex)=(e’TE)x=0*，因此*e’TE=0*，同样*Ee=0*。虽然当有正则化相机时，我们导出了*p*和*p′*之间的关系，但当相机不再是正则相机时，则应该推导出更一般的表达式。根据上述内容，已经得到了投影矩阵：

M=K[I, 0]M’=K’[R, T]

首先，此处定义*pc=K-1p*和*p’c=K’-1p’*是点*P*到相应相机图像的投影，在典型标准的相机下，满足：



将*pc*和*p’c*的值代入上式，那么可以得到：



将矩阵*F*=*K’−T* [*T×*]*RK−1*称为基础矩阵，其作用类似上述描述的本质矩阵，其编码相机矩阵*K、K’*以及相机之间的相对平移*T*和旋转*R*的信息。因此，即使相机矩阵*K、K’*和变换*R、T*未知，*F*也有助于计算与*p*和*p’*相关的对极线。与本质矩阵类似，可以只用基础矩阵和对应点来计算对极线*l’=Fp*和*l=FTp’*。基础矩阵和本质矩阵之间的一个主要区别是，与本质矩阵的5个自由度相比，基础矩阵包含7个自由度。

但是基础矩阵有什么用处呢？就像本质矩阵一样，如果我们知道了基础矩阵，那么只需要简单地知道图像平面上的一个点，就可以轻松地约束另一个图像平面上的对应点(即对极线)。因此，即使在不知道P在3D空间中的实际位置，或相机的任何外在或内在特征的情况下，就可以建立任何P和P’之间的关系。

从上述几何和数学分析中可以得出，极线约束是描述像点与其极线之间的关系。通过这种约束，我们可以确定一个相机图像中的某个像点在另一个相机图像中对应位置所在的直线，但无法知道同源像点在该直线上的具体位置。尽管，极线约束将图像中对应点的匹配范围限定在一条直线上。即使考虑噪声，如果我们将搜索范围扩大到该直线附近的邻域内，极大的减少图像对应点匹配过程的计算量。因此，极线约束对图像对应点的匹配过程具有指导作用。

12.2.2 基于立体视觉的三维重建

通过进行单相机和立体视觉系统的标定过程，我们可以确定同源像点在不同相机坐标系或图像坐标系中的坐标关系。然而，要使用视差来计算空间点的深度信息，我们还需要对图像进行进一步的调整(Image Rectification)，使两个相机的图像共面且“行对齐”，以简化计算过程。

如图12-8所示，当立体视觉系统可被配置或抽象为平视双目系统模型时，空间点的深度信息计算就会被简化。平视双目系统模型具有以下主要特点：

(1)左右相机类型相同，或至少焦距相同。

(2)左右相机且平行安装。即光轴平行。

(3)左右相机坐标系的*x*轴共线，并与基线平行。

这意味着两相机的图像共面，光轴与图像的交点(投影中心)在左右图像中的坐标相同。若不考虑畸变，则两图像中各行像素不仅水平方向相同，且具有相同的空间*y*坐标，即图像“行对齐”。

平视双目系统模型对于深度信息的计算提供了很大的简化，但它却是一个非常理想化的模型。在现实世界中，几乎无法使两个相机的像平面完全共面，因此物理上实现完全平视双目系统是不可行的，我们只能尽可能地根据其特点进行近似配置。然而，通过对图像进行畸变校正、旋转平移和行对齐等调整过程，我们可以将这种“近似的平视双目系统”转换为近似理想的平视双目系统。以图12-8为例，通过应用旋转矩阵*R*，我们可以使两个图像共面。然后，通过平移和行对齐操作，我们可以使两个相机的图像具备平视双目视觉系统的特点。

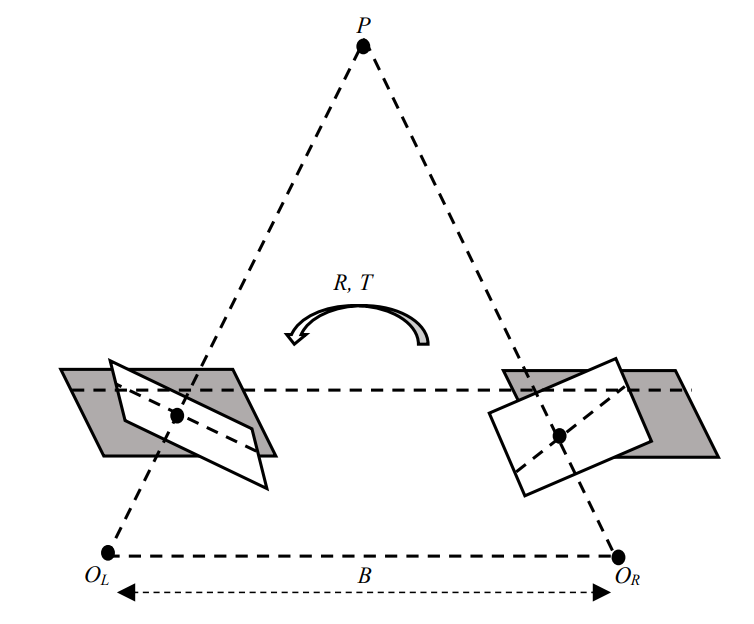
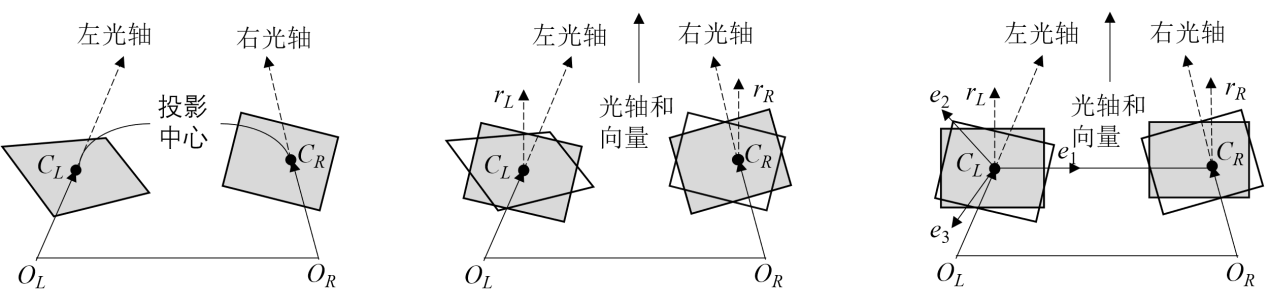


图12-8 图像调整目标

假设左右相机图像之间的位置关系可以用旋转矩阵*R*和平移向量*T*表示。我们可以首先通过旋转其中一个图像坐标系，使左右图像共面。为了减小仅旋转其中一个图像坐标系可能对整个系统的影响，可以分别对左右图像坐标进行旋转*rL*和*rR*，使它们均与两相机光心和投影中心所构成向量的和(亦为向量)平行。虽然分别旋转左右相机坐标可使两图像共面，但两共面图像中的像素行却并不一定对齐。这对实现后续的图像对准过程不利，因此还需要对两图像进行调整。图像的行对齐可通过矩阵*Rrectify*来实现：



其中*T*为两图像投影中心之间的平移向量，*e1、e2*和*e3*为3个正交向量，用于将共面的两个图像调整为行对齐。是当左图像中以投影中心为图像坐标原点时左极点的单位方向矢量。若使该单位矢量与两图像投影中心所构成向量的方向相同，就能使左极点在右图像中对应的极线为水平方向(极点位于无穷远处)。，它是e1和与光轴方向的单位向量的叉积，也就是说*e2*和*e1*与光轴均正交(垂直于*e1*与光轴构成的平面)。*e3*=*e1*×*e2*，即*e3*是*e1*和*e2*的叉积，它与*e1*和*e2*均正交。获得矩阵*Rrectify*后，就可以利用它继续对左右图像旋转后所得图像进行变换，使它们对齐。图12-9显示了图像调整过程中左右图像的旋转和行对齐步骤。



(a)原图像 (b)左右图共面 (c)左右图对齐

图12-9 图像调整过程的主要步骤

若将左右图像的旋转和行对齐过程综合，就可得到下面两个分别施加到左右图像的变换矩阵：



对左右相机图像分别应用RL和RR，可以获得对齐的图像。然而，在实际应用中，还需要确保左右相机图像的同步采集。如果左右相机不能同步采集图像，那么对于场景中的移动目标，将会导致错误的深度信息。这种情况下，系统的应用将受限于只能观测静止目标。

图12-10的示意图显示了图像调整过程。当被测目标的原始图像被左右相机采集后，图像调整过程会先对各种畸变进行矫正，然后调整左右图像使它们共面并且行对齐。在实际中，对于每个调整后的图像(c)中的整数坐标，图像调整过程都会从矫正图像(b)中找到对应坐标，然后映射到原始图像(a)中的浮点数坐标。接着再用其周围的整数坐标像素值通过插值算法(0阶插值或双线性插值)得到浮点坐标处的像素值。而该像素值就作为调整后的图像(c)中整数坐标处的像素灰度。最后通常需要对调整后的左右图像进行剪裁(d)，以提取用于计算深度信息的交集区域。

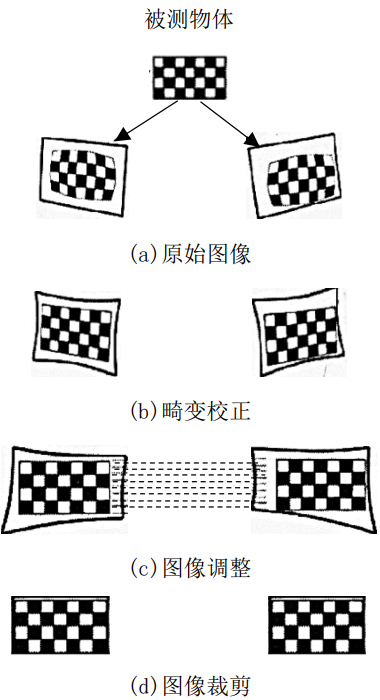


图12-10 图像调整过程

图12-11展示了图像调整过程前后左右图像的实例。因此，可以将实际中的图像调整过程看作是一个从原始图像到调整后图像的插值过程。为了提高速度，可以为左右图像的插值过程分别创建查找表，从而加快计算速度。这也意味着需要占用更多的内存空间。

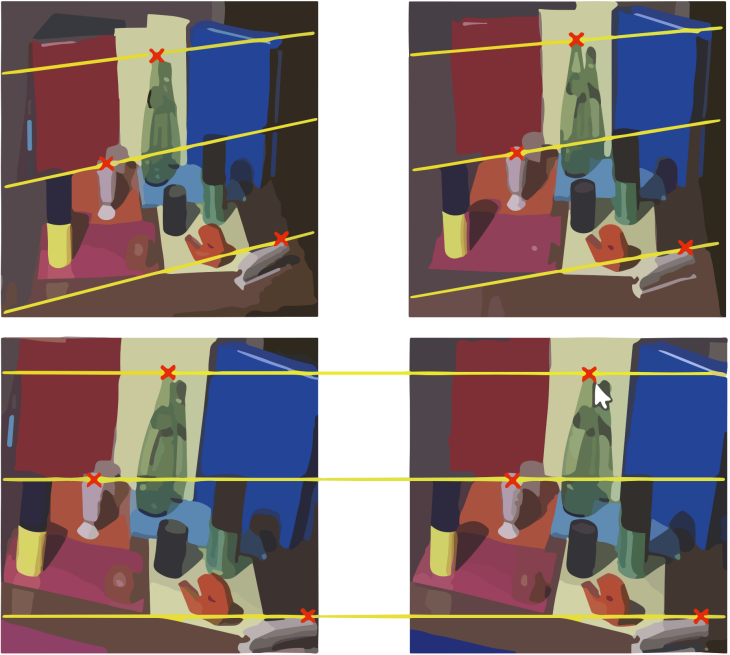
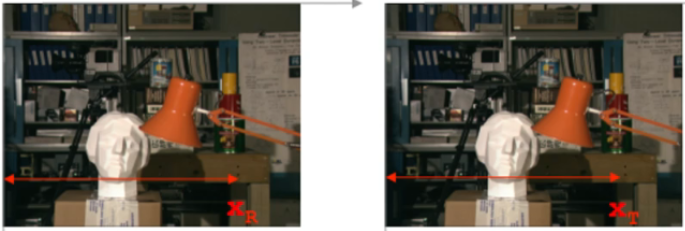


图12-11 图像调整实例

经过图像调整后，立体视觉系统的左右图像不仅共面，而且在每一行上都能够对齐。但是，在计算视差时，两幅图像中用于确定特征点位置的仍然无法直接对应。因此还要基于调整后的左右图像进行图像对应点匹配才能得到视差，并最终实现被测目标的深度信息计算或进行三维重建。

对应点匹配过程通常输出左右图像的视差图。视差图以立体视觉系统左右图像中任一幅图像为参考(通常取左图)，其大小与参考图像大小相同，但各元素值为左右图像同源点视差值。例如，在图12-12中，(a)和(b)为立体视觉系统的左右图像，*XR*为左视图横坐标，*XT*为右视图横坐标。图12-13是以左侧图像为参考的二维视差图，其中每个像素的灰度代表了左图中像素与其在右图中同源点间的图像坐标之差(即视差*d*)，即d=*XR-XT*。若以三维显示来观察该视差图，则有图12-14。考虑到两幅图像的同源点之间的视差值通常较小，通常会通过乘以一个因子(如16)来放大视差值，以便更好地显示视差图。在平面双目视觉系统中，空间点到相机的距离与视差呈反比关系。因此，视差图中像素的亮度(灰度值)越高，表示对应的空间点离相机越近；相反，较暗的像素表示空间点离相机较远。



(a)左图像(参考图像) (b)右图像

图12-12 同一场景的左右视图

****

图12-13 视差图2D显示

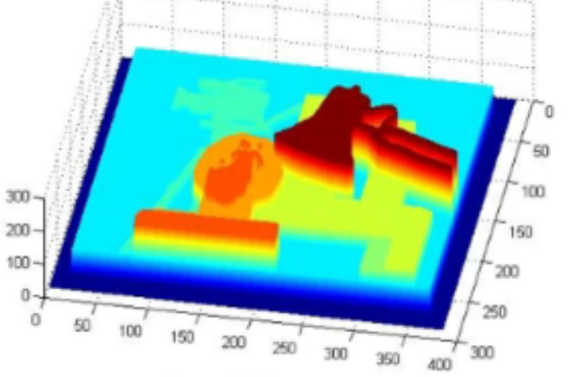
****

图12-14 视差图3D显示

目前存在两种常见的对应点匹配算法：块匹配算法和半全局匹配算法。块匹配算法具有较快的执行速度，而半全局匹配算法则可以提供更细腻的视差图，并适应纹理较少或几乎无纹理的图像。

块匹配算法可以分为前置滤波、匹配和后置滤波三个阶段。前置滤波的主要目的是减小图像中的亮度差异，增强图像的高频细节。常用的方法是应用Sobel滤波器或归一化响应滤波器。为了避免水平线误导匹配过程，通常只使用以下Sobel滤波器内核：



而使用该滤波器内核对图像增强时，采用以下方法取值：



其中*I*为图像增强后的像素值，也是应用Sobel滤波器内核时的中心像素。*ISobel*为对图像应用Sobel滤波器内核后得到的值(窗口内像素的加权和)，*Icap*是用来限定像素最大值的正参数。归一化响应滤波器通过在图像中滑动尺寸为5×5、7×7或其他奇数大小的窗口(最大为21×21)，并通过以下方法来计算窗口中心位置的新像素值：



其中*I*为图像增强后的像素值，也是窗口中心像素的新值。*ICenter*为原窗口中心像素的值，*Iavg*是窗口中像素值的平均值，*Icap*是用来限定像素最大值的正参数。

经过前置滤波以增强图像细节并减少亮度差异后，就可以进行左右图像的对应点匹配。由于在对左右图像进行调整后，左右图像中的同一行像素互为极线，因此对于左图像中的任意一个特征点，只需要在右图像的相同行(即*y*坐标相同)进行搜索，即可找到对应的匹配点。为降低噪声和其他因素的影响，在搜索匹配点时可在右图像中沿极线滑动一个小窗口，并计算左图像中同样大小的窗口内各特征像素点与右侧滑动窗口内各像素之间的绝对值差之和(Sum of Absolute Differences,SAD),最后从中选出SAD最小的窗口作为匹配窗口。该过程可通过下述步骤描述：

(1)构造一个小窗口(类似卷积核)。

(2)用窗口覆盖左边的图像，选择出窗口覆盖区域内的所有像素，包括特征像素。

(3)同样用窗口覆盖右边的图像并选择出覆盖区域的像素点。

(4)左边覆盖区域减去右边覆盖区域，并求出所有像素点差的绝对值之和(SAD)。

(5)移动右边图像的窗口，重复步骤(3)和(4)。

(6)找到这个范围内SAD值最小的窗口，即找到了左边图像的最佳匹配的像素块。

可以进一步基于极线约束对搜索范围进行限定。假定左图中特征像素的坐标为(*x0, y0*),则对于平视双目视觉系统来说，该特征点在右相机图像中只能出现在(*x0, y0*)坐标的左侧(同一行)。只有当立体视觉系统的两相机之间有一定夹角时(非平视双目视觉系统)，匹配点才可能出现在坐标(*x0, y0*)的右侧。若将(*x0, y0*)坐标处的匹配点视差为0，则其左侧匹配点的视差为正，右侧匹配点的视差为负。由此，可以通过最小视差和视差像素数两个参数来控制搜索范围。其中最小视差规定了在右图像中极线上搜索匹配点的起始位置(默认值为0)，而视差像素数则规定了搜索的最大区域(默认值为64个像素)。这样就可以将搜索范围限定在极线上x0-MinDisparity到x0-MinDisparity-NumDisparity范围内。图12-15显示了平视双目视觉系统中对应点匹配过程的搜索范围。

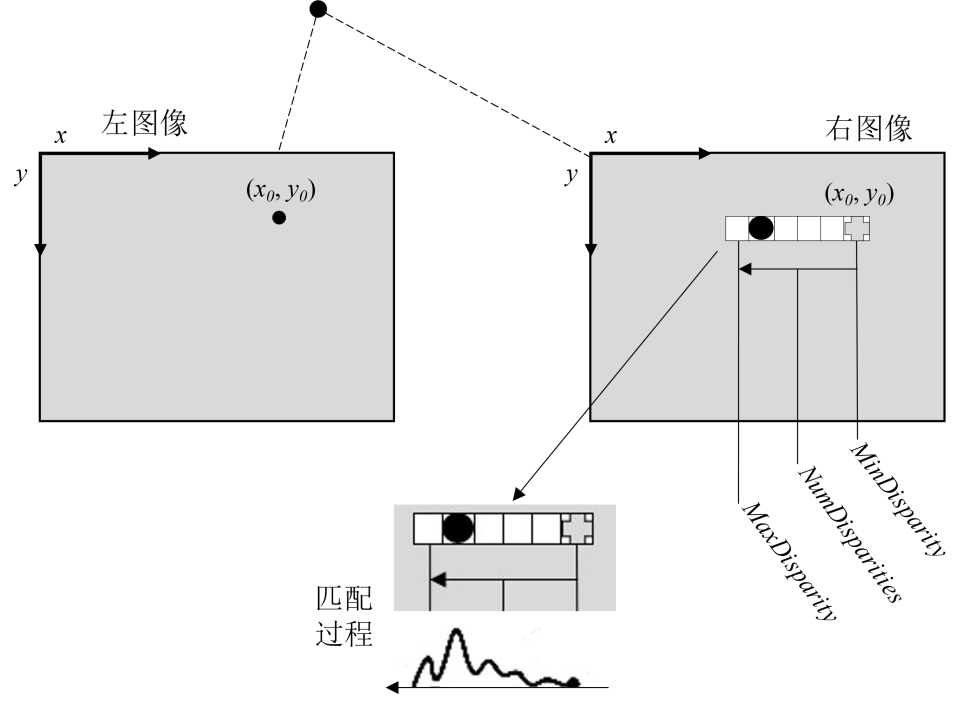


图12-15 平行双目视觉系统对应点匹配

在从右图像中搜索到与左图像中特征点匹配的像素点后，由于噪声等因素的影响，需要对这些像素点进行后置滤波，以消除错误的匹配结果。后置过滤包括两个步骤。首先，根据用户设定的视差独特性比例参数和纹理阈值参数，滤除不符合条件的像素点。然后，使用斑点滤波对剩余的像素点进行过滤。视差独特性比例参数用于描述有效匹配点的视差相对于其他像素点的独特性。其取值范围为0到100，数值越小表示匹配点相对于其他像素点的独特性越低。纹理阈值参数规定了某一像素对应的有效匹配点处的SAD应达到的最小值。通过设置视差独特性比例参数和纹理阈值参数，可以对搜索得到的潜在匹配像素进行过滤，排除不合格的部分。斑点滤波器检查每个像素周围用户指定的窗口范围内的像素，并去除超出用户设定的斑点范围或窗口视差范围阈值参数的像素。

半全局匹配算法会先计算每个像素点的视差，组成视差图，然后再设置下面和视差图相关的全局能量函数，并使这个能量函数最小，以求解每个像素的最优视差。



其中：*D*指视差图，*E*(*D*)为该视差图对应的能量函数。*p、q*为像素的坐标。*Np*为像素在*p*处的邻域大小，一般为8连通邻域。*C*(*p,Dp*)指视差为*Dp*时像素匹配过程的成本(Cost)。*P1*为用户指定的惩罚系数(Penalty)，用于更改相邻像素中视差值为1的像素。*P2*也是一个用户指定的惩罚系数，用于更改相邻像素中视差值大于1的像素。*I*[. ]是一个函数，它在中括号内的参数非0时返回1，否则返回0。

通过调整惩罚系数*P1*和*P2*，并最小化上述能量函数，可以得到理想的视差图。然而，利用该能量函数在二维图像中寻找最优解是一个极其复杂和耗时的问题。因此，需要将函数的求解近似分解为多个一维的线性问题，并使用动态规划来解决每个一维问题。由于一个像素存在着8个相邻像素，因此通常将问题分解为8个独立的一维问题。

若用*Lr*(*p,d*)为从当前像素沿*r*方向的路径上像素匹配的最小成本，则有：





则8个方向上的匹配成本*S*(*p,d*)可表示为



完成整个对应点匹配过程后，即可获得左右图像之间的视差图。而基于这些对应点之间的视差值，就能进一步计算得到被测目标的三维信息。若用*d*表示左图像中(*x,y*)点处的视差，则有：



其中，*f*为左图像对应的焦距，(*cx,cy*)为光心在左图像中的对应坐标*，*光心在右图像中的对应*x*坐标，*Tx*光为平移矩阵的*x*分量。

计算得到(*X,Y,Z,W*)T后，就可以计算得到目标点的三维坐标(*Xw,Yw,Zw*):



12.3 点云处理基础

基于立体视觉的三维重建可以实现从双目视觉图像中恢复三维点云。在三维重建过程中，需要对双目视觉图像进行图像匹配、深度估计和三维坐标计算等处理，以获取场景中每个像素点对应的三维坐标信息，从而生成点云数据。点云处理是三维重建技术的重要组成部分，可以有效地提取有用的信息、去除不必要的噪声和误差、提高点云数据的准确性和可用性。

点云(PointCloud)是由一组三维坐标表示的离散点集合，通常用于描述三维空间中物体的形状、表面和结构。点云数据可以通过各种传感器(如激光雷达、深度相机)或算法生成，具有广泛的应用领域，包括计算机图形学、机器人感知、三维重建、虚拟现实等。

点云数据可以包含各种信息，主要包括点的三维坐标和可能的其他属性，如法线、颜色、强度等。这些附加属性可以提供关于点的更多信息，例如点的法线可以用于表面重建、点的颜色可以用于渲染和可视化。

对于点云数据的处理，常见的任务包括点云滤波、点云配准、点云分割、点云特征提取等。点云滤波可以用于去除噪声或稀疏化点云数据；点云配准可以将多个点云对齐到同一个坐标系；点云分割可以将点云分割为不同的部分或对象；点云特征提取可以提取点云的描述性特征，如表面曲率、法线方向等。

点云处理的算法和工具有很多，例如PCL(Point Cloud Library)、Open3D等，这些工具提供了方便的接口和算法库，可以帮助开发者进行点云相关的处理和分析。

12.3.1 点云数据表示与结构

点云的无序性、非结构化使其与二维图像卷积存在差异，因此二维检测中研究成熟的网络不能直接用于处理点云数据，并且点云的表示形式对模型的性能有着直接影响，因此了解点云数据的表示形式是很有必要的。

点云是指获取物体表面每个采样点的空间坐标形成的点的集合。用于三维目标检测的点云通常由激光雷达扫描得来，包含点的三维坐标、强度等信息。点云的表示形式一般用点表示形式、体素表示形式和图表示形式。

点表示形式直接对点云进行处理，即采用最原始的点作为输入，这种表示形式通常基于 PointNet 网络，骨架网由点编码层和点解码层构成，编码层下采样提取语义信息，解码层将采集的语义信息传递给未采样的点，使其具备特征信息，从而保证全部的点都包含特征信息.最后基于这些点得到候选框。典型方法如 F-PointNet (Frustum-PointNet) 、Point RCNN等。

点表示方法因为使用最原始的点云数据，保留最丰富细致的信息，在所有方法中输入信息损失最小。但是，点表示方法需要处理的数据量较大，运行速度较慢，并且一般使用多层感知器，感知能力较差。

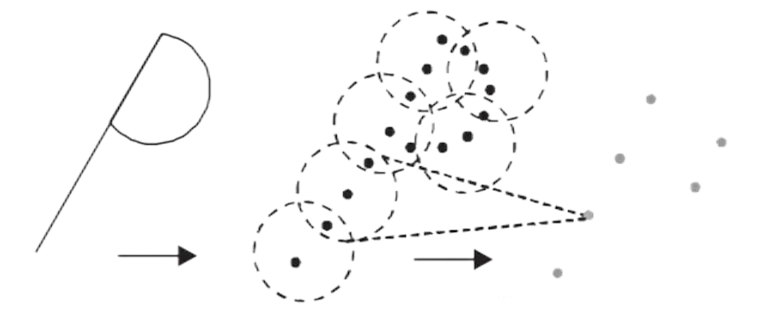


图12-16 点表示形式示意图

体素是体积元素的简称，是数字数据在三维空间分割上的最小单位，类似于二维空间的最小单位像素，数据表示形式如图12-17所示。体素表示形式将点云转化为规则的体素形式，对点云进行处理。点云体素化首先需要设置参数，包括体素大小及每个体素可容纳的点云数量。然后依次根据坐标得到每点在体素的索引，并根据索引判断此体素种类是否已达到设置的最大值。若达到，丢弃此点；未达到，保留。最后提取体素特征，进行回归预测。基于体素的典型方法包括：VoxelNet( Voxel Network)、Voxel-FPN(Voxel-Feature Pyramid Network)等。

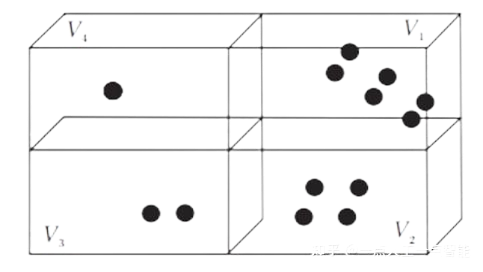


图12-17 体素表示形式示意图

基于体素的方法不仅性能较优，计算速度也较可观，尤其是稀疏卷积的发展，促进体素方法的应用。但是，基于体素的方法受设置参数的影响，不可避免地丢失一部分点云信息。

Shi 等提出图表示形式，如图12-18所示。核心在于构建图神经网络，再通过图卷积进行特征提取。图神经网络的计算费时，对于应用是一个严重限制。但是图表示方法能较好地适应点云的不规则性，并且可得到更多局部信息，因此是一种有发展潜力的点云数据表示形式。

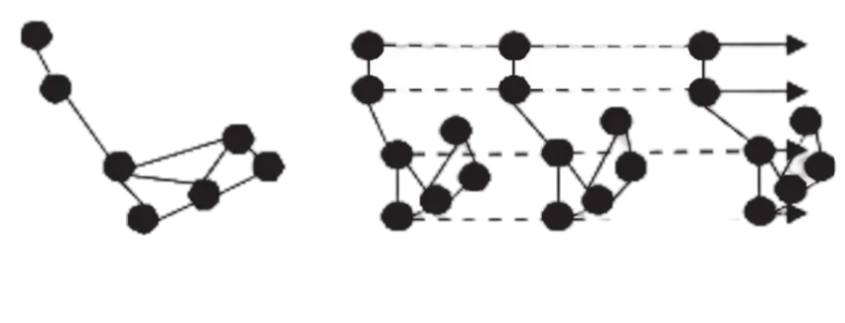


图12-18 图表示形式示意图

除上述3种表示形式以外，还有将点云投影为二维鸟瞰图、点云与体素混合等方式。与点云不同，图像的表示形式较单一。图像和点云作为两种常见的数据模态，各有优缺点：图像可提供丰富的纹理，但缺乏深度信息，对光照要求较高；点云可提供深度信息但较稀疏:两者的具体差异如表所示。

表12-1 图像和点云的差异性

|  |  |  |
| --- | --- | --- |
| 表示形式 | 图像 | 点云 |
| 数据结构 | 规则 | 不规则 |
| 数据类型 | 离散 | 离散 |
| 维度 | 二维 | 三维 |
| 分辨率 | 高 | 低 |

大体上，基于点云的三维检测方法精度高于基于图像的三维检测方法。在造价方面，激光雷达(LiDAR)价格远高于摄像机，这是研究基于图像的三维检测方法的一个重要原因。学者们关注图像和点云融合方法，以期达到更好的检测效果。

点云数据的结构会根据不同的采集设备、应用场景和数据处理任务而有所差异。在处理点云数据时，了解数据的结构有助于选择合适的算法和工具，并进行有效的数据操作和分析。

点云的数据结构可以采用多种方式来表示，其中最常见的有以下几种:

XYZ表示法是点云数据只最简单的表示形式，每个点由其三维坐标值(x,y,z)唯一确定。这种表示法适用于无需额外属性或特征的简单场景。

在XYZ表示法的基础上，XYZRGB表示法为每个点添加了RGB颜色信息。每个点的属性由其三维坐标和对应的RGB颜色值(x,y, z,r,g,b)组成。该表示法广泛应用于计算机图形学领域。

XYZN表示法在XYZ表示法的基础上，为每个点添加了法线向量信息。法线向量描述了点在曲面上的方向，常用于表征点云的表面几何特征。每个点的属性由其三维坐标和对应的法线向量(x,y,z,nx,ny,nz)组成。

12.3.2 点云数据预处理

在实际的点云数据模型获取中，由于物体本身的遮挡、光照不均匀等原因，三维激光扫描设备对复杂形状物体的某些区域容易扫描为视觉盲点，因此容易造成扫描“盲区”，形成孔洞。同时由于扫描设备测量范围有限，对于大尺寸物体或者大范围场景，不能一次性进行完整测量，必须多次扫描测量，因此扫描结果往往是多块具有不同坐标系统且存在噪声的点云数据，不能够完全满足人们对数字化模型真实度 和实时性的要求，所以需要对三维点云数据进行去噪、简化、配准以及补洞等预处理。

点云数据预处理中涉及到的几个主要研究内容，即点云去噪声、点云简化、点云配准以及点云补洞等。通过数据预处理，可以有效剔除点云中的噪声和外点，在保持几何特征的基础上实现点云数据简化，并将不同角度扫描的点云统一到同一坐标系下，为后续的曲面构建及三维实体模型生成提供稳健的数据基础。

12.3.3 点云配准与坐标变换

点云配准(Point Cloud Registration)指的是输入两幅点云Ps (source) 和Pt(target) ，输出一个变换矩阵T使得T(*Ps*)和*Pt*的重合程度尽可能高。或者说，对于两个不同视角下的坐标系，比如世界坐标系和相机坐标系，我们需要求出一个变换矩阵T使得两个坐标系变换到统一视角下。我们这里只考虑刚性变换，即变换只包括旋转、平移。

点云配准早先主要应用于建筑行业中的建筑信息模型(BIM)、采矿行业中的矿区开采等，而在自动驾驶领域内的作用主要有三类，分别为三维地图构建、高精地图定位、姿态估计。

第一，三维地图构建。在建高精地图时，自动驾驶系统通过激光雷达采集回来的相邻帧点云进行点云配准后，将不同位置采集回来的点云统一到一个坐标系下，然后构建出一个三维的高精度地图。

第二，高精地图定位。在自动驾驶车辆行驶时，车辆需要做到厘米级的精准定位。例如，自动驾驶车辆需要估计其在地图上的精确位置及车与道路路沿的距离。自动驾驶系统通过点云配准技术，将实时采集到的点云数据与高精地图的数据做匹配，为自动驾驶车辆给出精确的定位。

第三，姿态估计。自动驾驶系统通过点云配准技术来估计车辆的相对姿态信息，然后并有利于对车辆做决策规划。

如何使点云配准方法更加快速准确已成为一个点云研究的热点和难点。点云配准要应对点云数据的无序性、非结构化、不均匀和噪声等干扰。如何有效地利用已有的信息实现精确、鲁棒的点云配准算法具有重要的研究意义和价值。

点云配准的阶段,点云配准分为粗配准(Coarse Registration)和精配准(Fine Registration)两个阶段。点云配准的过程就是矩阵变换的过程。

粗配准是在源点云与目标点云完全不知道任何初始相对位置的情况下，所进行的配准方法。该方法的主要目的是在初始条件未知的情况下，快速估算一个大致的点云配准矩阵。对于任意初始状态的两片点云，使得两片点云大致对齐，给旋转矩阵R和平移向量T提供初值。整个计算过程要求比较高的计算速度，对于计算结果的精确度则不做过高的要求。常见的粗配准算法的思路包括了：基于全局搜索思想的配准方法和基于几何特征描述的配准方法。

RANSAC(RAndom SAmple Consensus,随机采样一致)方法原理：该算法从给定的样本集中随机选取一些样本并估计一个数学模型，将样本中的其余样本带入该数学模型中验证，如果有足够多的样本误差在给定范围内，则该数学模型最优，否则继续循环该步骤。

RANSAC算法被引入三维点云配准领域，其本质就是不断的对源点云进行随机样本采样并求出对应的变换模型，接着对每一次随机变换模型进行测试，并不断循环该过程直到选出最优的变换模型作为最终结果。

具体步骤：

1)对点云进行降采样和滤波处理，减少点云的计算量。

2)基于降采样和滤波处理后的点云数据，进行特征提取。

3)使用RANSAC算法进行迭代采样，获取较为理想的变换矩阵。

4)使用所获得的变换矩阵进行点云变换操作。

优点：适用于较大点云数据量的情况，可以在不考虑点云间距离大小的情况下，都能实现点云的粗配准。缺点：存在配准精度的不稳定的问题。

4PCS(4-Points Congruent Sets，全等四点集)，方法原理：该算法利用刚体变换中的几何不变性(如向量/线段比例、点间欧几里得距离)，根据刚性变换后交点所占线段比例不变以及点之间的欧几里得距离不变的特性，在目标点云中尽可能寻找４个近似共面点(近似全等四点集)与之对应，从而利用最小二乘法计算得到变换矩阵，基于RANSAC算法框架迭代选取多组基，根据最大公共点集(LCP)的评价准则进行比较得到最优变换。

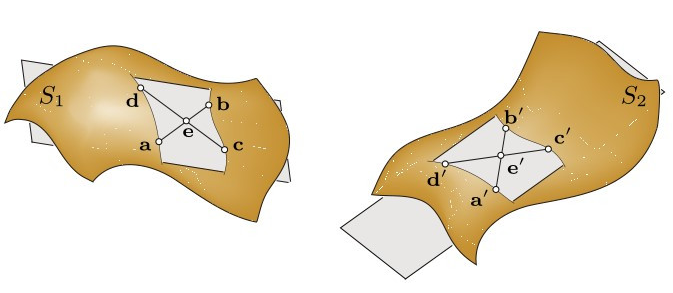


图12-19 全等四点集

具体步骤：

1)在目标点云集合中寻找满足长基线要求的共面四点基(基线的确定与输入参数中overlap有很大关系，overlap越大，基线选择越长，长基线能够保证匹配的鲁棒性，且匹配数量较少)。

2)提取共面四点基的拓扑信息，计算四点基间的两个比例因子。

3)计算四种可能存在的交点位置，进而计算所有中长基线点对的交点坐标，比较交点坐标并确定匹配集合，寻找到对应的一致全等四点。

4)寻找点云中所有的共面四点集合，重复上述步骤可得到全等四点集合，并寻找最优全等四点对。

优点：适用于重叠区域较小或者重叠区域发生较大变化场景点云配准，无需对输入数据进行预滤波和去噪。缺点：不适合工程化应用。

精配准是在粗配准的基础上，进行更精确、更细化的配准，进一步计算两个点云近似的旋转平移矩阵。精配准的主流方法包括ICP、NDT、深度学习等。

ICP(迭代最近点算法)是一种古老而又经典的点云匹配算法，其思路也很清晰。点云配准就是求解两堆点云之间的变换关系——也就是旋转关系R和平移关系t。ICP算法的思路就是：找到两组点云集合中距离最近的点对，根据估计的变换关系(R和t)来计算距离最近点对经过变换之后的误差，进过不断的迭代直至误差小于某一阈值或者达到迭代次数来确定最终的变换关系。

点到点ICP方法实现过程用数学语言可以描述为：给定待配准点云 Ps，也就是当前传感器扫描到的点云，里面由 n 个三维点pi;组成同时我们有目标点云Pt，可以是预先建好的高精点云地图，也可以是其他位置传感器采集的但有重叠区域的点云，里面同样包含一系列三维点pi;组成。

2

而我们要做的就是寻找一个3x3旋转矩阵R和3x1平移矩阵t，使得下式取值最小。

1

ICP算法的求解，ICP算法求解可以分为两种情况已知对应点的情况和未知对应的情况。

已知对应点时，如果已知两个点云中点的对应关系，那么ICP算法的求解过程将非常简单，这个在视觉[slam](https://so.csdn.net/so/search?q=slam&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/u014709760/article/details/_blank)中较为常用，因为我们可以根据特种匹配的方式找出空间点的对应关系。

具体方法为：

(1)计算两组点云的质心

3

(2)计算两组点云中的点以质心为原点的坐标

4

(3)计算w并对其进行SVD分解

5

(4)则两组点云之间的变换关系为(即ICP算法的解)：

6

已知对应点的情况下，我们可以一次性计算出点云之间的变换关系。但是在激光slam中，我们并不知道两组点云之间的对应关系。因此也就不能通过一次计算就求解出点云之间的变换关系。这时候我们的策略就是使用迭代的方式。

具体方法为：首先，寻找两组点云中距离最近的点对。然后根据找到的距离最近点对，来求解两组点云之间的位姿关系。其次根据求解的位姿关系对点云进行变换，并计算误差。最后若误差没有达到要求，则重复1，2，3步直至误差满足要求或达到最大迭代次数。

坐标变化，一般点云生成过程所涉及到的坐标系统主要包括激光扫描仪坐标系、惯导坐标系、当地水平坐标系、地心地固坐标系。坐标系的变换如下所示：

扫描坐标系(Scanner’s Own Coordinate System)，原点*O*为激光发射点，*X*轴指向载体前进方向，*Y*轴垂直向上, *Z*轴垂直于*X*轴，构成右手系。

惯性平台IMU坐标系(Body Frame System),原点*O*为惯性平台参考中心，坐标系按惯性平台内部参考标架定义，*Ｙ*轴指向载体纵轴向前，载体前进方向向右为*Ｘ*轴，*Ｚ*轴垂直向上构成右手系。

当地水平坐标系坐标原点(Local-Level Frame System)，原点位于GNSS相位中心，*Ｘ*轴指向东、*Ｙ*轴指向真北方向、*Ｚ*轴沿椭球法线方向构成右手系。根据坐标轴方向不同，当地水平坐标系可选为东北天、北东地、和北西天等右手坐标系，不同的选择方式主要和应用场景有关。

地心地固坐标系(Earth Center Earth Fixed Frame System)，简称地心坐标系，是一种以地心为原点的地固坐标系。原点*O*为地球质心，*Ｚ*轴与地轴平行指向北极点，*Ｘ*轴指向本初子午线与赤道的交点，*Ｙ*轴垂直于XOZ平面(即东经90°与赤道的交点)构成右手坐标系。

扫描仪坐标系与点坐标获取，三维激光扫描仪通过非接触式测量方式获得扫描仪中心点到目标点的距离以及角度值,根据转换公式即可求得目标点在扫描仪坐标系下的坐标。

7

扫描仪坐标系到惯导坐标系，仪坐标系与惯导坐标系之间的转换参数包括三个旋转角(α、β、y)以及三个平移量X、Y 、Y。扫描点坐标为(x、y、z)P，在惯导坐标系下的坐标为(X、Y、Z)IMT，根据公式12-21通过平移和旋转，得到惯导坐标系下激光点坐标。

8

其中：

9

惯导坐标系到当地水平坐标系，惯导坐标系前进方向为*Y*轴、前进方向向右为*X*轴、向上为*Z*轴。惯导记录实时姿态角:横滚角Rol1、俯仰角Pitch、偏航角Heading。设Rol1、Pitch、Heading 分别为*r、p、h*。扫描点在当地水平坐标系下的坐标为(*X,Y,Z*)*LOCAL*则惯导坐标系下的坐标转则可换为当地水平坐标系下的坐标。

10

其中：

111

当地水平坐标系到ECEF坐标系，当地水平坐标系原点*О*在ECEF坐标系下的大地坐标为(*B,L,H* )。设扫描点在ECEF坐标系下的坐标为(*X,Y,Z*)ECEF，当地水平坐标系下的坐标可转换为ECEF坐标系下的坐标。

122

222

其中：(*X,Y,Z*)O,为当地水平坐标系原点在ECEF坐标系下的空间直角坐标。综上所述，最终扫描点定位方程可为:

22222

其中：是扫描仪坐标系下扫描点三维坐标系，是点到目标点的扫描距离，是扫描水平角度，是扫描竖直角度。是扫描仪坐标轴系到载体坐标系的旋转矩阵。α、β、是视准轴角度。是扫描仪坐标轴系与载体坐标系原点之间的杆臂值。是载体坐标系到当地水平坐标系ENU之间的旋转矩阵。当地水平坐标系到ECEF坐标系之间的旋转矩阵。组合导航系统中心在ECEF参考框架下的三维坐标。激光扫描点*Р*在ECEF参考框架下的三维坐标。B、L、H是纬度、经度、椭球高，r、p、h是滚动、俯仰、航向角，均由组合导航系统提供。

思考与本章习题

1. 已知一个平视双目视觉系统中，同源像素点的坐标为(130, 120)和(65, 120)，相机的焦距为15mm，基线长度为0.3m，尺寸为dx×dy=5μm×5μm，图像分辨率为1200pixel×1200pixel，主点位于图像中心，请计算同源点所对应三维点的世界坐标系坐标。
2. 在一个双目视觉相机系统中，左右相机具有相同的焦距f=25mm，且具有相同的图像分辨率1000pixel×1000pixel和相同的像元尺寸du=dv=6μm，并且右相机相对于左相机的姿态变化以及卡尔丹角的方式表示为x=1°，y=0.1°, z=0.03°，且右相机相对于左相机的平移变换表示为T=(0.65,0.13,0.11)，在左相机图像坐标系中，某点的坐标为(ui, vi)=(0.11, 0.05)。求该点在右相机图像坐标系中的极线方程，并写出该立体视觉系统的本质矩阵和基础矩阵。
3. ICP算法的具体步骤。
4. 为什么要进行点云数据预处理？

