

## 第五章 距离信息的检测

### 5.1 双目立体视觉

#### 5.1.1 概述

##### 5.1.1.1 工作原理

##### 5.1.1.2 匹配特征的选择

##### 5.1.1.3 匹配规则

##### 5.1.1.4 算法简介

#### 5.1.2 Marr-Poggio-Grimson 算法

#### 5.1.3 Baker-Binford 算法

#### 5.1.4 摄象机的标定

### 5.1 双目立体视觉

#### 5.1.1 概述

如果能从两个不同的位置观察同一物体，我们就能用三角计算方法测量摄象机到该物体的距离。这种方法被称为立体视觉或双目立体视觉（Stereo 或 binocular Vision），或简称为体视。体视是人类获取环境三维信息的主要途径。人类的许多能力，如识别和定位物体，回避障碍物，和搜索物体等都要依靠体视。因此人类视觉系统具有高度发达的体视功能，可以在相当大的范围内实时地提供关于周围物体相对位置的准确信息，体视的距离可远达 1000 米。体视可经受各种干扰，在各种光照条件和光度学及几何学畸变的条件下仍能可靠地提供立体信息。体视可经受对比度的变化，在一幅图相对于另一幅图有明显的模糊或扩展时，仍能工作良好。体视的处理是快速的，并能很好地处理物体运动的情况。体视对深度信息检测的分辨率很高。在理想条件（孤立边缘）下能可靠地分辨小于 1 秒弧的视差<sup>[Wes 78]</sup>。这相当于在 1 米的观察距离上确定大约相距 0.8 毫米的两个特征的相对深度，或在 50 厘米远处的 0.2 毫米的相对深度。

在计算机视觉研究中立体视觉也是很重要的，因为它可适用于各种条件。例如，体视可被用于根据航空照片获得地形信息，这时就难以应用主动式的测距方法。对体视的研究从根本来说有两种不同的目的和方法：一种是为了理解人类双目立体视觉的机理；另一种是找寻获得距离信息的实用方法。前者寻求人类立体视觉的计算机模型，它可适用于各种情况，因此需要进行大量的计算；与此相反，后者希望开发可实用的立体视觉系统，由于它是适用于特定的领域，因此通常是不通用的。由于体视处理需要进行大量计算，因此目前在实用中还未被广泛采用。近来由于在高速信号处理器硬件研究方面取得迅速进展，以及并行处理技术的发展，使得有可能应用通用的并行处理器来解决体视处理中的计算量问题。此外，上述两种方法之间的相互渗透和启发能为发展实用的通用立体视觉系统指出新的途径。因此，对体视的研究再次引起各方的重视。

##### 5.1.1.1 工作原理

图 5.1(a)所示为用双摄象机观测同一景物时的情形。物体上的点 P 在摄角机 1 中的成

象点为 A，它是通过从 P 点发出的光线经过透镜中心  $C_1$  与图象平面相交而形成的。相反地，若已知图象平面上的一点 A 和透镜中心  $C_1$  可唯一地确定一条射线  $AC_1$ 。所有可成象在 A 点的物体点必定在这条  $AC_1$  射线上。但问题是不知道物体在这条射线上的什么地方，也就是不知道离得多远。如果我们能找到同一物体点 P 在另一摄象机中的成象点 B，那么根据第二个图象点 B 与相应透镜中心  $C_2$  决定的第二条射线  $BC_2$  与  $AC_1$  的交点就可以确定物体点的位置。因此，如果已知两台摄象机的几何位置，并且摄象机是线性的，同时知道同一物体在两个摄象机中的成象位置，那么利用三角原理就可以计算物体在空间的位置。射线

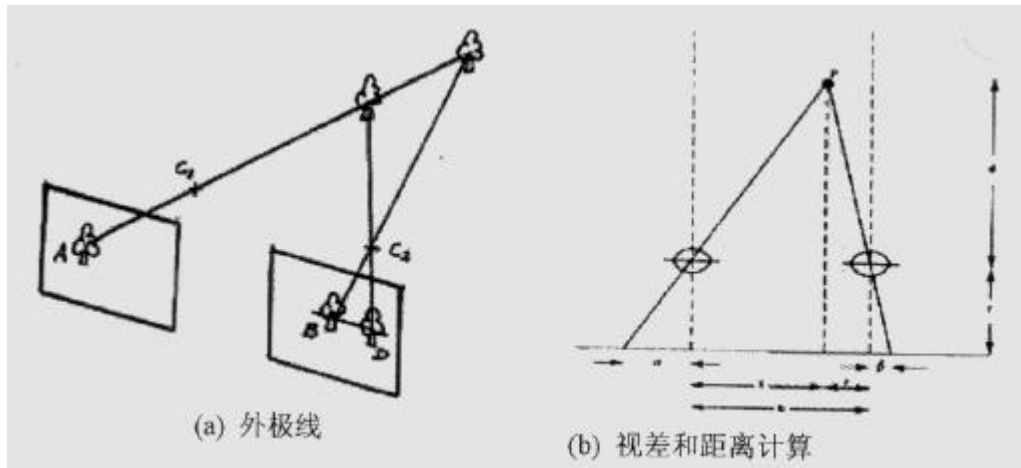


图 5.1 立体视觉原理

$AC_1$  上各点在右摄象机图象平面中的成象是一条直线 (BD)，这条线被称为外极线 (epipolar)。同理， $BC_2$  在左摄象机图象平面中的成象也形成外极线。因此，如果已知空间点在一个图象平面中的成象点要寻找在另一图象平面中的对应点时，只需沿此图象平面中的外极线搜索即可。图 5.1(b) 所示为两摄象机的光轴平行，并且摄象机的水平扫描线位于同一平面时的简单情形。P 点在左、右图象平面中成象点相对于坐标原点  $O_1$  和  $O_2$  ( $O_1$  和  $O_2$  是左、右摄象机透镜光轴与图象平面的交点) 的距离分别为  $\alpha$  和  $\beta$ 。P 点在左、右图象平面中成象点位置差  $\alpha + \beta$  被称为视差 (disparity)。在图 5.1(b) 所示情况下，P 点距透镜中心的距离  $d$  等于

$$d = \frac{fb}{a+b} \quad (5-1)$$

其中  $f$  是透镜的焦距， $b$  是两透镜中心之间的距离，当摄象机的几何位置固定时，视差  $\alpha + \beta$  只与距离  $d$  有关，而与 P 点离摄象机光轴的距离无关。视差越大说明物体离透镜的距离越近；反之，则越远。

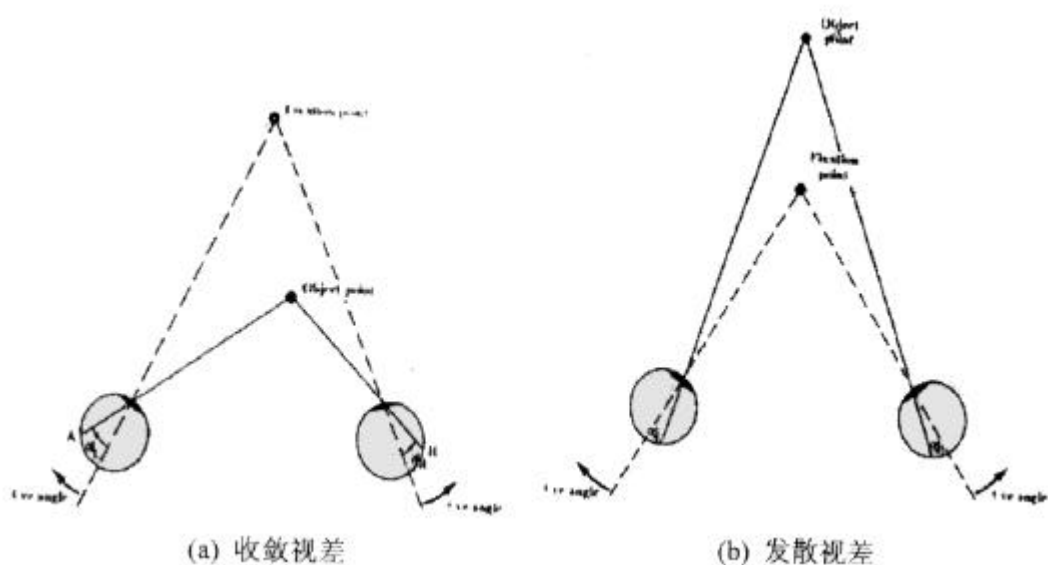


图 5.2 双目光轴不相平行时的视觉

在一般情况下左、右摄象机（双目）的光轴不平行，而是相交于某一点（称为固定点），如图 5.2 所示。固定点的视差为零。如果物体点在固定点的前方（离透镜较近，图 5.2 (a)）这时的视差称为收敛视差（convergent disparity）。在图 5.2 中用双目代表摄象机，在研究人的视觉时常用视角差来表示视差。如果物体点在固定点后方（图 5-2(b)），这时的视差称为发散视差（divergent disparity）。这时的视差直接反映物体点距固定点的距离，而固定点的位置可通过改变摄象机光轴的夹角进行调整。因此，改变摄象机光轴的夹角可以调整距离测量范围。例如，人在观察近处的物体时就需要把双目的光轴会聚在近处。

为了避免混淆需要说明一下距离和深度的定义。距离是指从观察者到物体的客观实际距离；深度（depth）是指由观察者感觉到的主观距离，通常是测量相对于定位点或某个空间点的距离。

如上所述，从原理上讲根据“立体图象对”抽取深度信息的处理应包括以下四部分：

1. 在图象中寻找在两幅图象中都便于区分的特征，或用于匹配的基元（primitive）。
2. 把左、右两幅图象中的有关特征进行匹配，即解决特征匹配的方法问题。
3. 确定摄象机的相对几何位置和有关参数，即摄象机的校准（Calibration）。
4. 根据视差计算成像物体相对摄象机的距离。

这些问题中最重要和困难的是前两个问题。也就是在左、右图象中发现与同一空间点对应的成像点对，这说是所谓的对应性（Correspondence）问题。一旦确定了对应关系就可容易地计算出这些图象点所代表的物体点在空间的位置。但是对一幅图中的给定匹配基元来说在另一幅图中经常可发现不止一个可能的匹配基元与之匹配。这样就产生了匹配中的多义性或匹配假目标问题。这是个关键而困难的问题。

### 5.1.1.2 匹配基元的选择

对应性问题不是简单地把两幅图中象素的灰度作比较就能解决的。左、右图象中单个象素点的灰度不够稳定，即使认为它们是稳定的话，也很容易出现在相当大的区域里象素具有相同灰度的情况，这样就造成了严重的多义性问题，或假目标（false target）问题。

立体视觉处理中对搜索对应点时的多义性问题可分两步来解决。第一步，是在单幅图象作预处理时通过抽取图象局部结构较为丰富的描述来减少错误对应的可能性；第二步，是在两幅图的对应点间作匹配时应用选择性规则来限制搜索空间。各种算法间的区别主要在于它们在匹配时选择什么样的匹配基元（matching primitive）作为表面位置标志的基本元素，

以及选用什么规则来限制搜索空间和删除不合适的匹配。

选择匹配基元时要考虑基元的稳定性和敏感性。由于图象对中不可避免地存在光度学和几何学的畸变。为使在这种情况下仍能可靠地检测所需的位置标志，所选的匹配基元应能经受上述两方面的畸变，也就是应有较高的稳定性。同时，从减少出现不正确匹配的可能性来看，所选的匹配基元应能灵敏地反映两个匹配基元之间的差别，这就是说要有较高的敏感性。人类的立体视觉经受图象对之间对比度差别的能力很强，这意味着在人的立体视觉中可能应用了如局部灰度梯度最大点这样的与对比度大小无关的匹配基元。在选择匹配基元时还应考虑便于检测、能准确定位和允许在较大的视差范围内进行匹配等因素。目前所用的匹配基元可以分成两大类：

#### 1. 在所有图象点上抽取的量测

这类匹配基元一般是在每个象素位置处都产生一个描述，所以这时把匹配基元看成是一种量测比看成一种特征更为确切。这些特征表示图象中的局部结构状态，在数量上要比象素少得多。属于这类的匹配基元有以下几种：

(1) 象素灰度。象素灰度可由成象系统直接得到，因此是最简单的。目前被用于大多数商用的视觉系统中。

(2) 局部区域的灰度函数。在各种大小窗口中求得的灰度分布的导数可用于产生描述各点周围结构的矢量<sup>[Kas 83]</sup>。

(3)  $\nabla^2 G$  卷积图象的符号<sup>[Nis 83]</sup>。把图象与各种大小的  $\nabla^2 G$  算子卷积后，图象中各点的符号可作为原始图象特征的描述。在卷积后的图象中可得到正号区和负号区。这两个区域之边界接近于灰度梯度局部极大值的位置。

#### 2. 图象特征

这种匹配基元较为符号化，它检测图象中包含丰富信息的结构所在的位置，例如图象中的边缘，这些边缘可能与景物中表面之间的边界相对应。与象素相比图象特征数量较少。

(1)  $\nabla^2 G$  卷积图象中的过零点。这种方法是由 Marr 和 Poggio<sup>[Mar 79]</sup>，Marr 和 Hildreth<sup>[Mar 80]</sup> 提出和发展的。它虽然也可用于检测边缘，但是更确切说这种方法的目的是检测稳定的、稠密的表面标志。按这种方法任何小的影调变化或小的纹理变化只要稳定都是一个特征。

(2) 边缘。这种基元试图抽取景物中表面之间或不同颜色区域之间的实际边界。这种匹配基元上还可以带有如边缘方向、对比度、长度、边缘曲率等附加信息。检测边缘的算子如第四章中所述种类很多。在选择边缘作为特征时有两点需要考虑：第一，由于对于给定的特征点来说，对应的外极线上的点都是可能的匹配点。所以与外极线方向平行的边缘线段无法作为匹配的特征，只有其方向与外极线交叉的边缘点才能作为匹配基元。第二，因为边缘经常代表深度的不连续点，在从不同位置所取得的图象中，边缘两侧的区域情况将会不同。因此，基于边缘特征的立体视觉算法通常只利用边缘的位置和方向的信息，而对边缘两侧的灰度信息用得很少。

### 5.1.1.3 匹配规则 (matching rules)

在研究具体的匹配规则以前需要先讨论在匹配过程中应遵循的约束条件。这些约束条件是根据对匹配环境所作的假设产生的，约束条件主要包括以下三条：

#### 1. 相容性 (Compatibility) 约束

如果两个匹配基元确实是由同一物理标记产生的，那么它们就可以匹配起来。如果不是这样，它们就不能匹配。在判断两个匹配基元是否相容时要根据它们之间的相似性。问题是如何度量匹配基元的相似性。有两种相似性的假设。一种是基于光度学不变性的性质。即左、右图象对应区域中灰度的变化情况相似。如果景物中表面的深度变化比较平缓，同时由于双眼相隔的距离不大，作这样的假设是有道理的。例如，用立体视觉原理，通过航空摄影测地

形时，由于地形的起伏与飞机的高度相比较小，因此可采用这样的假设。但在机器人视觉应用中，景物的深度分布经常有急剧变化，在这样的区域附近容易产生与左图中相对应的区域在右图中被遮挡，或反之的情况。这时光度学不变性的假设就难以保持。另一种相似性的假设是根据几何学不变性，即两幅图象中描述对象的几何结构相同。例如，在以边缘作为匹配基元时，沿外极线上任何扫描方向，在左、右图象中边缘出现的次序相同（虽然由于存在遮挡，出现在左图中的边缘可能不出现在右图中，或反之）。

## 2. 唯一性约束

由于在任何时刻位于某一物质表面上的一个给定点在空间只占有一个唯一的位置，所以，除了极个别的情况以外，某个匹配基元只能与另一幅图象中的一个匹配基元相匹配。这样，图象中的每个匹配基元最多只能有一个视差值。

## 3. 连续性约束

这条约束条件的含义是匹配得到的视差值的变化在图象中几乎处处平滑。这个约束条件是以下述假设为前提的：和表面到观察者的总距离相比较，物体表面凹凸引起的变化或由观察者到表面的距离变化造成的差异都很小。因此，物体表面可看成是平滑的。也说是说，除物体的边界外，从观察者到可见表面的距离的变化是连续的，而物体的边界只占图象面积的很小部分。

上述约束条件对减小匹配多义性的作用可用下述例子来说明。如图 5.3 所示，左、右眼都可以看到 4 个点，那么左图中任意一个点到底与右图中哪一个点相对应呢？如果匹配不是一对一的，则在 4 个点的情况下，对每个点来说有  $2^4=16$  种可能的方式与另幅图象中的点相匹配。所以，从原理上来说，4 个点总共有 65536 种可能的匹配方式。根据唯一性的约束条件，来自两个眼睛的任何一条视线上都不能有多于一个的匹配，也即每条视线或无匹配点，或有一个匹配点，这将使匹配方式降为 209 种。如果进一步限制沿每条视线只有一个匹配点，那么在图 5.3 中所有的 16 个可能匹配中可以有 24 种排列组合方式。这时需要应用连续性来进一步减小匹配的多义性。连续性约束条件说明在这 24 种可能的匹配方式中最可能的是视差变化最平滑的物体表面，在图 5.3 中用实心圆表示。因此， $R_1-R_4$  应顺序地与  $L_1-L_4$  相匹配。

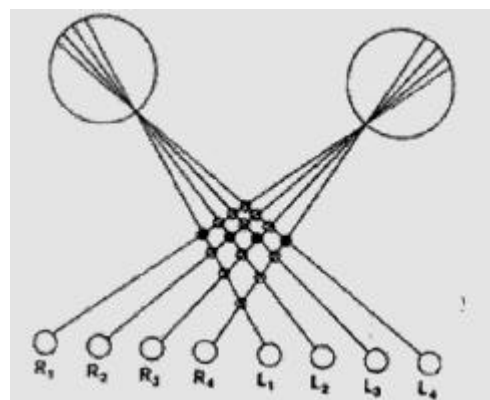


图 5.3 两个视网膜上成像对应关系的多义性

以下的问题是如何把上述一般性的约束条件结合到算法中去。这方面的规则可分成两大类，一类规定相似性测量的本质，另一类对相邻匹配基元的视差的关系作出限制。每种匹配算法至少利用这两类规则中的一种。具体来说有以下两大类规则。

### 1. 对相似性测量的本质作出规定的规则有以下几种：

(1) 区域的统计量。把图象中小区域里得到的统计量与另一幅图象作比较，以得到相似性测量。例如，在两幅图象之间进行小区域灰度分布的相关运算和视差方差的统计分析就是属于这一类。一般来说，采用这种相似性测量时要求假设在这个小区域中的视差为常数，因此相当于作了很强的表面连续性的假设。

(2) 边界的统计量。这条规则与区域统计的规则相似，不同之处是把图象特性的比较仅限于表面的边界上。边界内的表面被认为是连续的，沿着边界的视差变化是平滑的。把统计量仅限于边界有利于扩大这条规则的适用范围。

(3) 点的统计量。这条规则实际是区域统计量规则的变型，不是在一个区域里把逐点的相似性比较综合起来，而是以同一位置为中心作空间测量。

## 2. 视差梯度限制规则

这类规则根据以下两个实际观察来限制候选的匹配关系：(1)很少出现相对于两个摄像机的表面梯度很陡的情况；(2)在这种情况下基元的测量很可能是不稳定的。因此，视差梯度急剧变化的候选匹配应被抛弃。因为这样的匹配很可能是不正确的。具体来说有以下几种规则：

### (1) 排序约束

这条规则所作的限制是：在两幅图象中沿相应的外极线上的匹配基元必须以相同的次序排列。这相当于假设成象表面是不透明的，并且是连续的。这条规则可有效地减少候选匹配基元的数量。

### (2) 视差梯度范围限制

这条规则对相邻匹配基元之间允许的最大视差梯度作出限制。在人的视觉系统中视差梯度被限制在 1 单位以内，即单位长度上视差变化小于 1 个感光细胞大小（0.4' 弧度）。这条规则把由于成象条件所限难以正确匹配的对应点排除在外。

### (3) 由粗到细的匹配规则

用不同尺度（分辨率）的算子来检测匹配基元，在较粗分辨率下进行匹配所得到的信息用于限制在较细分辨率下基元匹配的搜索范围。这样既提高了匹配的可靠性，又达到了较高的分辨率。

## 5.1.1.4 算法简介

立体视觉算法可分成两大类：一类以密集的基元测量为基础，称为基于区域（area-based）的算法。这类算法的典型例子是利用小区域上的相关技术。另一类以在图象中相对比较稀少的、较为符号化的特征为基础，称为基于特征（feature-based）的算法。

在立体航空摄影的应用中对基于灰度的区域相关技术进行了深入研究。Moravec<sup>[Mor 80]</sup>和 Gennery<sup>[Gen 80]</sup>在研究用于自主式移动机器人导航的立体视觉系统时提出了两种以像素灰度为基础的算法，Tai<sup>[Tai 83]</sup>也提出了两种属于这一类的算法。这种算法使用摄像机从 8 个已知方位取得的透视图象。用他的方法得到的相关函数比一般两帧图象间的区域相关函数尖锐得多，并可使用小得多的窗口。

基于特征的立体视觉算法要求使用较为严格的匹配规则以删除不正确的匹配。

在匹配的不同阶段分别使用上述两类算法以期达到更高性能的混合算法的代表性例子是 Baker 和 Binford<sup>[Bak 80, 81]</sup>提出的算法。这种算法的初始匹配基元是带有空间分辨率、对比度、方向和灰度等属性的边缘。匹配过程被局限于根据成象几何学计算的外极线，并使用排序约束。算法从低分辨率开始，先找到两幅图象间大致的对应关系，然后对中间结果作改进，对更精细的细节作分析。接着在第一阶段所得对应关系的导引下作基于灰度的匹配。这两种匹配都依靠动态规划技术 viterbi 算法。最后对两幅图象利用如边缘连接性这样的全局约束来去除错误的边缘对应关系。

## 5.1.2 Marr-Poggio-Grimson 算法（MPG 算法）

这个算法的目的是试图模拟（至少在计算理论的层次上）人类视觉系统双目立体视觉敏

感深度信息的能力。它的特点是以不同大小的  $\nabla^2 G$  算子与图象卷积，并从中抽取过零点作为匹配基元；采用从粗到细的匹配策略，应用在低分辨率下匹配得到的信息来限制高分辨率时匹配的搜索空间。这样做的优点是既具有较大的深度敏感范围，又有较高的空间定位准确性。具体来说，这个算法的主要内容包括：

### 1. 匹配基元的选择

作为匹配基元的特征点的选择根据以下考虑：

(1) 因为在灰度均匀区域内的点难以在另一幅图中找到对应点，所以只有在其附近灰度急剧变化的点才能作为匹配基元。

(2) 灰度急剧变化点对应于图象与 Laplacian 算子卷积后的过零点。所以，特征点将是  $\nabla^2 f(x, y)$  中的过零点，其中  $\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ ， $f(x, y)$  是图象函数。

(3) 在作 Laplacian 卷积以前与 Gaussian 函数作平滑滤波，以区别不同尺度变化。

$$G(x, y) = s^{-2} \exp\left(-\frac{x^2 + y^2}{2s^2}\right)$$

(4) (2)和(3)中的算子可合并为

$$\nabla^2 G(x, y) \equiv \frac{r^2 - 2s^2}{s^4} \exp\left(-\frac{r^2}{2s^2}\right)$$

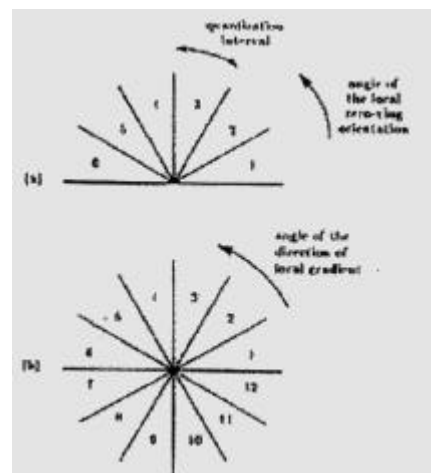
$$r^2 = x^2 + y^2$$

第二章中图 2.7 所示为  $\nabla^2 G$  算法的截面图。这个算子的宽度  $w_{2D}$  是用原点左、右第一个零点之间的距离来表示。为了避免截短 (truncate) 效应，算子的窗口宽度要大于  $3.6w_{2D}$ 。

### 2. 匹配基元的属性

可以用过零点两边的符号变化和过零点轮廓来表示匹配基元的特征。在与  $\nabla^2 G$  算子作卷积后所得图象中，图象值的符号从正到负的变化表示原图中有一个从低到高的灰度上升，反之则有一个灰度下降。显然灰度的上升和下降是两种不同的变化。

要估计局部的过零点轮廓的方向需要知道在其周围过零点的位置。Grimson<sup>[Gri 81]</sup> 用 6 个值 (图 5.4) 来表示过零点轮廓的方向。当然，也可以用计算各点灰度梯度的方向来确定它的方向。



(a)过零点轮廓的方位被量化到 6 个角度间隔；

(b)如化为 12 个角度间隔则可自动地把过零点的符号包括在内。

图 5.4 过零点轮廓的方向

### 3. $w_{2D}$ 的选择

如何选择  $w_{2D}$  是 MPG 算法的关键。因为由此可以很容易解释为什么需要进行多通道的匹配。按多通道算法，首先在  $w_{2D}$  最大的通道（粗通道）中寻找特征点，并使左、右图中的特征点相匹配得到低分辨率的景物深度图。粗通道时得到的信息被用来控制较细通道中特征点的匹配。我们首先研究粗通道时特征点的匹配。如在左图中选择一个特征点 A（图 5.5），它在右图中真正的匹配点是 B。为寻找真正的匹配点，可先把 A 点的坐标传递到右图用 X 表示。如已知最大的视差为  $d_{max}$ ，那么围绕 X 建立一个  $d_{max}$  大小的搜索区。影响粗通道时特征点匹配的因素分析如下：

(1) 搜索区内所有与 A 具有相同符号变化和过零点轮廓方向的过零点都被认为有可能与 A 匹配，但在所有可能匹配中只有一个是真实的，其余都是假目标。

(2) 假目标的数量与搜索区的大小以及  $w_{2D}$  的大小都有关。为此，Marr 和 Poggio 曾经研究了过零点的统计分析，以确定滤波图象中相邻同符号过零点之间间距的概率分布。设在图中某一过零点 L 与右图中某些一过零点 R 相匹配。Marr 和 Poggio 在随机点立体图象对中得到的概率分布表明在 R 的  $w_{2D}/2$  间距内有另一个同符号过零点的概率低于 0.05。这意味着如果图象中这个区域的视差小于  $w_{2D}/2$ ，那么在  $\pm w_{2D}/2$  的范围内搜索时只发现正确的概率是 0.95。由此可知，如果要完全避免假目标问题，那么搜索区域的范围应限制在  $\pm w_{2D}/2$ 。但

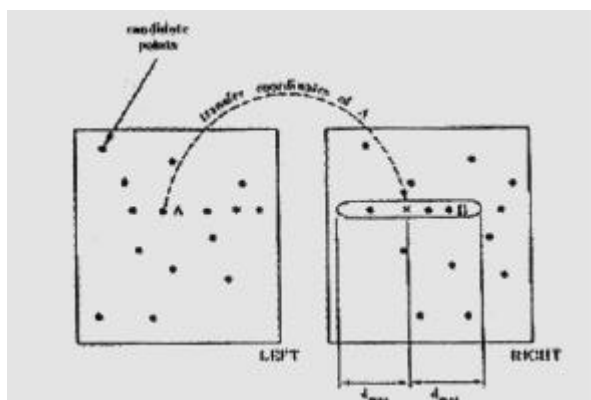


图 5.5 特征点的匹配

如把搜索区域的范围扩大到  $\pm w_{2D}$  也是可以接受的。Marr 和 Poggio 证明，如果搜索区域扩大到  $\pm w_{2D}$ ，所有匹配中的 50% 是正确和无歧义的。这意味着有的匹配是多义性的。一般这样的匹配有两个。其中一个是收敛视差（在  $(0, w_{2D})$  区域内）；另一个是发散视差（在  $(-w_{2D}, 0)$  区域内）。这两个匹配中的一个正确的。在有多义性的情况下，可利用相邻非多义性匹配的视差符号来确定那一个是正确的匹配。根据连续性约束，应取视差符号相同的匹配。这样可取  $w_{2D}$  大致等于  $d_{max}$ 。

(3) 取  $w_{2D} = d_{max}$  后，我们在  $d_{max}$  的距离内只能得到一个点的深度值。这只代表景物在粗通道时的深度图。



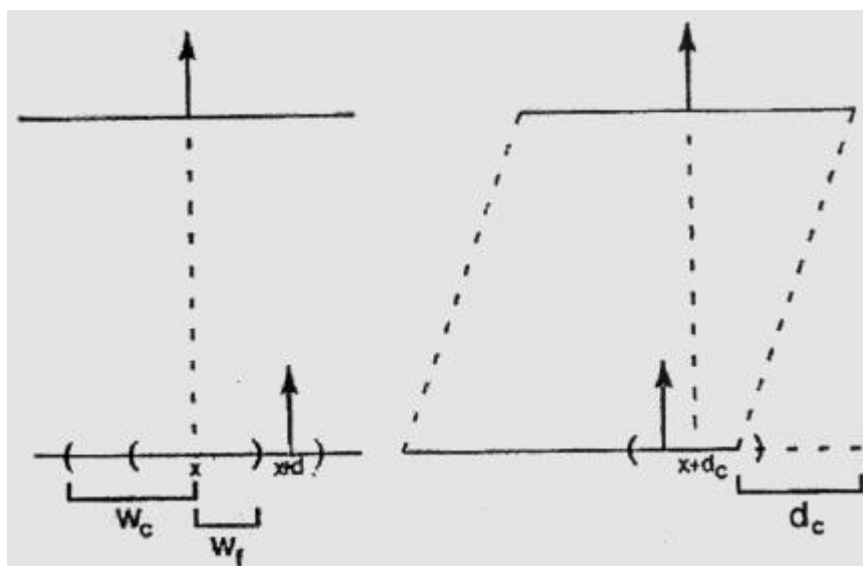


图 5.6 粗通道信息对细通道匹配的导引

左图所示为一维情况下的匹配。图中箭头表示过零点的位置，两个相应的过零点相距视差  $d$ ， $w_f < d < w_c$ ， $w_c, w_f$  分别为粗细通道的宽度。所以对应点在细通道的范围以外，但在粗通道的范围之内，右图表示可用粗通道时求得的视差  $d_c$  来对准图象，这样对应点就在细通道的搜索范围之内，并求得正确的视差  $d$ 。

#### 4. 粗通道信息对细通道匹配的导引

MPG 算法的主要想法是利用粗通道时的匹配信息来导引细通道时的匹配。下述例子可以说明这种导引的必要性。设在图中  $(x, y)$  处的点在右图中的对应点位置是  $(x+d, y)$ ， $d$  是视差。（图 5.6）。如果粗、细通道  $\nabla^2 G$  算子的宽度分别为  $w_c$  和  $w_f$ 。那么在粗细通道中合适的搜索范围分别为  $\pm w_c$  和  $\pm w_f$ 。假设  $w_f < d < w_c$ ，那么要发现这点的视差只根据细通道的信息是不行的，因为匹配点在搜索范围以外。但匹配点将在粗匹配的范围之内。因此，这时可求得粗通道中的视差  $d_c$ 。但由于是在粗通道中得到的，所以这点的视差不精确的，譬如说  $d_c \neq d$ 。然而，由粗通道得到的匹配可为精确的视差提供一个近似的估计值。如果所选的滤波器的大小合适，则可保证  $|d_c - d| < w_f$ 。这样，如果我们应用视差的初步估计来对准图象，则如图 5.6 中右图所示，细通道中的搜索可集中在点  $(x+d_c, y)$  附近，而不是在点  $(x, y)$  附近。这样，虽然细通道中的搜索范围减小了，但由于搜索是集中在大致正确的图象范围内，我们仍能在细通道之下实现成功的匹配。总之，通过粗、细通道处理的结合既可在大的视差范围内检测高分辨率的视差，又同时避免了假目标问题。

现在已经知道人类视觉系统中视差计算是在从粗到细的 5 个通道中进行的。相邻通道之间大致相隔一倍频程，即  $w_c = 2w_f$ ，其中  $w_c$  和  $w_f$  分别为相邻粗、细通道的宽度。这些通道所用  $\nabla^2 G$  算子的宽度  $w_{2D}$  大致为 63, 35, 17, 9 和 4 个像素。这里一个像素是指一个视网膜中央凹感光细胞的大小，这大致相当于  $0.4'$  弧度。人类视觉系统可确定图象中大约只有  $5''$  弧度大小的特征。Marr, Poggio 和 Grimson 认为对这种高度敏锐性（Hyperacuity）的一种可能解释是假设在确定孤立过零点的位置时进行了内插，使定位精度高于感光细胞的分辨率。过零点是一个像素与相邻像素之间有符号变化的情形。在这样的两像素之间作内插就可把边缘定位到分像素（subpixel）的精度。

### 5.1.3 Baker-Binford 算法

如前所述，立体视觉的两大类算法：基于特征的算法和基于区域的算法各有优缺点，并适用于不同的领域。Baker-Binford 算法试图把这两种方法结合起来，希望能有更大的适用范围和更高的性能。上一节中的 MPG 算法就属于基于特征的算法类型。对于基于区域的方法我们把它作为 Baker-Binford 算法的一部份，介绍它的相似性检测方法（图象灰度相关）和对应点搜索方法（动态规划匹配方法）。然后介绍 Baker-Binford 算法本身。

#### 1. 基于图象灰度相关的相似性量测

在基于区域的立体视觉算法中，给定左图中的一个点后在右图中是根据邻域的相似性来寻找其对应点的。邻域可被称为窗口。窗口通常以检测点为中心。已提出了许多种相似性量测函数，其中最常用的是左、右窗口中灰度的相关函数。如窗口是  $M \times N$  的矩形，设，左、右窗口中第  $i$  行、第  $j$  列象素的灰度分别为  $I_L(i, j)$  和  $I_R(i, j)$ ，则相关函数  $C$  可定义为

$$C = \frac{s_{LR}^2}{\sqrt{s_L^2 + s_R^2}} \quad (5-2)$$

其中  $s_L^2$  和  $s_R^2$  分别左、右窗口中灰度的方差。

$$s_K^2 = \sum_{j=1}^n \sum_{i=1}^m [I_K(i, j) - m_K]^2 / (m * n) \quad (5-3)$$

$K = L \text{ 或 } R$

其中  $m_K$  是有关窗口中灰度的均值， $s_{LR}^2$  是  $I_L(i, j)$  和  $I_R(i, j)$  的协方差，表示为

$$s_{LR}^2 = \sum_{j=1}^n \sum_{i=1}^m \{[I_L(i, j) - m_L][I_R(i, j) - m_R]\} / (m * n) \quad (5-4)$$

最简单的相似性量测函数可直接取作灰度差之和。这时我们把它称为差异性量测函数  $D$ 。为实现归一化可把它除以方差，即定义为：

$$D = \sum_{j=1}^n \sum_{i=1}^m [I_L(i, j) - I_R(i, j)]^2 / (s_L s_R) \quad (5-5)$$

为简单起见也可取一个方差作除数

$$D = \sum_{j=1}^n \sum_{i=1}^m [I_L(i, j) - I_R(i, j)]^2 / s_L^2 \quad (5-6)$$

左、右图象的灰度差也可能是由于摄像机特性不同造成的。为克服这个困难可用灰度均值和方差来归一化，即把差异性量测函数定义为

$$D_n = \sum_{j=1}^n \sum_{i=1}^m \{[I_L(i, j) - m_L] / s_L - [I_R(i, j) - m_R] / s_R\}^2 \quad (5-7)$$

$D_n$  越大说明左、右图象中对应点的相似性越差。

#### 2. 动态规划 (dynamic Programming) 匹配方法

在另一幅图象中寻找给定点的对应点的基本策略是计算外极线上所有候选点的差异性量测值，并选择其中的最小点作为对应点。在此过程中有两个问题要解决：第一，是选择计算相似性量测时窗口的大小；第二，是如何判断是否存在正确的对应点。一般来说，大的窗口适宜于得到全局的深度信息，但难以准确地确定对应点的位置。相反，小的窗口可较准确地定位，但对噪声敏感，或可能出现多个匹配峰点的情况。一种克服这些困难的方法是动态规划匹配法。这种方法不是孤立地寻找单独的匹配，而是按某种准则使水平扫描线上的一组点与另一幅图中相应水平扫描线上的一组点相匹配。

两个波形之间的对应关系可被概括成路径规划问题（图 5-7）。图中纵轴表示左图中扫

描线的位置，横轴表示右图扫描线上的位置。路上  $(i, j)$  处点 P 表示右图在  $i$  处的点与左图  $j$  处的点相匹配。搜索对应点的问题被归结为发现花费代价最小的最佳路径问题。路径成本可用前段中所述差异性量测函数  $D_n$  来定义，把路径成本定义为  $D_n$  沿扫描线的积分。动态规划技术可有效地解决这样的路径规则问题，并把它称为动态规划 (dynamic programming, DP) 匹配问题。如果景物中的表面是连续的，并且表面的每一部分在两幅图中都可被观测到，那么规划所得路径是连续的，并且是单调增长的。

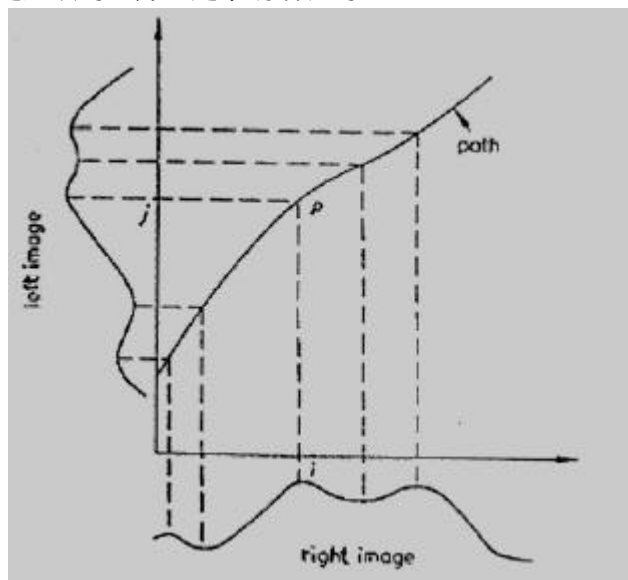


图 5.7 动态规划匹配问题

设， $D(i, j)$  表示右图中在  $i$  处周围灰度分布与左图在  $j$  处周围灰度分布的相似性量测，路径成本可定义为沿路径的相似性量测之和。设  $F(i, j)$  是从起点到  $(i, j)$  点最佳路径的最小成本。DP 算法可表示为

$$(1) \quad F(i, 1) \leftarrow D(i, 1) \\ F(1, j) \leftarrow D(1, j)$$

$$(2) \quad \text{For } 2 < i, j < n$$

$$F(i, j) \leftarrow \min [D(i, j) + F(i-1, j-1), D(i-1, j) + F(i-2, j-1), D(i, j-1) + F(i-1, j-2)]$$

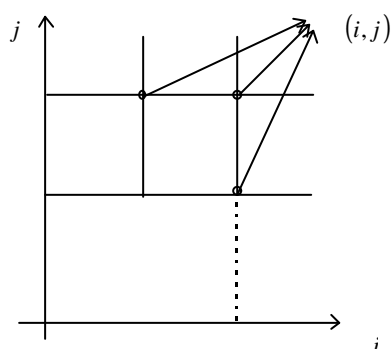


图 5.8 可能的连接弧

其中  $n$  是包含在扫描线上的象素数量，在第二步中计算过程是从小的  $i$  和  $j$  到较大的  $i$  和  $j$  进行，一种合理的前进路径是沿  $i+j$  之和的方向。按这种匹配算法，如图 5.8 所示，在迭代的每一步只允许沿三条弧的方向。当然，也可以允许沿更多不同方向的弧前进，这时算法公式要复杂一些。总之，通过限制下一步可允许的前进方向可有效地压缩迭代中每步的搜索空间。这样的限制是根据连续性提出的。DP 匹配的立体视觉算法很适合于由航空立体照片得到连续地形图的应用场合。

如果景物中包含不连续的表面，那么算法中的第(2)步要修改。如果部分表面在左图中可被观察到而在右图中观察不到，那么与这部分表面对应的弧应是垂直的。这样在第(2)步中就需要考虑更多的路径。困难的问题是如何定义这部分表面上灰度分布的差异。因此，应用其它的方法来删除表面上的不连续的部分。只有在表面

已知是连续的区域中才能应用 DP 匹配方法。

### 3. (Baker-Binford) 算法<sup>[Bak 81]</sup>

算法分成两部分。第一部分是建立左、右图象中边缘的对应关系。边缘的描述给出景物的基本结构，产生较为稀疏的视差测量。然后算法的第二部分通过图象灰度的相关，求得更多细节处的距离信息。最后得到整幅距离图。

#### (1) 基于边缘的相关

图 5.9 所示为左、右图象中沿某条相应外极线的灰度分布，用边缘检测算子得到的边缘位置也重叠在此图中。相关过程就是沿外极线搜索边缘的最佳匹配，由于同一条外极

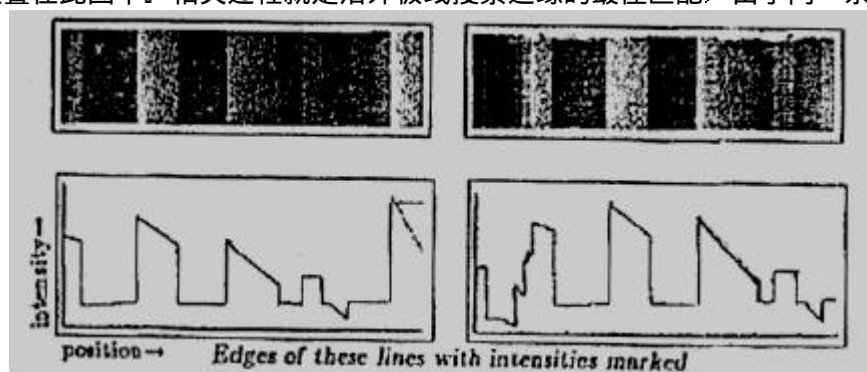


图 5.9 外极线上的边缘以及灰度分布

线上的边缘数一般可高达 30 条，这时即使应用启发式的搜索方法计算量也太大，为此采用被称为 Viterbi 算法的动态规划方法<sup>[For 73]</sup>。Viterbi 算法与一般搜索算法的不同之处是它要求把原始的问题分解成两个子问题。每个子问题可求出最优解，然后可对所得结果继续处理以求得原始问题的全局最优解。每个子问题又可以递归地被分解和求解。

为使相关过程可靠和高效的进行，采用了从粗到细的方法。先在降低分辨率把细节删除的情况下确定大致的视差，并利用这些近似的视差来导引较高分辨率时的相关计算。在作全分辨率（分辨率未降低）的相关时，应用了边缘方位，两边的灰度，相对视差（低分辨率时所得的视差）和间隔对应等信息。在较低分辨时的相关计算中使用的特征是两边的灰度，对比度和间隔对应。所有这些参数最后都被计算为概率加权系数  $P$ ,  $0 \leq P \leq 1$ 。由于上述的相关计算中利用了一条线上的信息，所以求得的结果难免还有误差，这表现为存在明显的水平锯齿状交叉。通常可以假设物体表面和表面轮廓是连续的，因此在一幅图中连接的边缘在另一幅图中也是连接的边缘，这被称为连接性约束。利用这个约束条件可去除不正确的匹配，使匹配结果比较整齐。

#### (2) 基于灰度的相关

经过上述基于边缘的相关计算以后确定了两幅图中边缘的对应关系。从而确定了边缘之间外极线间隔的对应关系。间隔中如有未被匹配的边缘则可与另一幅中相应间隔中未被匹配的边缘重新匹配。这样就使更多的边缘得到匹配。最后可利用两个相邻边缘之间的灰度分布信息进行相关计算从而求得整幅图的视差。

## 5.1.4 摄像机的标定 (calibration)

摄像机标定的目的是确定摄像机的图象坐标系与物理空间中的三维参考坐标系之间的对应关系。只有当摄像机被恰当地标定以后，才能根据图象平面中的二维坐标推论对应物体在物理空间中的实际位置，或反之。为了确定上述对应关系需要知道摄像机的光学和几何参数（内部参数）以及摄像机相对外部参考坐标系的位置和方向（外部参数）。摄像机标定过程就是根据一组已知其参考坐标系坐标和图象坐标系坐标的控制点来确定摄像机的内部和外部参数。有时若已知内部参数，那么只需确定其外部参数。

在立体视觉系统中为根据成象点的坐标确定目标点的空间位置必须对摄象机进行标定，并且进行成象点向三维空间的投影和求取交点。在这一节中我们也将讨论这个问题。

## 1. 摄象机模型<sup>[Tho 81]</sup>

先要建立摄象机的结构模型以便描述透视变换。为避免涉及复杂的透镜，我们先把描述局限于最简单形式的摄象机，即针孔摄象机（图 5.10）。在针孔摄象机中，三维空间中一个点的图象是从这点发出的并经过针孔的光线与摄象机后平面的交点。这样就在后平面的底片上产生了图象。对理想针孔来说，从点发出的光线是理想的直线。所以，空间点的图象本身是一个点，这就消除了聚焦的必要性。当然针孔是摄象机的焦心（focal center），这点被作为投影点。虽然在此我们作了针孔摄象机的假设，但这个模型对大多数通用透镜来说也相当适合。图 5.11 中所示为具有针孔经透镜的摄象机的光路图。这时空间点 P 被透镜聚焦在图象平面中成象。由于从 P 点发出并落在透镜上的光线被聚焦在图象上，所以得到亮得多的图象。对固定焦距摄象机来说，摄象机模型是固定的。焦距相应于透镜的聚焦长度。在可变焦距摄象机中，可把透镜移得离图象平面远些，以便对离摄象机近的空间点聚焦。这样就移动了焦心，并改变了焦距。

焦心的位置可用在参考坐标系中量测的矢量  $C$  来描述（图 5.11）。下一件事是规定图象

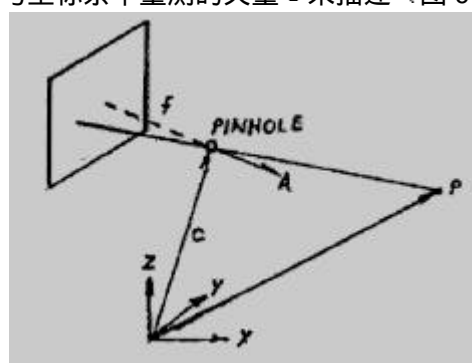


图 5.10 针孔摄象机

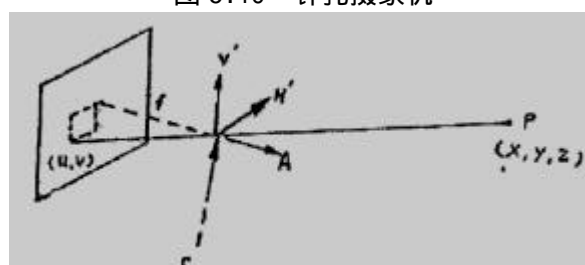


图 5.11 针孔摄象机光路图

平面相对于焦心的位置，图象平面的方向可用垂直于此平面的方向来描述。在图 5.12 中把这个方向表示为摄象机的对准矢量  $A$ ，矢量  $A$  与图象平面垂直，并通过焦心。焦距  $f$  测量沿矢量  $A$  从焦心到图象平面的距离。为确定图象平面上点的位置需要规定两个垂直于矢量  $A$  的坐标轴矢量  $H$  和  $V$ 。 $H$  和  $V$  分别表示图象的水平和垂直方向。对照相机来说，这两个方向可以相当任意。但对电视或 CCD 摄象机来说，矢量的方向是由图象被扫描和数字化的方式来确定的。

利用  $A, H, V$  这三个矢量定义了一个原点在焦心  $C$  的坐标系，这就完成了针孔摄象机的模型。用这个坐标系可以用下述表达式准确地确定图象平面中任意点  $(u, \mathbf{u})$  在参考坐标系中的位置为：

$$C - fA + uH' + \mathbf{u}V \quad (5-8)$$

其中  $u$  和  $\mathbf{u}$  是比例系数，也就是图象平面中点的坐标。当然，在实际的摄象机中图象平面的敏感元件在  $u\mathbf{u}$  空间中是有一定范围的，这就限制了摄象机的视场。以下我们用这个模型来

确定空间中任意点 P 与这个点在图象平面中成象点之间的几何关系。

## 2. 摄象机的透视变换

我们从研究摄象机几何关系的侧视图开始，先只研究投影的垂直分量。在图 5-12 中摄象机的对准矢量  $A$  和图象平面的垂直矢量  $V$  位于书本的平面上。首先作焦心  $C$  到目标点  $P$  的位移矢量  $D$ ，并用求矢量  $D$  与  $A$  或  $V$  的矢量点积的方法把矢量  $D$  分别投影到这两个矢量上去。根据相似三角形可求得：

$$v = f \frac{D \cdot V'}{D \cdot A} \quad (5-9)$$

其中  $D = P - C$

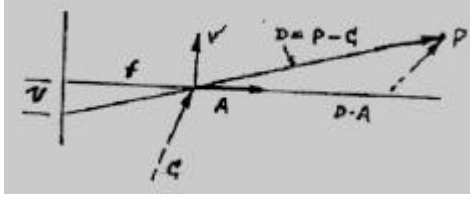


图 5-12 成象过程的侧视图

类似地还可得到

$$u = f \frac{D \cdot H'}{D \cdot A} \quad (5-10)$$

图象平面和焦心在三维空间中的位置是由  $C, A, H', V'$ ，这几个参数决定的，如已知这些参数，则按上述两个方程式就可以根据目标点的三维坐标确定它在图象平面上的象。而这些参数则要通过标定过程来确定。

要注意的是虽然我们假设  $A, H'$  和  $V'$  是单位矢量以便于推导。但事实上只要这些矢量的长度相同，那么实际的推导与这些矢量各自的长度无关。这个事实在讨论标定时有重要意义。

## 3. 数字图象的描述

这里我们要讨论的是成象表面的点与数字化后所得图象的行、列之间的关系。对固定摄象机来说，行、列与成象表面之间的对应关系是固定的，它由光敏感器件的单元所确定。而对电视摄象机来说，列数和行数分别由象素数时钟和扫描行数时钟的计数求得。实际图象坐标与象素地址之间的对应关系可表示为

$$u = n(j - j_0) \quad \text{和} \quad u = m(i - i_0) \quad (5-11)$$

其中  $n$  和  $m$  分别是图象水平和垂直方向上象素之间的距离长度， $i_0$  和  $j_0$  是图象平面坐标系原点的象素地址。通常希望  $n$  和  $m$  相同，这将使象素形成正方形，而不是矩形，从而使在图象平面中旋转时是各向同性的。

把(5-11)代入(5-9)中可得

$$i = \frac{f}{m} \cdot \frac{D \cdot V'}{D \cdot A} + i_0 \quad (5-12)$$

上述公式可改写为

$$i = \frac{D \cdot V}{D \cdot A} \quad (5-13)$$

其中

$$V = \frac{f}{m} V' + i_0 A \quad (5-14)$$

通过重新定义  $V$  我们可以在考虑图象数字化的原点和分辨率的情况下仍能用简单的公式表示投影方程。类似地可求得

$$j = \frac{f}{m} \cdot \frac{D \cdot H'}{D \cdot A} + j_0 \quad (5-15)$$

类似地可改写成

$$j = \frac{D \cdot H}{D \cdot A} \quad (5-16)$$

其中 
$$H = \frac{f}{n} H' + j_0 A$$

需要说明的是在电视摄象机中水平扫描线略有倾斜。所以，图象的水平轴并不严格地垂直，这将造成一定程度的畸变。

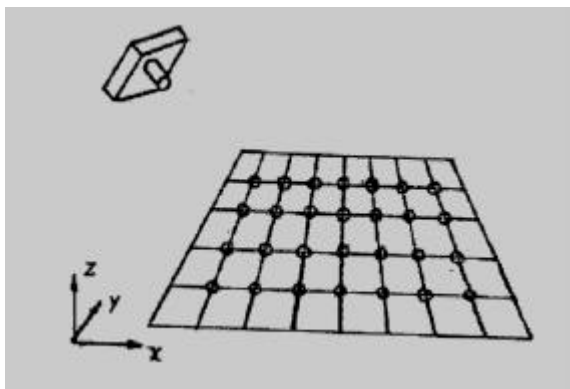


图 5.13 摄象机的标定

#### 4. 摄象机的标定

前面所用到  $C, A, H, V$  四个矢量可确定摄象机的位置和方向（外部参数）。但这些参数难以用机械的方法直接准确地测量。这里将介绍一种根据一组已知其空间位置以及对应图象坐标的目标点计算这些参数的方法。摄象机的内部参数， $f, i_0, j_0, m$  和  $n$  也可用标定的方法求得<sup>[Tai 87]</sup>，并可考虑到透镜造成的畸变。但限于篇幅，这里只介绍标定摄象机外参数的方法。

如图 5.13 所示，为标定摄象机在参考坐标系中的方位和位置，我们用格状图形来产生目标点。格状图形上的交点可被准确地定位，所以被选为目标点的标记。标定的第一件事是识别和定位这些交点，并记录每个标记  $K$  的图象坐标  $(i_k, j_k)$  相应的三维空间中的位置  $P_k$ 。对每个参考点可有 5 个数据，并由式(5-13)和(5-14)产生两个方程式。如<sup>[Qian 89]</sup>所证明的那样，理论上只需要 4 个参考点就足以求解全部未知矢量的问题。但在实际应用中为提高标定的准确性一般希望参考点的数量大于 20。并且在选择参数点的位置时要占满实际应用中的整个视场，所以网状图形应该放在不同的高度上。

标定过程是对摄象机模型的参数与所观察到的标定数据进行最小二乘法拟合。我们希望能找到最好地满足以下关系的标定矢量：

$$i_k = \frac{D_k \cdot V}{D_k \cdot A} \quad (5-17)$$

和 
$$j_k = \frac{D_k \cdot H}{D_k \cdot A} \quad (5-18)$$

其中  $D_k = P_k - C, \quad k = 1, 2, \dots, L$ 。

如果所研究的系统是理想的，那么  $2L$  个方程中的每一个都可准确地被满足，但实际上由于四舍五入和数字化的误差、摄象机的非线性等原因使得难免存在误差。对式(5-17)中的  $i_k$  来说误差的平方和  $S$  等于：

$$S = \sum_{k=1}^L (D_k \cdot V - i_k D_k \cdot A)^2 \quad (5-19)$$

我们希望通过选择合适的标定向量使平方误差的和最小。这个解可通过应用标准的线性代数方法求得。(5-17)式可改写成

$$D_k \cdot V - i_k D_k \cdot A = 0 \quad (5-20)$$

因为  $D_k = P_k - C$ ，所以上式可写成

$$P_k \cdot V - i_k P_k \cdot A - C_v + i_k C_A = 0 \quad (5-21)$$

其中  $C_v = C \cdot V$ ,  $C_A = C \cdot A$

因为  $P_k$ ,  $A$  和  $V$  是三维向量, 因此可写成

$$\begin{aligned} P_k &= [p_{k1}, p_{k2}, p_{k3}], \quad C = [c_1, c_2, c_3] \\ A &= [a_1, a_2, a_3], \quad V = [v_1, v_2, v_3] \end{aligned} \quad (5-22)$$

把(5-22)式代入(5-21)式得

$$\begin{aligned} p_{k1}v_1 + p_{k2}v_2 + p_{k3}v_3 - i_k p_{k1}a_1 - i_k p_{k2}a_2 \\ - i_k p_{k3}a_3 - C_v + i_k C_A = 0, \quad k=1, 2, \dots, L \end{aligned} \quad (5-23)$$

方程(5-23)中包含 8 个未知数即,  $a_1, a_2, a_3, v_1, v_2, v_3, C_v, C_A$ 。为简化求解过程可设  $A$  矢量的任意一个坐标分量为 1。例如,  $a_1 = 1$ 。这样未知数可降为 7 个。如果需要, 以后可用比例系数来调整解的值, 使  $A \cdot A = 1$ 。以满足原来所作  $A$  是单位矢量的假设, 若  $a_1 = 1$ , 则令  $i_k p_{k1} = b_k$ ,  $i_k$  和  $p_{k1}$  已知, 所以  $b_k$  是常数。这时(5-23)式可被重写成以下矩阵形式

$$QW - B = 0 \quad (5-24)$$

其中  $Q$  是  $L \times 7$  的矩阵,  $W$  是  $7 \times 1$  的矢量,  $B$  是  $L$  维常数矢量。

$$\begin{aligned} W &= [v_1, v_2, v_3, a_2, a_3, C_v, C_A]^T \\ Q &= \begin{bmatrix} p_{11} & p_{12} & p_{13} & -i_1 p_{12} & -i_1 p_{13} & -1 & i_1 \\ p_{21} & p_{22} & p_{23} & -i_2 p_{22} & -i_2 p_{23} & -1 & i_1 \\ & & & \dots & & & \\ p_{L1} & p_{L2} & p_{L3} & -i_L p_{L2} & -i_L p_{L3} & -1 & i_L \end{bmatrix} \end{aligned}$$

$$B = [i_1 p_{11}, i_2 p_{21}, i_3 p_{31}, \dots, i_L p_{L1}]^T$$

方程(5-24)的最小二乘法解就是寻找使

$$|QW - B|^2$$

最小的  $u_1^0, u_2^0, u_3^0, a_2^0, a_3^0, C_v^0, C_A^0$ , 根据线性代数原理这相当于求解下面的代数方程

$$Q^T QW = Q^T B \quad (5-25)$$

方程(5-25)可用任何线性方程解法求解。

到此为止我们已求得  $A, V, C_A$  和  $C_v$ , 利用(5-18)式可对  $j_k$  建立与(5-23)式相似的方程, 即

$$\begin{aligned} p_{k1}h_{1+} + p_{k2}h_2 + p_{k3}h_3 - C_h - j_k p_{k1}a_{1-} - j_k p_{k2}a_2 \\ - j_k p_{k3}a_3 + j_k C_A = 0 \end{aligned} \quad (5-26)$$

由于已求得  $A$  和  $C_A$ , 所以(5-26)式中只有 4 个未知数, 即三维矢量  $H$  和  $C_h = C \cdot H$ , 然后应用上述的最小二乘法可求得  $H$  和  $C_h$ 。

在求得  $C_A, C_h$  和  $C_v$  以后, 通过求解下述线性方程组即可求得摄像机的空间位置  $C$ 。

$$\begin{aligned} C \cdot A &= C_A \\ C \cdot H &= C_h \\ C \cdot V &= C_v \end{aligned}$$

这样就求得了完整的摄象机模型。如果程序在求解时失败(上溢、下溢、迭代的发散), 则可能是标定过程不够精确或原始数据不充分。另一种可能的问题是如果令  $a_1 = 1$ , 这个任意的选择与你选择的参考坐标系有矛盾, 即在你固定的摄象机位置下  $a_1 = 0$ , 这时可移动坐标系的方位进行补救。

## 5. 确定物体在三维空间中的位置

这一段要讨论的问题是根据物体点在立体图象平面中坐标确定物体点在三维空间中的  $P$  位置。设三维空间的  $P$  点在左、右两个摄象机中成象点的坐标分别为  $(i_1, j_1)$  和  $(i_2, j_2)$ 。(图 5-14)。从原理上讲,  $P$  点应位于直线  $L_1$  和  $L_2$  的交点处。  $L_1$  和  $L_2$  是由这两个成象点向三维



空间投影产生的。因此，我们需要先求得  $L_1$  和  $L_2$  的表达式。 $(i_1, j_1)$  能满足方程(5-17)和(5-18)。这里把这两个方程重新写出，即

$$i_1 = \frac{D \cdot V_1}{D \cdot A_1} = \frac{(P - C_1) \cdot V_1}{(P - C_1) \cdot A_1} \quad (5-27)$$

$$j_1 = \frac{D \cdot H_1}{D \cdot A_1} = \frac{(P - C_1) \cdot H_1}{(P - C_1) \cdot A_1} \quad (5-28)$$

其中  $V_1 = \frac{f}{m} V'_1 + i_0 A_1$ ,  $H_1 = \frac{f}{n} H'_1 + j_0 A_1$ ,

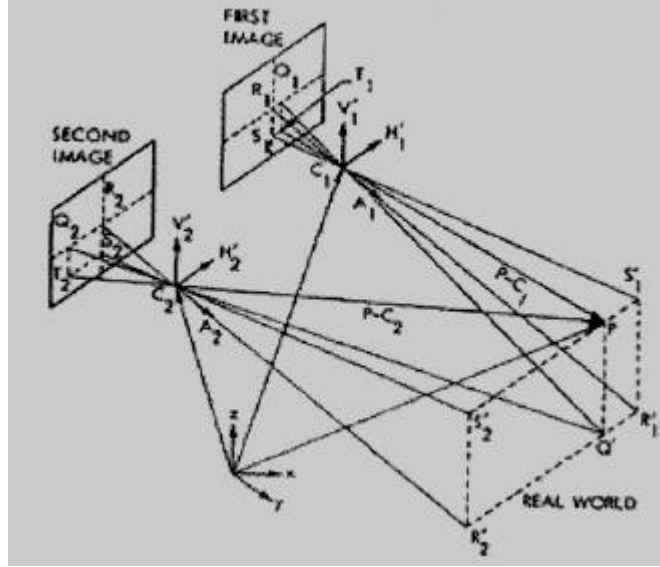


图 5.14 根据成像点求目标点的空间位置

由(5-27)式得

$$(P - C_1) \cdot V_1 - (P - C_1) \cdot i_1 A_1 = 0$$

所以  $(P - C_1) \cdot (V_1 - i_1 A_1) = 0$  (5-29)

同理由(5-28)式得

$$(P - C_1) \cdot (H_1 - j_1 A_1) = 0 \quad (5-30)$$

上两式说明  $(P - C_1)$  与  $(V_1 - i_1 A_1)$  和  $(H_1 - j_1 A_1)$  这两个矢量垂直，所以

$$(P - C_1) // (V_1 - i_1 A_1) \times (H_1 - j_1 A_1) \quad (5-31)$$

其中  $//$  表示平行， $\times$  表示矢量叉积符号，由上式可得

$$P = C_1 + \mathbf{I} R_1(i_1, j_1), \mathbf{I} > 0 \quad (5-32)$$

其中， $R_1(i_1, j_1) // (V_1 - i_1 A_1) \times (H_1 - j_1 A_1)$  同时  $R_1(i_1, j_1)$  是单位矢量。

换言之，若已知空间点在左摄像机中的成像  $(i_1, j_1)$ ，则可推知  $P$  必定位于空间直线  $L_1$  上。

$$L_1 = C_1 + \mathbf{I} R_1(i_1, j_1), \quad 0 < \mathbf{I} < +\infty \quad (5-33)$$

其中  $R_1(i_1, j_1) // (V_1 - i_1 A_1) \times (H_1 - j_1 A_1)$

并且  $R_1(i_1, j_1)$  归一化为 1

同理，由  $P$  在右摄像机中的成像  $(i_2, j_2)$  可确定  $P$  位于空间直线  $L_2$  上。

$$L_2 = C_2 + \mathbf{I} R_2(i_2, j_2), \quad 0 < \mathbf{I} < +\infty$$

由于不可避免地存在误差，空间直线  $L_1$  和  $L_2$  不一定能相交。因此可根据到  $L_1$  和  $L_2$  这两条直线的距离之和最小的原则算出  $P$  点的位置，也即寻找满足下式的空间点：

$$\min_{x \in R^3} \left\{ \|X - C_1\|^2 - [(X - C_1) \cdot R_1(i_1, j_1)]^2 + \|X - C_2\|^2 - [(X - C_2) \cdot R_2(i_2, j_2)]^2 \right\} \quad (5-34)$$

$$= \|P - C_2\|^2 - [(P - C_1) \cdot R_1(i_1, j_1)]^2 + \|P - C_2\|^2 - [(X - C_1) \cdot R_2(i_1, j_1)]^2$$

这是因为

$$\|X - C\| - [(X - C) \cdot R]^2$$

表示空间点  $X$  与直线  $C + \mathbf{I}R$  之间的距离，其中  $\|X - C\|$  表示向量  $X - C$  的模。  $\|R\| = 1$

讨论

以上用矢量代数方法介绍了摄像机标定问题。在处理摄像机标定问题时也可用坐标变换的方法。如图 5.15 所示，摄像机坐标可看成是由参考坐标系经过坐标变换得到。这个变换可被唯一地分解成绕参考坐标系坐标轴的旋转和三维平移。即

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix} + T \quad (5-35)$$

其中  $R$  是  $3 \times 3$  旋转矩阵

$$R \equiv \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} \quad (5-36)$$

$$r_1 = \cos \mathbf{y} \cos \mathbf{q}, \quad r_2 = \sin \mathbf{y} \cos \mathbf{q}, \quad r_3 = -\sin \mathbf{q}$$

$$r_4 = -\sin \mathbf{y} \cos \mathbf{f} + \cos \mathbf{y} \sin \mathbf{q} \cos \mathbf{f}, \quad r_5 = \cos \mathbf{y} \cos \mathbf{f} + \sin \mathbf{y} \sin \mathbf{q} \sin \mathbf{f}$$

$$r_6 = \cos \mathbf{q} \sin \mathbf{f},$$

$$r_7 = \sin \mathbf{y} \sin \mathbf{q} + \cos \mathbf{y} \sin \mathbf{q} \cos \mathbf{f}$$

$$r_8 = -\cos \mathbf{y} \sin \mathbf{f} + \sin \mathbf{y} \sin \mathbf{q} \cos \mathbf{f}, \quad r_9 = \cos \mathbf{q} \cos \mathbf{f}$$

$\mathbf{f}$ ,  $\mathbf{q}$  和  $\mathbf{y}$  分别为绕参考坐标系  $x$ ,  $y$ ,  $z$  坐标轴的转角。  $T$  是位移矢量

$$T \equiv [T_x \quad T_y \quad T_z]^T$$

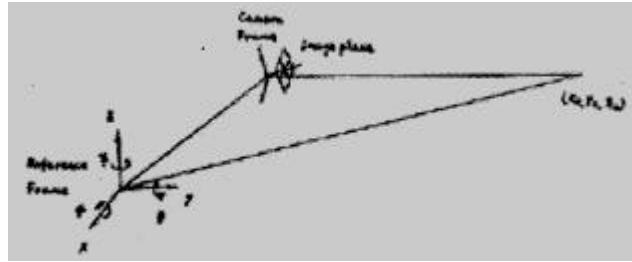


图 5.15 摄像机坐标系相对参考坐标系的坐标变换

空间点  $P$  在摄像机坐标系中经透视投影所得成像点坐标为  $(x_{im}, y_{im})$ ，利用 (5-35) 和 (5-36) 式可得，

$$x_{im} = f \frac{x_c}{y_c} = f \frac{r_1 x + r_2 y + r_3 z + T_x}{r_7 x + r_8 y + r_9 z + T_z} \quad (5-37)$$

$$y_{im} = f \frac{x_c}{y_c} = f \frac{r_4 x + r_5 y + r_6 z + T_y}{r_7 x + r_8 y + r_9 z + T_z}$$

摄像机标定的过程就是根据上述两个公式确定旋转矩阵  $R$  和位移矢量  $T$  的有关参数。

由于摄像机标定是实际应用中的关键技术，所以人们已对它进行了广泛深入的研究。已发表的算法大致可分成以下 5 类<sup>[Tai 87]</sup>，其中包括：(1) 非线性优化技术<sup>[Abe 71][Fai 75][Sob 74]</sup>。这种方法的优点是对任何复杂的成像模型都能达到很高的精度，缺点是为了进行非线性搜索需要有良好的初始猜测，这样就难以实现自动操作，此外由于要进行非线性搜索，需要进行大量

计算。(2)用线性方程求解计算透视变换矩阵<sup>[Yak 78][Hal 82][Gan 84]</sup>。这种方法的优点是避免了复杂的非线性优化计算,但不能处理透镜畸变的问题,同时要求解的未知数比方程实际的自由度数量要多许多。这是由于这些未知数不能线性独立地求解造成的。(3)双平面法<sup>[Isa 85][Mar 81]</sup>。这种方法的优点是只需求解线性方程,缺点是未知数起码有 24 个,比方程自由度多得多。同时图象和物体坐标系之间的变换公式要靠经验来确定。(4)几何学方法<sup>[Fis 80]</sup>。这种方法不需要非线性搜索,但不能处理透镜畸变,需要知道焦距和图象的比例系数。(5)两阶段的标定方法<sup>[Tai 87]<sup>×2</sup></sup>。这种方法应用径向对准约束把标定参数分成外部参数和内部参数两组。第一组参数用线性方程组求解来确定,第二组参数用迭代求解透视投影方程来确定。这种方法能处理透镜畸变问题,达到较高的精度,并且计算简单。

#### 参考文献:

- [Abe 71] Abel-Aziz and Karara, H. M., Direct linear transformation into object space coordinates in close-range photogrammetry, in Proc. Symp. Close-range photogrammetry, Univ. Of Illinois at Urbana-Champaign, 1971,1-18
- [Bak 80] Baker, H. H., Edge-based stereo correlation, Proc. ARPA Image Understanding workshop, Univ. Of Maryland, 168-175,1980
- [Bak 81] Baker, H. H. And Binford, T. O., Depth from edge and intensity based stereo, Proc. of IJCAI-81, Vancouver, British Columbia,1981,631-63
- [Bin 83] Binford, T. O., Stereo vision: complexity and constraints, in Robotics research, Ed. By R. Brady and R. Paul,1983
- [Fai 75] Faig, W., Calibration of close-range photogrammetry systems: mathematical formulation, Photogrammetry Eng. Remote Sensing, Vol. 41, 1479-1486,1975
- [Fis 80] Fishler, M. and Bolles, R., Random sample consensus: A paradigm for model fitting application to image analysis and automated cartography, in Proc. Image Understanding Workshop, Apr. 1980,71-88
- [For 73] Forney, G. And David, Jr., The viterbi algorithm, Proc. IEEE, vol. 61, no. 3, March 1973
- [Gan 80] Ganapathy, S., Decomposition of transformation matrices for robot vision, in Proc. Int. Conf. Robotics and Automation, Apr. 1980,71-88
- [Gen 80] Gennery, D.B., Modeling the environment of an exploring vehicle by means of stereo vision, Ph.D. thesis, Stanford AI Lab., 1980
- [Gri 81] Grimson, W. E. L., A computer implementation of a theory pf human stereo vision, Philosophical Transaction Royal Society London, B. 292, 1981,217-253
- [Isa 85] Isaguirre, Pu, P. And Summers, J., A new development in camera calibration: calibrating a pair of mobile cameras, in Proc. Int. Conf. Robotics and Automation, 1985,74-79
- [Kas 83] Kass, M., A computational framework for the visual correspondence problem, Proc. ARPA Image Understanding workshop. Washington, D. C., 1983
- [Len 87] Lenz, R. K. And Tsai, R. Y., Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology, Proc. of 1987 IEEE Int. Conf. On Robotics and Automation, 1987,68-75
- [Mar 79] Marr, D. and Poggio, T., A computational theory of human stereo vision, Proc. R. Soc. Lond B 204: 301-328,1979
- [Mar 80] Marr, D. And Hildreth, E., Theory of edege detection, Proc. R. Soc. Lond. B 207:187-217,1980
- [Mar81] Martins, H. A., Birk, J. R. And Kelly, R. B., Camera models based on data from two

- calibration plans, Computer Graphics Image Processing, Vol. 17, 173-180,1981
- [Mor 80] Moravec, H. P., Obstacle avoidance navigation in the real world by a seeing robot rover, Ph.D. thesis, Stanford AI Lab. 1980
- [Nis 83] Nishihara, H. K. and Poggio, T., Stereo vision for robotics, in Robotics Research, Ed. By R. Brady and R. Paul,1983
- [Sob 74] Sobel, I., On calibrating computer controlled camera for perceiving 3D scenes, Artificial Intelligence, Vol. 5, 185-188,1974
- [Tai 87] Tsai, R. T., A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-shelf TV camera and lens, IEEE Journal of Robotics and Automation, Vol. RA-3, no.4, Aug. 1987
- [Tho 81] Thompson, A. M., Camera geometry, Robotics Age, Vol. 20, Mar/Apr., 1981, 20-27
- [Yak 78] Yalimovsky, Y. And Cuanigham, R., A system for extracting three-dimensional measurements from a pair of TV camera, Computer Graphics & Image Processing, Vol. 7, 195-210,1978