

# 第十一章 深度图

获取场景中各点相对于摄像机的距离是计算机视觉系统的重要任务之一。场景中各点相对于摄像机的距离可以用深度图(Depth Map)来表示,即深度图中的每一个像素值表示场景中某一点与摄像机之间的距离。机器视觉系统获取场景深度图技术可分为被动测距传感和主动深度传感两大类。被动测距传感是指视觉系统接收来自场景发射或反射的光能量,形成有关场景光能量分布函数,即灰度图像,然后在这些图像的基础上恢复场景的深度信息。最一般的方法是使用两个相隔一定距离的摄像机同时获取场景图像来生成深度图。与此方法相类似的另一种方法是一个摄像机在不同空间位置上获取两幅或两幅以上图像,通过多幅图像的灰度信息和成像几何来生成深度图。深度信息还可以使用灰度图像的明暗特征、纹理特征、运动特征间接地估算。主动测距传感是指视觉系统首先向场景发射能量,然后接收场景对所发射能量的反射能量。主动测距传感系统也称为测距成像系统(Rangefinder)。雷达测距系统和三角测距系统是两种最常用的两种主动测距传感系统。因此,主动测距传感和被动测距传感的主要区别在于视觉系统是否是通过增收自身发射的能量来测距。另外,我们还接触过两个概念:主动视觉和被动视觉。主动视觉是一种理论框架,与主动测距传感完全是两回事。主动视觉主要是研究通过主动地控制摄像机位置、方向、焦距、缩放、光圈、聚散度等参数,或广义地说,通过视觉和行为的结合来获得稳定的、实时的感知。我们将在最后一节介绍主动视觉。

## 11.1 立体成象

最基本的双目立体几何关系如图 11.1(a)所示,它是由两个完全相同的摄像机构成,两个图像平面位于一个平面上,两个摄像机的坐标轴相互平行,且  $x$  轴重合,摄像机之间在  $x$  方向上的间距为基线距离  $b$ 。在这个模型中,场景中同一个特征点在两个摄像机图像平面上的成像位置是不同的。我们将场景中同一点在两个不同图像中的投影点称为共轭对,其中的一个投影点是另一个投影点的对应(correspondence),求共轭对就是求解对应性问题。两幅图像重叠时的共轭对点的位置之差(共轭对点之间的距离)称为视差(disparity),通过两个摄像机中心并且通过场景特征点的平面称为外极(epipolar)平面,外极平面与图像平面的交线称为外极线。

在图 11.1 中,场景点  $P$  在左、右图像平面中的投影点分为  $p_l$  和  $p_r$ 。不失一般性,假设坐标系原点与左透镜中心重合。比较相似三角形  $PMC_l$  和  $p_l LC_l$ , 可得到下式:

$$\frac{x}{z} = \frac{x'_l}{F} \quad (11.1)$$

同理,从相似三角形  $PNC_r$  和  $p_r RC_r$ , 可得到下式:

$$\frac{x - B}{z} = \frac{x'_r}{F} \quad (11.2)$$

合并以上两式,可得:

$$z = \frac{BF}{x'_l - x'_r} \quad (11.3)$$

其中  $F$  是焦距,  $B$  是基线距离。

因此,各种场景点的深度恢复可以通过计算视差来实现。注意,由于数字图像的离散特性,视差值是一个整数。在实际中,可以使用一些特殊算法使视差计算精度达到子像素级。因

此，对于一组给定的摄像机参数，提高场景点深度计算精度的有效途径是增长基线距离 $b$ ，即增大场景点对应的视差。然而这种大角度立体方法也带来了一些问题，主要的问题有：

1. 随着基线距离的增加，两个摄像机的共同的可视范围减小
2. 场景点对应的视差值增大，则搜索对应点的范围增大，出现多义性的机会就增大。
3. 由于透视投影引起的变形导致两个摄像机获取的两幅图像中不完全相同，这就给确定共轭对带来了困难。

在图 11. 1(b)中，图像中的每个特征点都位于第二幅图像中的同一行中。在实际中，两条外极线不一定完全在一条直线上，即垂直视差不为零。但为了简单起见，双目立体算法中的许多算法都假设垂直视差为零。

在实际应用中经常遇到的情况是两个摄像机的光轴不平行，比如，在某些系统中，调节两个摄像机的位置和姿态，使得它们的光轴在空间中相交在某一点，如图 11. 2 所示。在这种情况下，视差与光轴交角有关。对于任意一个光轴交角，在空间中总存在一个视差为零的表面。比这一表面远的物体，其视差大于零；反之，比这一表面近的表面，其视差小于零。因此，在一个空间区域中，其视差可被划分为三组：

$$+ \quad d > 0$$

$$- \quad d < 0$$

$$0 \quad d = 0$$

这三组视差可用于解决匹配不确定问题。

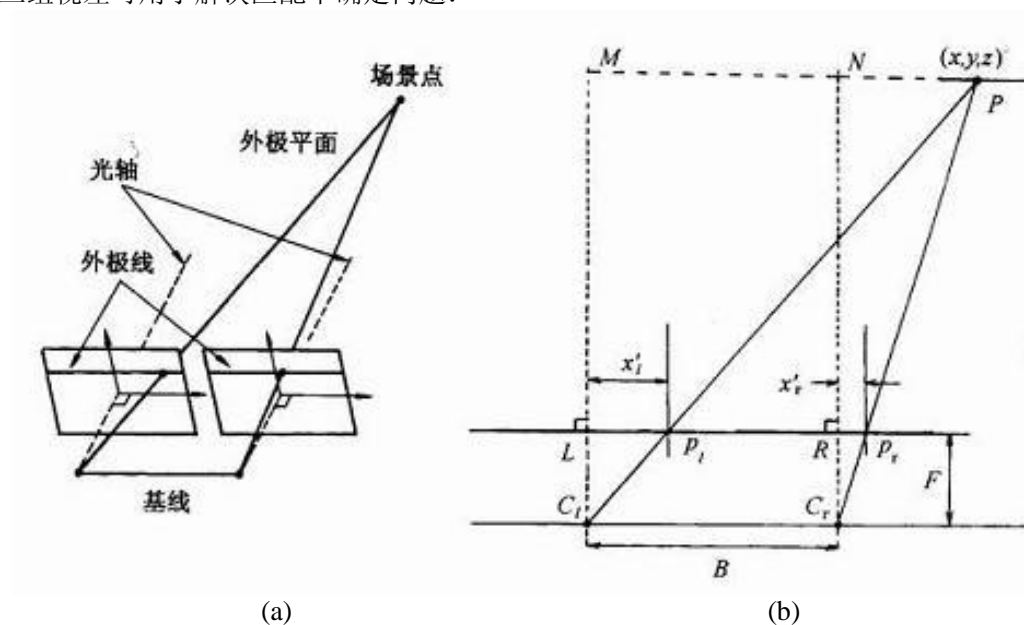


图 11. 1 双目立体视觉几何模型

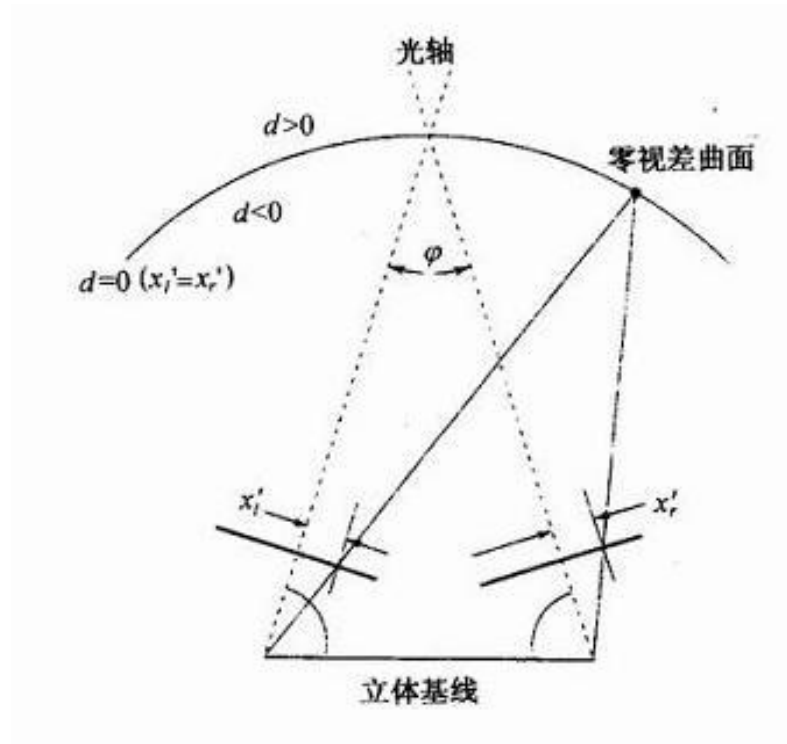


图 11. 2 摄像机光轴交会空间一点. 摄像机之间的夹角定义了视差为零的一个空间表面.

摄像机光轴不平行的另一种系统是后面将要介绍的一种叫做会聚式(convergent)立体视觉系统(见图 11. 5). 这种系统不要求光轴严格地相交于空间一点. 立体成象的最一般情况是一个运动摄像机连续获取场景图像, 形成立体图像序列, 或间隔一定距离的两个摄像机同时获取场景图像, 形成立体图像对.

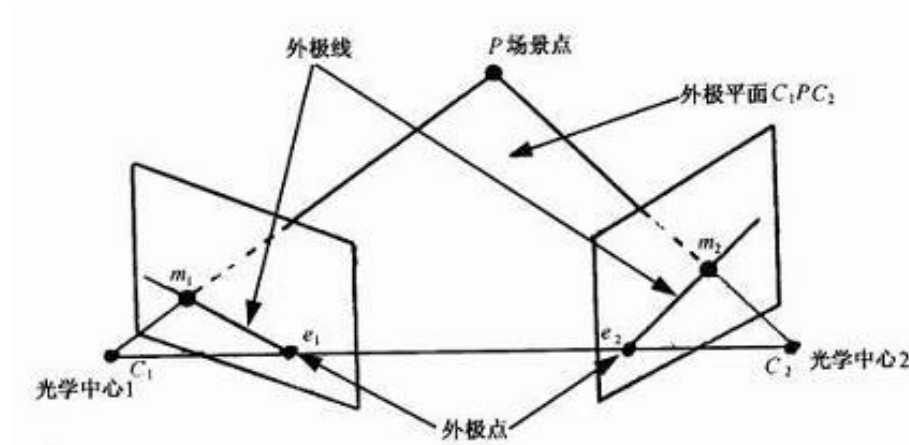


图 11. 3 外极线几何示意图

图 11. 3 所示的是处于任意位置和方位的两个摄像机, 对应于某一场景点的两个图像点位于外极线上. 这两幅图像也可以是一个摄像机由一点运动到另一点获取这两幅图像. 即使两个摄像机处于一般的位置和方位时, 对应场景点的两个图像点仍然位于图像平面和外极平面的交线(外极线)上. 由图不难看出, 外极线没有对应图像的某一行.

## 11. 2 立体匹配

立体成像系统的一个不言而喻的假设是能够找到立体图像对中的共轭对,即求解对应问题.然而,对于实际的立体图像对,求解对应问题极富有挑战性,可以说是立体视觉最困难的一步.为了求解对应,人们已经建立了许多约束来减少对应点搜索范围,并最终确定正确的对应.下面我们将讨论几个最基本的约束,然后讨论边缘特征和区域特征在立体匹配中的应用.

## 11. 2. 1 基本约束

### (1) 外极线约束

对于两幅从不同角度获取的同一场景的图像来说,传统的特征点搜索方法是首先在一幅图像上选择一个特征点,然后在第二幅图像上搜索对应的特征点.显然,这是一个二维搜索问题.根据成像几何原理,一幅图像上的特征点一定位于另一幅图像上对应的外极线上.因此,在外极线上而不是在二维图像平面上求解对应问题是一个一维搜索问题.如果已知目标与摄像机之间的距离在某一区间内,则搜索范围可以限制在外极线上的一个很小区间内,如图 11. 4 所示.所以,利用外极线约束可以大大地缩小寻找对应点的搜索空间,这样即可以提高特征点搜索速度,也可以减少假匹配点的数量(范围越小,包含假匹配点的可能性越小).请注意,由于摄像机位置及其方向的测量误差和不确定性,匹配点可能不会准确地出现在图像平面中对应的外极线上;在这种情况下,有必要在外极线的一个小邻域内进行搜索.

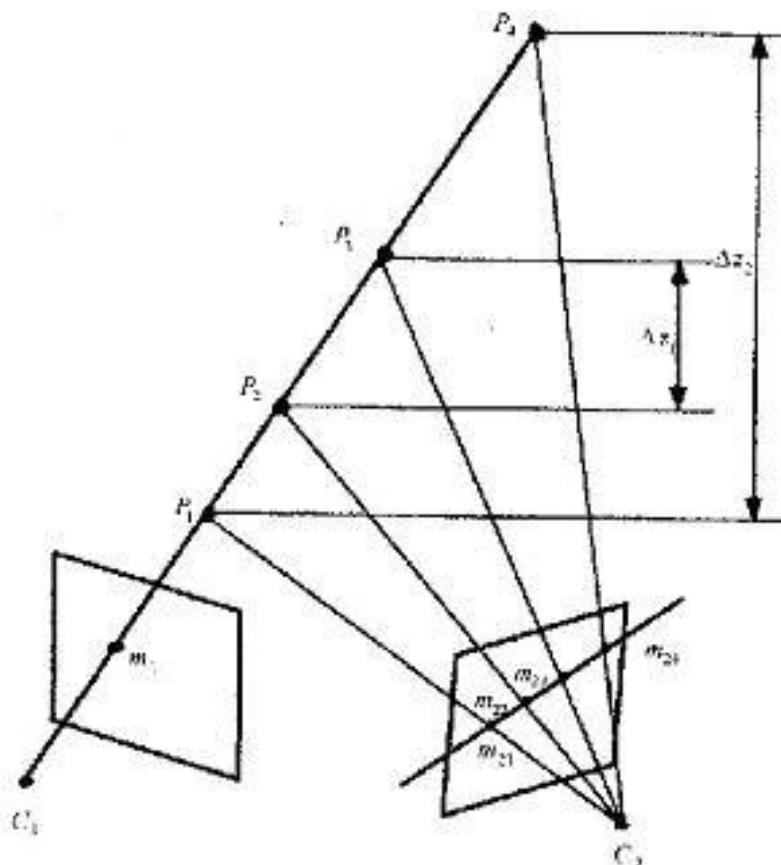


图 11.4 空间某一距离区间内的一条直线段对应外极线上的一个有限区间

### (2) 一致性约束

立体视觉通常由两个或两个以上摄像机组成,各摄像机的特性一般是不同的.这样,场景中对对应点处的光强可能相差太大,直接进行相似性匹配,得到的匹配值变化太大.因此,在进行匹配前,必须对图像进行规范化处理(Normalization).设参考摄像机和其它摄像机的

图像函数分别为  $f_0(i, j)$  和  $f_k(i, j)$ , 在  $m \times n$  图像窗内规范化图像函数为:

$$\bar{f}_0(i, j) = (f_0(i, j) - \mu_0) / \sigma_0 \quad (11.4)$$

$$\bar{f}_k(i, j) = (f_k(i, j) - \mu_k) / \sigma_k \quad (11.5)$$

其中  $\mu$  是图像窗内光强的平均值,  $\sigma$  是光强分布参数:  $\sigma^2 = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m (f(i, j) - \mu)^2$

相似估价函数为差值绝对值之和(sum of absolute difference, SAD):

$$\varepsilon_k = \sum_{i=1}^n \sum_{j=1}^m |\bar{f}_0(i, j) - \bar{f}_k(i, j)| \quad (11.6)$$

### (3) 唯一性约束

一般情况下, 一幅图像(左或右)上的每一个特征点只能与另一幅图像上的唯一一个特征对应。

### (4) 连续性约束

物体表面一般都是光滑的, 因此物体表面上各点在图像上的投影也是连续的, 它们的视差也是连续的。比如, 物体上非常接近的两点, 其视差也十分接近, 因为其深度值不会相差很大。在物体边界处, 连续性约束不能成立, 比如, 在边界处两侧的两个点, 其视差十分接近, 但深度值相差很大。

## 11. 2. 2 边缘匹配

本算法使用的边缘特征是通过 Gaussian 函数的一阶导数获得的。在有噪声的情况下, 使用 Gaussian 梯度来计算边缘更加稳定。立体算法的步骤如下:

1. 用四个不同宽度的 Gaussian 滤波器对立体图像对中的每一幅图像进行滤波, 其中前一次滤波的宽度是下一次滤波器宽度的两倍。这一计算可以反复通过对最小的滤波器进行卷积来有效地实现。
2. 在某一行上计算各边缘的位置。
3. 通过比较边缘的方向和强度粗略地进行边缘匹配。显然, 水平边缘是无法进行匹配的。
4. 通过在精细尺度上进行匹配, 可以得到精细的视差估计。

## 11. 2. 3 区域相关性

尽管边缘特征是图像的基本特征, 而且边缘检测算法也十分成熟。但边缘特征往往对应着物体的边界, 物体的边界深度值可以是 (前景) 物体封闭边缘的深度距离和背景点深度距离之间的任一值。特别是曲面物体, 其封闭边缘是物体的轮廓影象边缘, 在两幅图像平面中观察到的轮廓影象边缘与真实的物体边缘不是对应的。不幸的是, 图像平面的强边缘只能沿着这样的封闭边缘才能检测到, 除非物体有其它的高对比度的非封闭边缘或其它特征。这样, 恢复深度的基本问题之一是识别分布于整幅图像中的更多的特征点, 并作为候选对应点。还有许多用于寻找对应点潜在特征的方法, 其中的一种方法是在立体图像对中识别兴趣点 (interesting point), 而后使用区域相关法来匹配两幅图像中相对应的点。

### (1) 区域中感兴趣特征点的检测

两幅图像中用于匹配的点应尽可能容易地被识别和匹配。显而易见, 一个均匀区域中的点是不适合作为候选匹配点, 所以兴趣算子应在图像中寻找具有很大变化的区域。一般认为图像中应有足够多的用于匹配的分立区域。

在以某一点为中心的窗函数中, 计算其在不同方向上的变化量是这些方向上点的差异性

的最好测度。方向变化量的计算公式如下：

$$\begin{aligned}
 I_1 &= \sum_{(x,y) \in S} [f(x,y) - f(x,y+1)]^2 \\
 I_2 &= \sum_{(x,y) \in S} [f(x,y) - f(x+1,y)]^2 \\
 I_3 &= \sum_{(x,y) \in S} [f(x,y) - f(x+1,y+1)]^2 \\
 I_4 &= \sum_{(x,y) \in S} [f(x,y) - f(x+1,y-1)]^2
 \end{aligned} \tag{11.7}$$

其中  $S$  表示窗函数中的所有像素。典型窗函数尺寸从  $5 \times 5$  到  $11 \times 11$  个像素的范围。因为简单的边缘点在边缘方向上无变化，所以，选择上述方向变量的最小值为中心像素点  $(x_c, y_c)$  的兴趣值，可以消除边缘点。否则，在第一幅图像中的边缘点可能与第二幅图像中沿着同一条边缘的所有像素相匹配，由此使得准确确定视差变得十分困难（特别是当边缘是沿着外极线时更是如此）。于是，有如下公式：

$$I(x_c, y_c) = \min(I_1, I_2, I_3, I_4) \tag{11.8}$$

最后，为了避免将多个相邻点选为同一个特征对应的兴趣点，可以将特征点选在兴趣测度函数具有局部最大值的地方。一个点被认为是一个的“好”的好兴趣点，则对应的局部最大值一定大于原先设定的阈值。

一旦在两幅图像中确定特征后，则可以使用许多不同方法进行特征匹配。一种简单的方法是计算一幅图像以某一特征点为中心的一个小窗函数内的像素与另一幅图像中各个潜在对应特征点为中心的同样的小窗函数的像素之间的相关值。具有最大相关值的特征就是匹配特征。很明显，只有满足外极线约束的点才能是匹配点。考虑到垂直视差的存在，应将外极线邻近的特征点也包括在潜在的匹配特征集中。

考虑两幅图像  $f_1$  和  $f_2$ 。设待匹配的候选特征点对的视差为  $(d_x, d_y)$ ，则以特征点为中心的区域之间相似性测度可由相关系数  $r(d_x, d_y)$  定义为：

$$r(d_x, d_y) = \frac{\sum_{(x,y) \in S} [f_1(x,y) - \bar{f}_1][f_2(x+d_x, y+d_y) - \bar{f}_2]}{\{\sum_{(x,y) \in S} [f_1(x,y) - \bar{f}_1]^2 \sum_{(x,y) \in S} [f_2(x+d_x, y+d_y) - \bar{f}_2]^2\}^{1/2}} \tag{11.9}$$

这里的  $\bar{f}_1$  和  $\bar{f}_2$  是被匹配的两个区域中的所有像素灰度平均值，和式符号是在以特征点为中心的小窗函数中的所有像素上进行的。

在式(11.9)中，在每一个像素上使用阈值化处理后的具有正负符号的梯度幅值，而不是图像灰度值，可以改善相关性的计算精度。这可以通过下列过程来实现：计算未平滑的两幅图像中的每个像素的梯度幅值，然后使用两个阈值，一个大于 0，另一个小于 0，将这些值映射到三个值  $(-1, 0, 1)$ 。这样图像就被转换为波浪起伏阵列，可以产生更灵敏的相关值。如果使用上述方法，则无需在相关性计算公式中包括正则项， $r(d_x, d_y)$  可以简化为相对应的像素值的乘积之和。

在大多数情况中，相互靠近的特征点其对应的场景点深度也是十分靠近的。这种启发式方法来源于迭代式松弛方法，我们将在 14.3 节对这种方法进行介绍。

我们在早些时候已经注意到，基于特征的立体匹配算法产生对应于图像特征点的场景稀疏深度图。在稀疏深度图上进行表面内插或逼近运算，可以重建一个表面，这部分内容将在第 13 章讨论。

立体重建过程的主要难题之一是选择兴趣点。一种典型的兴趣点选择方法是基于灰度值的局部最大变化量。不幸的是，这些点经常出现在拐角处或不满足平滑约束条件的表面不连续处。在一些机器视觉应用中，这个问题是通过使用结构光来解决的。将模式图投影到表面上来产生兴趣点，并可使区域变得光滑(见下一节内容)。

最后要说明的是，从图像像素集合中选择用于求解共轭对的像素子集意味着仅仅能恢复这些像素对应的特征点深度。要想得到其它点的深度值，必须通过使用有关计算方法来估算，如内插值技术。

### 11. 3 多基线立体成象

一幅图像上的每一个特征点只能与另一幅图像上的唯一一个特征对应,通常将这一性质称为特征的唯一性约束. 在实际中,由于大多数特征点不是十分明显,特别是重复纹理的出现,常常会产生对应多义性(ambiguity),即一幅图像上的一个特征点对应另一幅图像的若干个对应点,其中的一个点是真正的对应点,而其它点是假对应点. 消除对应点多义性的一种有效方法是采用多基线立体成象[Okutomi 1993],如图 11. 4 所示.

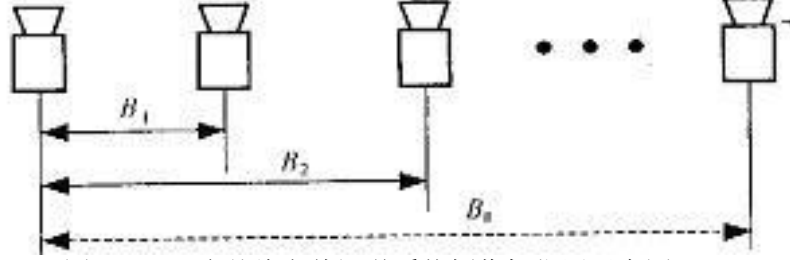


图 11. 5 多基线立体视觉系统摄像机位置示意图

假定  $n$  个摄像机具有相同的焦距  $F$ , 其位置分别为  $P_0, P_1, \dots, P_{n-1}$ , 对应的  $n-1$  个基线用  $B_1, B_2, \dots, B_{n-1}$  表示,  $f_0(x)$  和  $f_i(x)$  表示在位置  $P_0, P_i$  处同步获取的图像,称为一个立体图像对. 已知场景一点  $Z$  的深度值为  $z_r$ , 则根据式子(3),  $f_0(x)$  和  $f_i(x)$  形成的立体视差  $d_{r(i)}$  为:

$$d_{r(i)} = \frac{B_i F}{z_r} \quad (11. 10)$$

图像强度函数  $f_0(x)$  和  $f_i(x)$  在  $Z$  点附近可以表示为:

$$\begin{aligned} f_0(x) &= f(x) + n_0(x) \\ f_i(x) &= f(x - d_{r(i)}) + n_i(x) \end{aligned} \quad (11. 11)$$

其中  $n_0(x), n_i(x)$  服从正态分布  $N(0, \sigma_n^2)$ 。

在一定大小的窗口  $W$  上, 对立体视差变量为  $d_{(i)}$ , 求立体图像对函数差值平方和(sum of squared difference, SSD):

$$e_i(x, d_{(i)}) \equiv \sum_{j \in W} (f_0(x + j) - f_i(x + d_{(i)} + j))^2 \quad (11. 12)$$

设  $\zeta = 1/z$ , 则 (11.10) 改写为:

$$d_{r(i)} = B_i F \zeta_r \quad (11. 13)$$

$$d_{(i)} = B_i F \zeta \quad (11. 14)$$

上式中  $\zeta_r, \zeta$  分别是真实的和候选的场景点深度值的倒数.

用式(11. 13)和(11. 14)替换式(11. 12)得:

$$e_i(x, \zeta) \equiv \sum_{j \in W} (f_0(x + j) - f_i(x + B_i F \zeta + j))^2 \quad (11. 15)$$

将  $e_i(x, \zeta)$  称为匹配评估函数.  $e_i(x, \zeta)$  的期望值为:

$$E[e_i(x, \zeta)] \equiv \sum_{j \in W} (f(x + j) - f(x + B_i F(\zeta - \zeta_r) + j))^2 + 2N_w \sigma_n^2 \quad (11. 16)$$

其中  $N_w$  是窗内像素点的个数.

$n$  个摄像机形成  $n-1$  个 SSD 函数, 将所有的 SSD 函数相加(sum of SSD, SSSD)形成了总的评估函数:

$$e_{1,2,\dots,n-1}(x, \zeta) = \sum_{i=1}^n e_i(x, \zeta) \quad (11.17)$$

其期望值为

$$\begin{aligned} E[e_{1,2,\dots,n-1}(x, \zeta)] &= \sum_{i=1}^{n-1} E[e_i(x, \zeta)] \\ &= \sum_{i=1}^{n-1} \sum_{j \in W} (f(x+j) - f(x + B_i F(\zeta - \zeta_r) + j))^2 + 2nN_w \sigma_n^2 \end{aligned} \quad (11.18)$$

下面讨论将评估函数表示成深度值倒数函数是如何消除多义性的. 设强度函数  $f(x)$  在  $x$  和  $x+a$  处周围具有相同的模式, 即

$$f(x+j) = f(x+a+j) \quad j \in W \quad (11.19)$$

则根据式 (11.16), 有

$$E[e_i(x, \zeta_r)] = E[e_i(x, \zeta_r + \frac{a}{B_i F})] = 2N_w \sigma_n^2 \quad (11.20)$$

由上式可见, 在  $\zeta_r$  和  $\zeta_f = \zeta_r + a/(B_i F)$  处都将得到评估函数的最小值, 因此产生多义性. 但注意一个事实, 假的匹配深度值的倒数  $\zeta_f$  随着基线变化而变化, 而真的匹配深度值的倒数  $\zeta_r$  则与基线长度无关, 这样累加多个基线评估函数, 则会得到唯一的最小值.

比如有两个基线的长度分别为  $B_1, B_2$  ( $B_1 \neq B_2$ ), 则有:

$$\begin{aligned} E[e_{1,2}(x, \zeta)] &= \sum_{j \in W} (f(x+j) - f(x + B_1 F(\zeta - \zeta_r) + j))^2 \\ &\quad + \sum_{j \in W} (f(x+j) - f(x + B_2 F(\zeta - \zeta_r) + j))^2 + 4N_w \sigma_n^2 \end{aligned} \quad (11.21)$$

可以证明:

$$E[e_{1,2}(x, \zeta)] \geq 4N_w \sigma_n^2 = E[e_{1,2}(x, \zeta_r)] \quad (11.22)$$

也就是说,  $e_{1,2}(x, \zeta)$  在  $\zeta_r$  处具有极小值. 图 11.6 是一个验证 SSSD 的简单例子, 其中的原函数为

$$f(x) = \begin{cases} \cos(\frac{\pi}{4}x) + 2 & -4 < x < 12 \\ 1 & \text{其它} \end{cases}$$



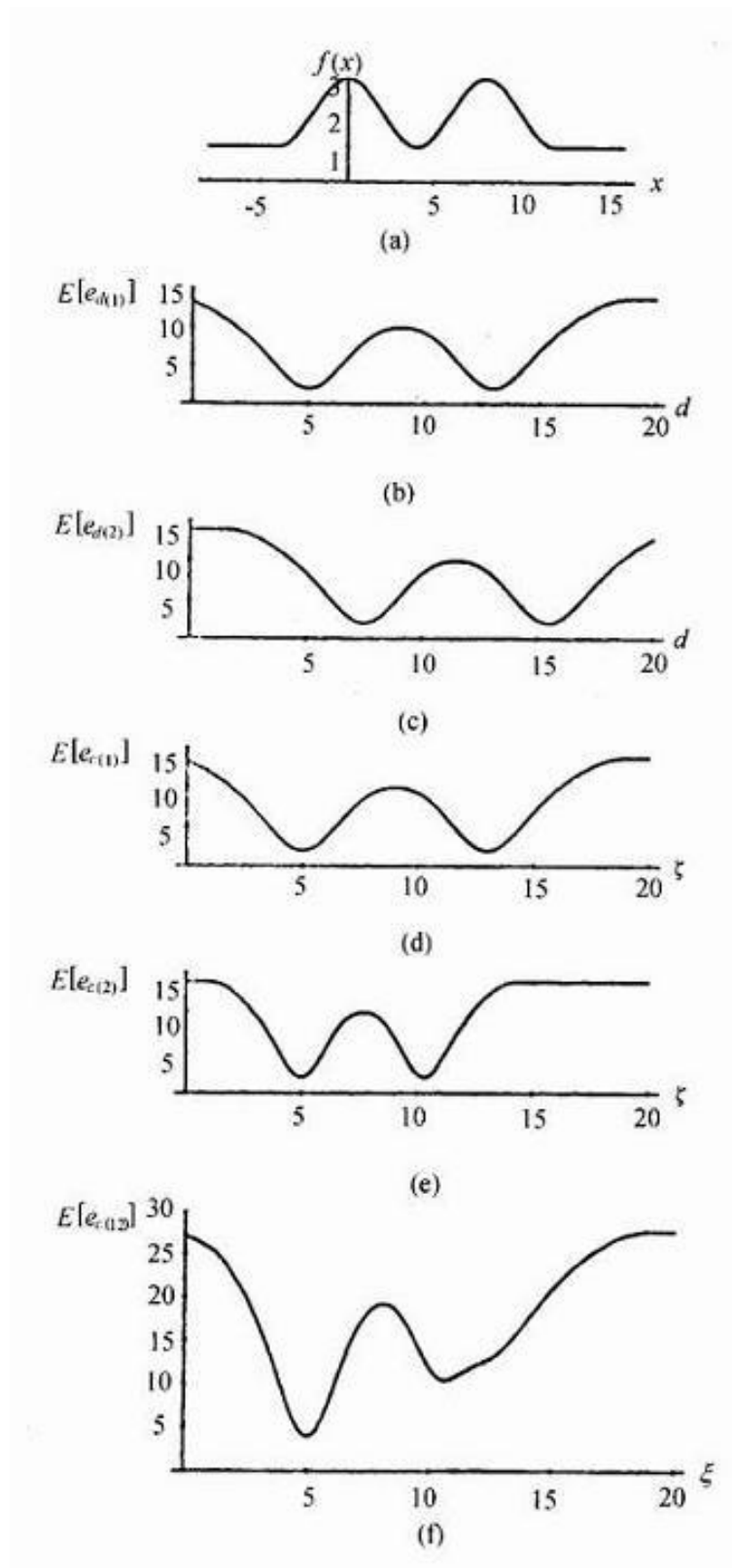


图 11.6 不同评估函数的曲线，

多基线立体视觉可以有效地消除一般场景纹理和重复纹理的对应多义性，但仍然无法求解弱纹理和无纹理物体对应点。图 11.8 是兵马俑模型的立体图像对，除了少部分特征外，其它部位都无法作为特征来求解对应点。为了恢复兵马俑的稠密深度图像，可以采用专用光源向场景投影条纹模式[Kang 1995]，如图 11.8 所示，每一个条纹的强度变化服从正弦分

布. 这样做的效果相当于无纹理的模型上有了条纹特征, 由此实现对应的特征点共轭对的求解.

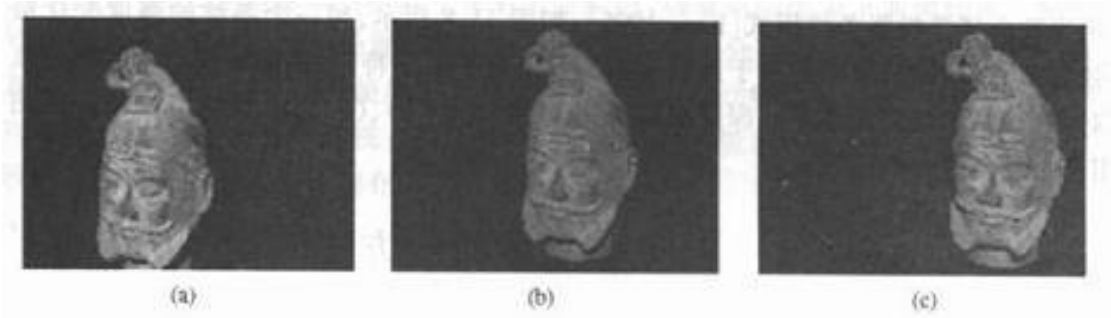


图 11. 7 用三个摄像机获取的兵马俑石膏模型多立体图像对

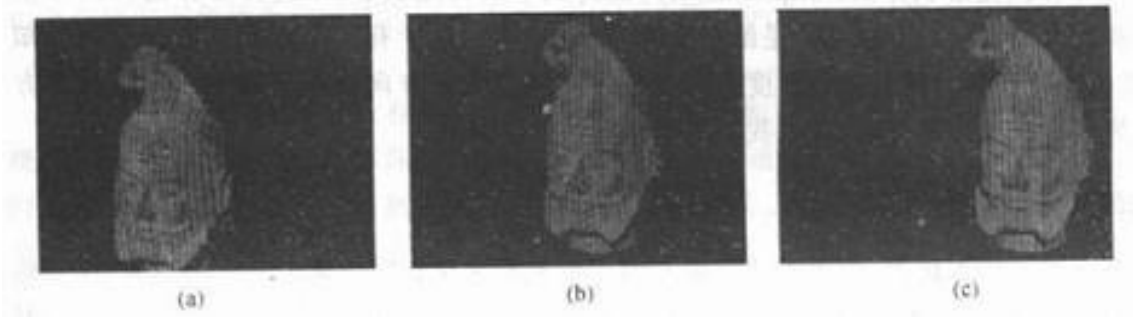


图 11. 8 主动式条纹光源照射下的多立体图像对

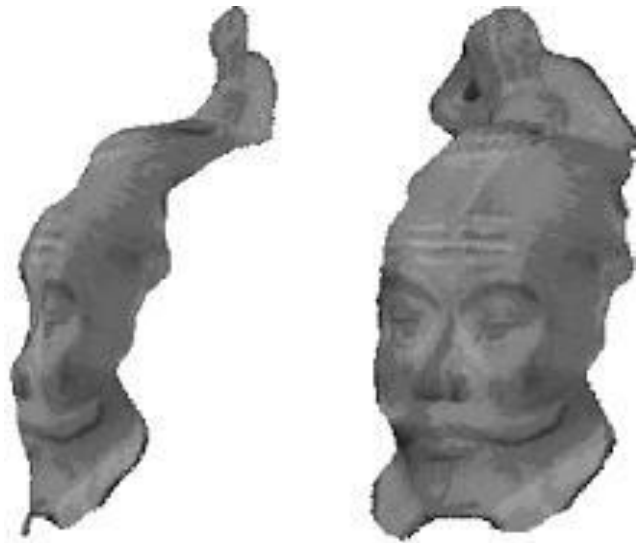


图 11. 9 基于基本约束恢复的秦俑头像的深度图像

### 11. 4 从 X 恢复形状的方法

除了上述讨论的立体图像匹配方法外, 从灰度图像中提取形状信息的许多其它方法, 统称为由 X 恢复形状方法, 也得到了大力地发展. 不过许多方法只用来估计一点的局部表面方位而不是绝对深度. 如果物体上至少有一个点的实际深度是已知的, 那么同一目标上其它点的深度值可以通过对局部表面方向求积分得到. 因此, 这类方法称为间接深度计算方法. 这

里我们将简要地介绍一些方法，更详细的阐述参见在其它章节。

#### (1) 光度立体

在光度立体方法中，使用不同方向上的三个光源来获取同一场景的三幅图像。在三幅图像的获取过程中，要求场景中的摄像机和目标静止不动。知道了场景中物体的表面反射特性，就可以计算由这三个光源照明的所有点的局部表面方位。这种方法的详细描述见第 9 章。光照立体方法的重要优点之一是三幅图像中的点是完全套准的，这是因为摄像机和场景是固定不变的。因此，这种方法不存在对应性问题。这种方法的最大缺点就是它是一种间接的方法，也可能由于照明无法得到精细控制而不能用于实际的成像系统。

#### (2) 由明暗恢复形状

由明暗恢复形状的方法主要使用图像灰度（阴影）变化来恢复物体形状的信息。这一过程是通过计算图像中每个点  $(x', y')$  对应的场景表面方位来实现的。除了由辐射测量原理产生的约束之外，阴影恢复形状的方法还假设了表面是光滑的，以便于计算表面方位参数。这种方法在第 9 章中已经详细介绍过了。很明显，阴影恢复形状方法是一种间接的深度计算方法。进一步的，光滑度约束在所有点都不能满足，表面反射特性也不总是知道的很精确，从而导致了不精确的形状重建。

#### (3) 由纹理恢复形状

图像的纹理特性变化，如密度、尺寸、和方位，都是建立纹理恢复形状的算法的基础。例如，被定义为纹理元素基元尺寸的最大变化方向和幅值的纹理梯度，决定了表面方位。量化纹理元素的形状（如圆显现为椭圆）的变化对确定表面的方向十分有用。构成规则网格纹理结构的表面图像（很可能是由于结构光照射的结果）可以通过寻找投影点（透视画中平行线条的会聚点）唯一地确定方位。由纹理恢复形状方法除了是一种间接的深度计算方法外，还存在精确定位和量化纹理基元及其特性的困难。由纹理恢复形状方法在第 7 章中已经进行了介绍。

#### (4) 由运动恢复形状

使用一个运动摄像机获取静止场景的图像时，场景点的图像平面坐标从一幅图像到另一幅图像的位移取决于场景点到摄像机的距离。这十分类似于前几节描述的立体成像方法。与之相对应的方法是使用一个静止摄像机获取运动物体的图像序列，从而在图像序列中产生运动视差。这样的视差取决于物体表面点的位置和速度。恢复物体结构和运动参数的方法将在 14 章详细介绍。

## 11. 5 测距成象

能够直接测量可视范围内每个点的距离，并将其记录为一个二维函数的系统称为测距成象系统，得到的图像称为距离图像或深度图。图 11. 11 所示的就是通过激光三角测距得到的一个深度图的例子。

最常用的测距成象方法有结构光测距和雷达。结构光测距成象系统使用三角测量原理来计算深度，而成象雷达系统则是通过比较发射信号与接收信号的时延或相位等方法实现距离测量的。按发射信号的波长，雷达可分为声雷达、激光雷达、毫米波雷达等。若按工作体制可分为脉冲飞行时间法，相位差或差频，测距仪来计算深度图。

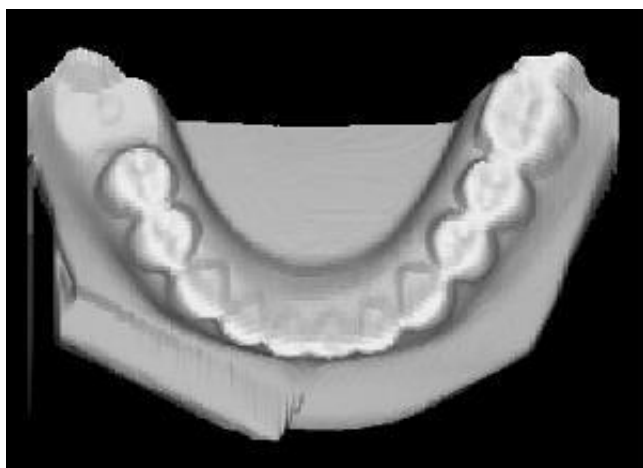


图 11. 10 一个牙齿石膏模型的深度图（北京医科大学口腔医院吕培军博士提供）

### 11. 5. 1 结构光测距

使用结构光照明的成象系统是指场景是由已知几何模式的照射光源照明的. 在一个简单的点投影系统中, 投影光源仪和摄像机之间相距一个基线距离  $b$ , 如图 11. 11 所示. 物体坐标  $(x, y, z)$  与被测量的图像坐标  $(x', y')$  和投影角  $\theta$  之间的关系如下:

$$[x, y, z] = \frac{b}{F \cos \theta - x'} [x', y', F] \quad (11. 23)$$

这样一个三角测量系统的距离分辨率可由投影角  $\theta$  和图像点水平位置  $x'$  的测量精度确定.

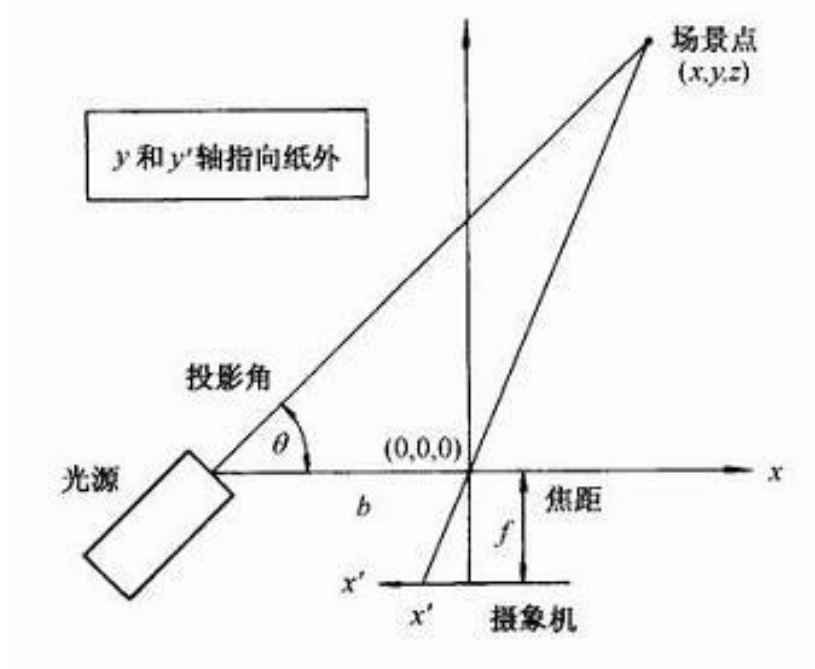


图 11. 11 以摄像机为中心的三角测量几何[Besl 1988]

为了计算所有点的深度, 可使用二维网格照明模式照明场景, 一次只照明一个点. 然后

使用上述方程计算该点的深度，由此得到二维距离图像。显然，由于这种方法获取数据的序列特性，其计算速度很慢，不适用于动态变化的场景中。一个典型的结构光照系统将光平面或者二维网格模式投影到场景中，与光源相隔一定距离的摄像机获取投影到场景中物体表面的光照模式。所观测到的光照模式图像含有变形，该变形由光照模式和物体表面的形状和方向决定，如图 11. 13 所示。注意，摄像机看到的网格包含了在方向和曲率上的不连续和变化。对应于图像平面中的任意点的三维物体坐标可以通过计算摄像机视线和光照平面的交点来确定。为了得到物体完整的描述，或者摇摆光源（如图所示），或者目标在一传送带上运动，以获取多幅图像。物体上的不同表面可以通过对相似空间属性的光条的聚类来探测。

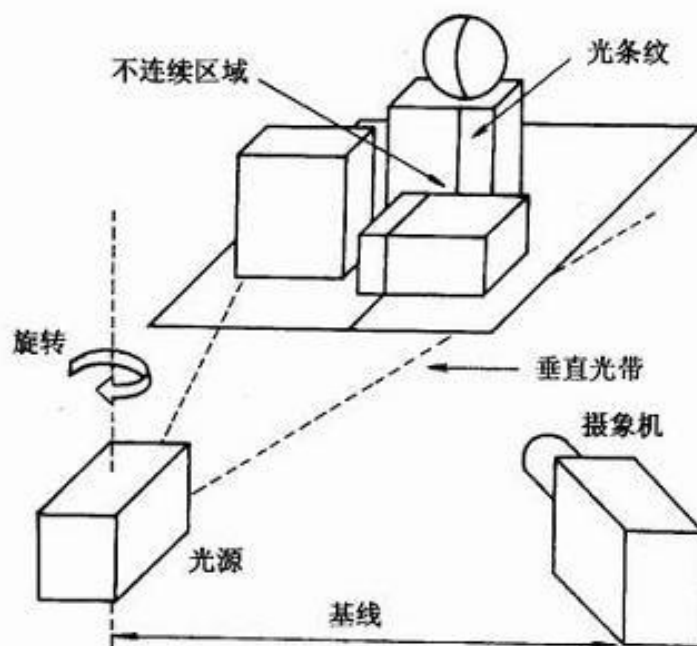


图 11. 12 结构光测距原理示意图[Jarvis 1983]

在动态变化场合下，依次向场景投射光带来获取覆盖整个场景的完整深度图像是不太实际的。如果为了获取完整的深度图像而将多条光带同时投影到场景中，则在匹配不同深度物体表面的光带时会产生潜在的混淆。在这种情况下，所投影的每条光带必须具有唯一性的编码。例如，用二进制编码方案，有可能只需投影  $\log_2 N$  个模式就可以获取完整的深度数据集，其中  $(N-1)$  是光带的总数。这种方法示意在图 11. 14 中，其中  $N=8$ 。

使用从 (001) 到 (111) 二进制码可以给七条光带中的每一条光带赋予唯一的编码。因为  $\log_2 8$  等于 3，所以仅能获取三幅图像。每幅图像可以通过三位二进制编码中比特(bit)位置 1, 2 或 3 来确定的。在图像中某一特定光带，如果其对应的比特位为 1，它就处于开的状态 ON。例如，光带 2(010) 只有在第二幅图像中才是 ON，但光带 7(111) 在所有三幅图像中都是 ON。在所有三幅图像中的光带都被唯一地确定，因此在匹配光带段时，不会产生混淆。在快速运动场景中，使用单色编码图像来取代多个二进制编码图像。

结构光照明技术已经广泛地用于工业视觉，因为在这种应用中可以很容易地控制场景照明。一个典型的应用是放在传送带上的运动物体经过一个光束平面，在光带图像中产生变形，然后计算光束平面上物体的轮廓。以规则间隔重复上述过程即可恢复物体的形状。

结构光照明技术的主要缺点是无法获取摄像机或光源看不到的物体点的数据。

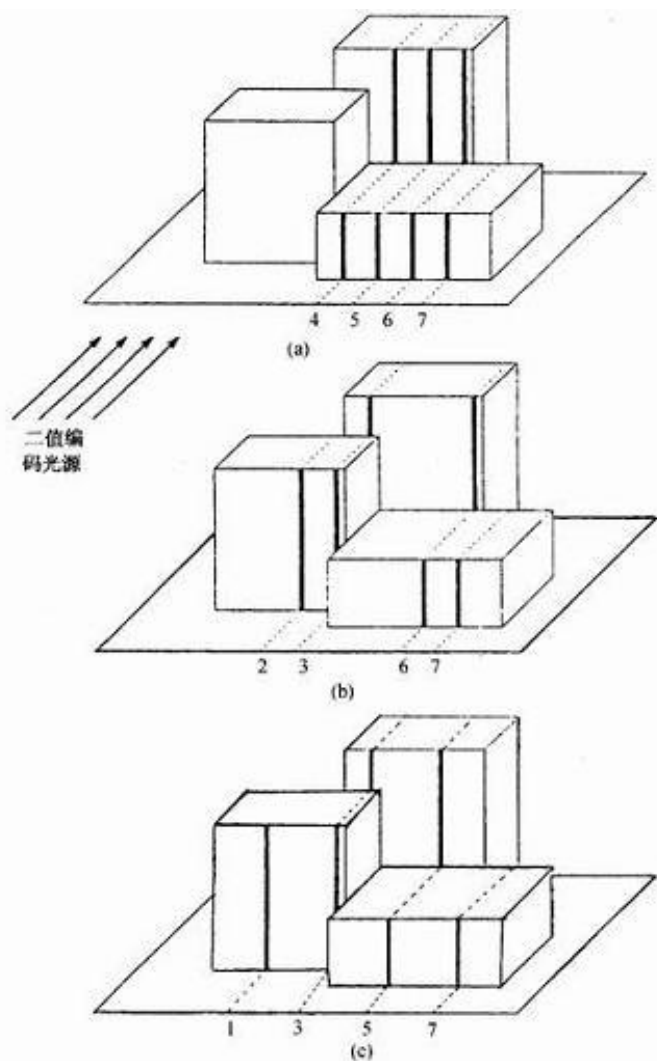


图 11. 13 二进制编码结构光照示意图，这里的投影序列决定了每条光带的二进制编码

## 11. 5. 2 激光测距雷达

在生物界，蝙蝠和海豚是通过接收自己发出的超声波来确定其周围物体的存在。根据生物的这种感知距离的能力，人类发明了各种测距系统，称之为雷达，即向空间发射信号，然后接收反射信号并与发射信号进行比较，以确定目标的距离和方位。激光测距（laser range finder）是在二十世纪 60 年代发展起来的一种新型测距雷达，现在已经进入实用化阶段。激光雷达具有波束窄、波长短等独特优点，因而具有极高的角分辨能力、距离分辨能力和速度分辨能力。激光雷达可以获取目标的多种信息，如反射特性、距离信息、速度信息等。在机器视觉领域，一般需要探测目标的细节及其精确的位置和方位，而声波甚至是电磁波都无法达到这一要求，只能使用波长很短的激光测距雷达。激光测距的主要问题是技术实现难度较大。

下面介绍三种工作体制的测距原理：脉冲飞行时间测距、相位差测距和差频测距。

### (1) 脉冲飞行时间测距

我们知道，声波和光波信号在空气介质中传播的速度是一定的，如果知道它们在空间中的传播时间，则可以计算波形在空气中的传播距离。人们将这种直接通过测量波形（或脉冲）飞行时间来计算距离的方法称为飞行时间法（time of flight）。激光脉冲测距雷达就属于这类测距系统。激光脉冲测距雷达向空间发射激光脉冲信号，并接收由于碰到物体表面而反射回

来的信号，然后测量发射和接收脉冲的时间差，就可以得到空间物体的距离。

图 11. 15 是一种典型的激光脉冲测距雷达原理框图，是 1977 年加州理工大学喷推实验室研制的 [Lewis 1977]。激光源是砷化镓固体脉冲激光器，波长为  $840nm$ 。激光源发出脉冲激光束，经光学校准系统到达光学扫描镜，对预定的空间进行扫描；光学扫描镜还对反射光同轴地接收，以保证测量的一致性。反射光经光电探测器转换进入脉冲前沿甄别电路，产生一个独立于信号强度的到达时间脉冲。发射参考脉冲与到达时间脉冲之间的时间间隔通过时间—脉冲高度转换器转换成具有一定宽度( $2\mu s$ )，其高度正比与时间间隔的脉冲信号。为了提高测量精度，可以对多次测量的高度值取均值。这种激光测距雷达在 1—3 米的测量范围内，测距精度为 2 厘米。获取  $128 \times 128$  图像需要 3 分钟。尽管上述的飞行时间测距原理十分简单，但要构造一个实用的系统比较困难，主要原因是测距系统的工作频率必须很高，才能得到较满意的距离图像，比如，为了得到测距精度为  $\pm 0.5cm$ ，则时间间隔的测量精度误差应小于  $\pm 30ps$  ( $1ps = 10^{-12}s$ )，要达到这样高的精度，只有采用原子物理仪器中的精密测量仪，如用于  $\gamma$  射线的测量的 Anticompton 分光计。

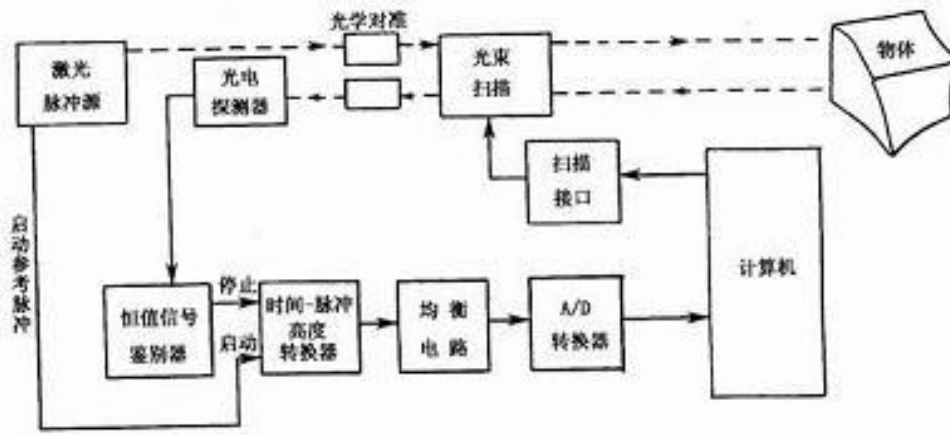


图 11. 14 脉冲飞行时间激光测距原理图

## (2) 相位差测距

相位差测距属于连续波雷达。深度信息可以通过检测调幅光波发射和接收的相位差来得到，这种原理的激光雷达的早期样机是美国 Stanford 研究所研制的[Nitzan 1977]，图 11. 16 是这种激光雷达原理框图。使用 15mw 氦氖激光器发出的激光的幅度被调制，调制频率为 9MHz，调制光经过平面镜分成两路，一路经过光学扫描装置向三维空间扫描，另一路进入相位差检测装置与经过物体反射后接收的光信号比较相位，如，调制频率为  $f_m$ ，相位差为  $\phi$ ，则物体到传感器的距离为：

$$d = \frac{c}{4\pi f_m} \phi = \frac{\lambda_m}{4\pi} \phi \quad (11.24)$$

式中  $c$  为光速， $\lambda_m$  为波长。由于相位差检测只能测量  $0-2\pi$  的相位差，因此这种类型的激光测距雷达的测距范围为  $D = \frac{c}{2f_m} = \frac{\lambda_m}{2}$ 。当测量距离超过  $D$  的整数  $N$  倍时，测量出的相位差是不变的，此时，式 (11.24) 变为：

$$d = \frac{c}{4\pi f_m} \phi + N \frac{c}{2f_m} \quad (11.25)$$

也就是说，这种激光测距雷达具有多义性，消除多义性的方法有两种，一是事先知道待测距离的大致范围，二是使用多台具有不同调制频率的激光测距雷达。

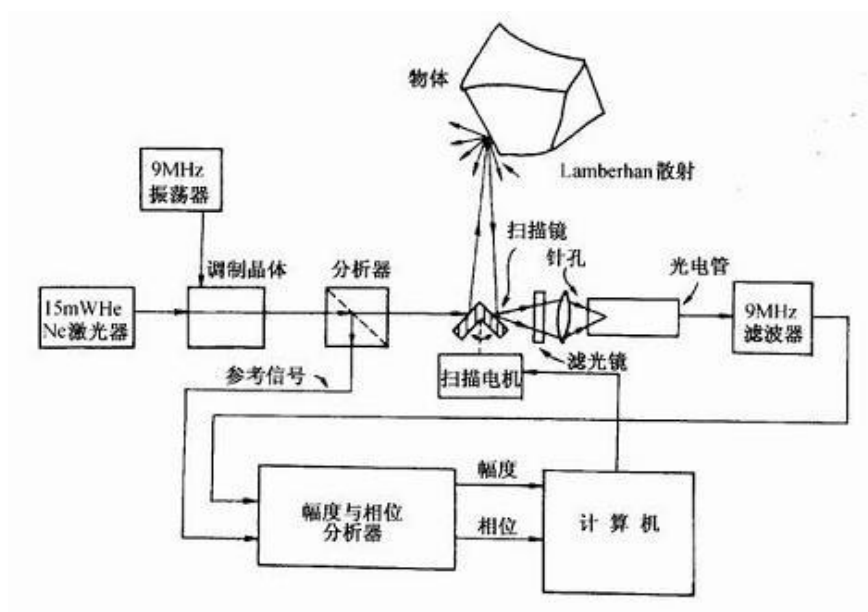


图 11. 15

### (3) 差频测距

差频测距也属于连续波雷达。当发射信号的频率在一个周期  $1/f_m$  内线性地扫过  $(\nu - \Delta\nu/2, \nu + \Delta\nu/2)$  区间，回波信号将在接收机与参考信号相混频产生一个差频  $f_b$ ，该差频与目标距离的关系为：

$$d = \frac{cf_b}{4f_m\Delta\nu} \quad (11. 26)$$

## 11. 5. 3 变焦测距

根据理想薄透镜定理，透镜中心（光学原点）到图象平面的距离  $z'$ 、透镜中心到场景点的距离  $z$  和透镜焦距长度  $f$  之间的关系为：

$$\frac{1}{z'} + \frac{1}{-z} = \frac{1}{f} \quad (11.27)$$

当  $-z \rightarrow \infty$ ，从图象平面到光学原点的距离  $z'$  等于焦距长度  $f$ 。当平行光线聚焦到图象平面上一点时，焦距长度是图象平面到光学原点的距离。在摄影测量学(Photogrammetry)中，光学原点称为投影中心， $z'$  称为摄像机常数。当透镜聚焦近距离点时， $z' > f$ ，因此使用  $f$  作为  $z'$  的近似实际上是低估了摄像机常数。当场景远点越远时，聚焦平面到光学原点的距离  $z'$  越接近  $f$ 。如图 11.17 是理想透镜成像几何示意图。



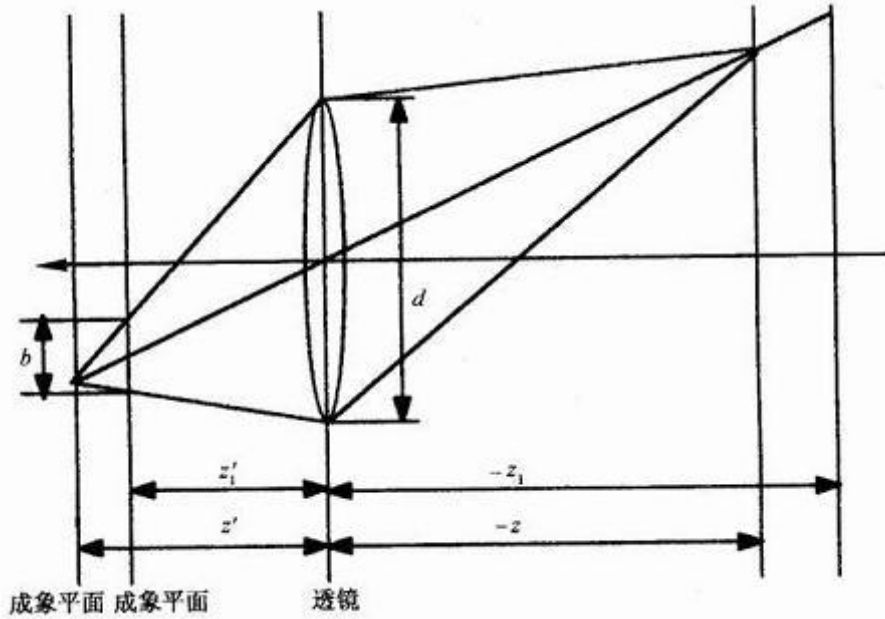


图 11. 16 薄透镜几何关系

对于图象平面到透镜中心的距离为  $z'$  的特定配置中，仅仅由方程 11.27 得到的在距离  $z$  处平面上的点是完全聚焦的。实际上，景深是由成像装置的空间分辨率决定的。低于成像分辨率的一定量散射可以容忍。在可接受散射水平下，存在一个图象平面距离  $z'$  的范围，以及一个对应的场景距离  $z$  范围，称为景深，其中的场景点在某一接受的程度上完全聚焦。

当某一场景点不聚焦时，它会在图象平面上产生一个圆斑图象，而不是一个点图像。如果圆斑的直径低于成像装置的分辨率，那么图象散焦量不会特别的显著。假定圆斑的直径为  $b$ ，透镜光圈直径为  $d$ ，焦距长度为  $f$ ，图象平面的理想位置为  $z'$ 。如果图象平面向透镜方向运动到达  $z'_1$  位置，即  $z'_1 < z'$  则图象的模糊程度由下式确定：

$$b = \frac{d(z' - z'_1)}{z'} \quad (11.28)$$

根据相似三角形性质， $b/2$  和  $z' - z'_1$  的比值必须等于  $d/2$  和  $z'$  的比值。我们可以从透镜方程(11.27)中求解  $z'$  和  $z'_1$ ，分别对应于  $z$  和  $z'$ ，并把这一表示式代入式(11.28)得到场景距离与模糊量关系的表示式

$$b = \frac{df(z - z_1)}{z(f + z_1)} \quad (11.29)$$

假定  $b$  是可接受散焦下模糊圆环的最大直径。解方程 11.29 求得  $z_1$ ，可以得到视野范围内远距离平面的距离表示式

$$z_1 = \frac{fz(d - b)}{df + bz} \quad (11.30)$$

计算视野范围内近距离平面的距离表示式为

$$b = \frac{d(z'_2 - z')}{z'} \quad (11.31)$$

其中， $z'$  是图象平面距离配置（在正确配置范围之外），它对应于模糊的最大量。请读者在图 11.17 中指出图像平面位于  $z'_2$  的位置。解透镜方程求  $z$  和  $z'$  并代入到方程求模糊直径：

$$b = \frac{df(z_2 - z)}{z(f + z_2)} \quad (11.32)$$

解上述方程求平面距离

$$z_2 = \frac{fz(d+b)}{df-bz} \quad (11.33)$$

景深  $D$  是近距离和远距离平面之差:

$$D = \frac{2bdfz(f+z)}{d^2f^2 - b^2z^2} \quad (11.34)$$

对于一个特定的配置, 焦距平面为  $z$ , 光圈直径为  $d$ , 焦距长为  $f$ , 最大可接受模糊直径为  $b$ , 这些方程提供了近距离和远距离平面位置。距离  $z = -df/b$  称为超焦距距离, 在此处, 远距离平(见(11.30)式)和场景深度(见(11.34)式)变成无穷大。

关于变焦测距以及从变焦恢复形状文献有[Pentland 1987, Nayer 1994, Krotkov 1987, Grossmann, 1987].

## 11. 5. 4 激光阵列成像

目前, 结构光测距系统技术和激光测距技术已经在物体建模和三维机器视觉导航方面获得相当成功的应用, 并且已经有商品化的系统。基于变焦测距技术的视觉系统, 由于其测距误差一般较大, 因此一直没有很好的应用系统。对于立体视觉系统, 由于系统的固有不适定问题, 发展十分缓慢。作为应用系统, 最重要的指标是可靠性高, 适应环境能力强。视觉系统的最主要应用场合是恶劣的环境, 如, 战场环境, 太空环境, 海洋环境等, 这些环境具有高冲击(比如火箭发射过程), 剧烈振动等特点。对于有运动部件或扫描部件的主动测距成像系统和主动视觉系统, 在这样恶劣环境下工作的生存能力很低。为此, 美国科学家研制出全固态(无扫描部件)的激光阵列成像雷达 Sandia 样机, 并正在实施应用。

激光阵列成像雷达能以很高的速率获取场景的稠密深度图, 该系统使用大功率激光二极管阵列向场景发射激光, 测量来自目标的相对于调幅载波相的反射光相移, 以便计算到目标的距离。接收器里的图像增强器的增益被和发送器同样的频率调制。到达 CCD 的光与返回信号的相位和它的强度无关, 而与目标的反射率有关。为了规范化反射率的变化, 返回波速的强度被采样两次, 一次是接收器调制增益作用, 另一次是接收器调制增益不作用。这样, 与每一个像素相联系的距离基本上同时在场景中测量。尽管这一系统具有能快速提供高精度的深度估计值的优点, 该系统使用了许多元件, 规模很大, 功耗大, 要想用于实际系统, 必须做进一步的工作。

## 11. 6 主动视觉

大多数计算机视觉系统都依赖于由固定参数的系统来获取数据, 包括所有的被动敏感系统(如视频摄像机)和主动敏感系统(如激光测距仪)。与这些数据提取方法不同, 我们认为主动视觉系统的数据获取特性和参数可以由场景解释系统动态地控制, 是感知系统中至关重要的部分。主动视觉不是一个新概念。生理系统通常是以主动方式获取数据的。主动视觉系统既可以用主动传感器也可以用被动传感器。然而, 在一个主动视觉系统中, 传感器的状态参数如焦距、光圈、聚散度以及照度都可以控制, 以获取有利于场景解释的数据。主动视觉是一个基本智能数据获取过程, 这一过程由场景测量和计算的参数与误差来控制。精确定义这些场景和内容相关参数不仅需要全面理解成像和处理系统的特性, 也需要全面理解相互依赖关系。

在文献[Jain 1990]中详细地介绍了距离图像传感、处理、解释以及应用。Besl[Besl 1988]给出了获取距离图像的各种方法, 并比较了它们之间的优缺点。[Jarvis 1983]早期的一个综述报告不仅包括直接距离测量方法, 还包括由二维图像特征信息计算距离的有关问题。Boyer 和 Kak[Boyer 1987]介绍了一种采用彩色编码从一幅图像中提取距离信息的方法。Will 和 Pennington 的经典论文[Will 1972]讨论了用网格编码和富里叶域处理技术, 以确定场景中不同二维表面的位置。

Bajcsy 在文献[Bajcsy 1988]提出主动视觉系统倾向于主动感知系统和控制策略方法。[Krotkov 1987]介绍了一个立体图像获取系统，该系统可以主动地控制焦距、缩放、光圈以及照明，以获取深度图。Aloimonos 等人[Aloimonos 1988]介绍了主动视觉系统可以完成各种视觉任务的优点，如从图像恢复物体形状。Ahuja 和 Abbot[Ahuja 1993]将用于表面估计的视差、摄像机以及镜头聚焦集成到主动视觉系统中。

## 思考题

- 11.1 使用一个测距摄像机获取一个边缘长为 10 厘米的正方体图像。摄像机轴线穿过立方体的两个相对顶点，其中最近的顶点距摄像机中心的距离为 10 厘米。摄像机所记录的灰度为  $1000/d$ ， $d$  表示沿摄像机轴到图像平面的距离(不是欧几里德距离)。请描绘通过摄像机获取的距离图像并计算由摄像机得到的立方体顶点灰度。
- 11.2 推导出图 11.2 所示的视差为 0 的表面方程式。
- 11.3 试阐述多基线立体视觉消除多义性对应点的基本思想。
- 11.4 比较结构光测距、激光雷达测距和变焦测距的优缺点。

## 计算机练习题

- 11.1 用两个摄像机步获取同一个场景的立体图像对，或用一个摄像机在不同的位置获取同一个场景的立体图像对。请编制一个程序来完成如下工作：
  - (1) 对第一幅图像进行边缘检测，然后使用图像强度相关性算法在第二幅图像中找到对应的边缘。
  - (2) 使用与第一幅图像相同的边缘检测器对第二幅图像进行边缘检测，比较边缘检测器输出的边缘图和通过相关性算法得到的对应边缘图。能否编制一个程序将图像边缘检测和相关性方法结合起来找出对应的边缘图。
  - (3) 请输出立体图像对的视差图。