

第七章 运动视觉

7.1 概述

7.2 视觉运动的检测和测量

7.2.1 以灰度为基础的方法

(1) 速度场和光流

(2) 光流

(3) 光流的估算

7.2.2 基于过零点的方法

7.2.3 基于特征的运动检测方法

7.3 运动理解

7.3.1 基于光流的运动理解

7.3.2 基于特征的运动理解

7.3.3 根据图象流动力学恢复表面结构和三维运动

7.3.4 基于光流的代数法恢复运动参数和景物深度^[Tek 96]

7.1 概述^[Ull 79]

对运动的分析在生理系统中占中心位置。甚至在简单的动物身上都存在着复杂的观察、跟踪和利用运动的功能。例如，青蛙可以有效地探测小飞虫。这样的动物可有选择地对视场中运动着的小的、深色的物体进行跟踪。家蝇可以跟踪运动物体和发现目标跟背景之间的相对运动，甚至当物体与背景在纹理上相同，因此如果没有相对运动是区分不开的情况下也是这样的。

在高等动物包括灵长目动物中，运动的分析从早期视觉处理阶段开始就渗透到视觉系统中。某些生物，例如鸽子和兔子在视网膜的水平上完成基本运动分析。在其他动物，包括猫和灵长目动物，视觉皮层中从眼睛接受输入的第一个神经里已经涉及到运动的分析：这些神经对沿某一方向运动的刺激反应很灵敏，但对沿相反方向的运动却反应很小，或根本不反应。

运动视觉在生理系统中占主要地位不令人奇怪的，因为运动揭示了关于环境的有价值的信息。本章的研究任务是根据运动发现结构和解释由运动所产生的光流场。对生理系统运动视觉的研究将有助于时变图象的计算机分析算法的研究。

反之，对时变图象解释的研究可以深化对一般原理的理解，这些理解可增进我们对生理视觉系统的理解。时变图象分析的基本计算问题可分成两大类：运动的检测和测量，以及视觉运动的解释。

7.1.1 运动探测和测量

图象中基元和区域的运动不是直接给出的，而是必须根据基本量测来计算得到。由眼睛或由电子成象系统记录的时变图象可描述为随时间变化的亮度值 $E(x, y, t)$ 的二维阵列。图象中的运动可用速度场 $V(x, y, t)$ 来表示。 $V(x, y, t)$ 给出某一时刻 t ，在图象中 (x, y) 处点的运动速度和方向。 $E(x, y, t)$ 是由原始的测量直接得到的，而 $V(x, y, t)$ 不是这样。因此，分析运动中的第一个问题是根据 $E(x, y, t)$ 计算 $V(x, y, t)$ 。这些计算是视觉运动的量测。

在某些情况下，只要检测速度场 $V(x, y, t)$ 的某些特性就足够了，而不要完全和精确的量测。例如，当希望迅速地对运动物体作出响应时，在这种情况下必须检测运动，但不需要测量。

但是，从这里讨论问题的分类来说这两个问题都重要。当研究深入时，这些问题将被证明比原来所预计的更为困难。因此，对有效和可靠的测量方法的研究是对时变图象分析方面的重要研究邻域。

1. 离散和连续运动

对由人类视觉系统进行的运动检测和测量的心理学研究确定了两种类型的运动：离散的和连续的。对感觉运动的观察者来说，刺激物不一定要连续地移过视场。选择合适的空间和时间显示参数，可以给出对顺序刺激的平滑、不中断运动的印象。视觉系统可以填补离散显示之间的间隙，甚至当刺激物相隔几度视觉角，和长的时间间隔（400ms）也是这样。所得到的运动，叫做表观的（*apparent*）或 β 运动，从感觉来说跟连续运动无法区分。此外填入的位置可为以后的处理，例如立体视觉所利用。表观运动机理在人类和低等动物中都是天生的。

表观运动现象引起了离散和连续运动是由相同的还是由分别的机理所记录的问题。视觉系统可以记录两种类型运动的事实不一定意味着分别的实现机理，因为从原理上来说一个记录离散运动的系统可以记录连续运动。最近的心理物理学方面的证据支持存在两种机理。Braddick^[Bra 74]提出把这两种机理叫做短范围和长范围（*short range and long range*）。短范围机理测量连续运动或大约 15 弧分（在视场中心）（*minute of arc*）和小于大约 60~100 msec 时间间隔的离散运动。长范围机理处理比较大的位移和时间间隔。这样的术语比用离散 / 连续这样的分类法更好些，因为大到 15 分的视觉弧度跳变的离散显示是由短范围机理来处理的。

这样的两个系统之间存在着比它们的范围差别更为基本的差别。它们似乎是根据不同的运动基元，在不同处理阶段进行各自的运动量测。在量测运动中，对这两种主要的处理方案进行区别是有益的。在最低层，运动量测是直接以灰度值的局部变化为基础，这称为以灰度为基础的方案。另外还有一种可能是首先识别象边缘、线、斑点或区域这样的特征，然后通过时间和位置的变化范围内匹配这些特征来进行检测，这种类型的方案被称为标记匹配方案（*token-matching schemes*）。在人类视觉系统中，似乎短范围过程是以灰度为基础的方案，长范围过程是标记匹配方案。

这两种运动检测和测量方式将产生不同的计算问题，并且随之在生理系统和计算机视觉系统中产生不同种类的过程。

2. 以灰度为基础的方案

已经提出了几种在生理系统中的以灰度为基础的运动测量方案。这些方案可分成两个主要类型：相关技术和梯度方法。

(1) 相关方案

通过把在两个相邻位置处光增量检测器的输出作比较可构成简单的运动检测器。在位置 p_1 和时间 t_1 的输出跟位置 p_2 在时间 $t - dt$ 的输出相比较。这种方法称为延迟比较方案，该方案的两种变种已经提出来作为生理系统的模型。按第一种模型，生理系统把两个值相乘，也即 $D(p_1, t) \cdot D(p_2, t - dt)$ ，其中 D 表示为单元（*subunit*）的输出（图 7.1）。如果一个光点由 p_2 点在时间 dt 内移到 p_1 ，这样在 p_2 点处造成光线增加的 dt 时间后，造成了 p_1 点处光线的增加。因此，上述乘积是正的。在这样的检测器阵列中，平均输出本质上等于输入的交叉相关（*cross-correlation*）。这种模型可成功地描述各种昆虫对在它们视场中的运动所作的反应。

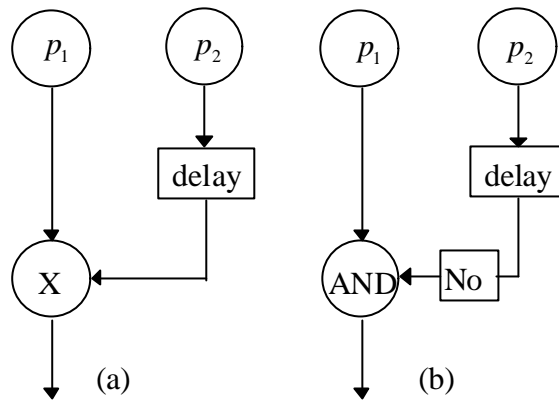


图 7.1 延迟比较运动检测方案

P_1 和 P_2 是对光点瞬时作出响应的检测器

(a) 对以合适的速度向左移动的光点, P_1 和 P_2 的响应相重合, 在组合单元处产生正的输出; (b) 禁止方案: 对从 P_2 和 P_1 的运动不产生响应, 因为 P_2 从的延迟响应抵消了来自 P_1 的响应。

另一种比较简单的方法是 And-Not 方法, 这是由 Barlow 和 Levick 针对在兔子视觉皮层中有方向选择性单元所提出的模型, 这也是 Emersso 和 Gerstein 针对猫的视觉皮层提出的模型。这些单元被称为方向选择性, 因为这些单元对沿所谓的优先方向运动的刺激物的反应要比对沿相反方向运动的刺激物的反应强烈得多。因为 Barlow 和 Levick 发现了方向选择性机理之间固有的相互影响的证据, 他们提出一种模型, 按这种模型运动检测器计算 $D(p_1, t)$ 和 $D(p_2, t - \delta t)$ “Not” 的 “And” (图 7.1(b))。按这种方案, 从 p_2 到 p_1 的运动由于来自 p_2 的响应被延迟而被禁止, 而从 p_1 到 p_2 的运动产生一个正的响应。

Torre 和 Reichardt 对家蝇的视觉系统提出一种相似的方案, 按这种方案, 用低通时间滤波代替延迟。Torre 和 poggio 描述了一种实现这种计算的很巧妙的触突 (synaptic mechanism) 机理。

延迟比较的某些通用特性值得加以注意。第一, 这些检测器不仅有选择地对连续运动作出响应, 而且对在 p_1 位置和 p_2 位置之间刺激物的跳跃也作出响应; 其次, 这样的检测器有一些明显的局限性, 例如, 运动的速度必须在某个由延迟(或低通滤波)接收器的分离距离决定的范围之内; 最后, 单个这种类型的检测器的输出不能可靠地决定运动测量。例如, 在有许多运动物体的视场里, 如果在 p_1 点的检测器由一个运动物体所激发, 而在 p_2 处的检测器又由另一个物体所激发, 这种类型的运动检测器就会错误动作。为了得到准确可靠的运动测量, 来自这样检测器阵列的输出应被组合在一起。

另外的相关技术, 例如利用原始灰度值上的交叉相关, 已被作为人类视觉系统中的运动测量模型提出来。Anstis 提一种相减的方法, 按这种方法把两帧相继图象移位相减, 所得图象中的减小点就标志偏离 (displacement)。一般来说, 还没有得到关于生理上实现这样技术的模型描述。似乎还没有很有说服力的证据说明在人类视觉系统中存在这样的模型。

(2) 梯度方案 (gradient schemes)

最近一种用于生理系统运动检测和量测运动的梯度方案, 它被作为单个皮层细胞进行运动分析的模型。这些在猫和猴子的主要视觉皮层 (perennial visual cortex) 中发现的细胞有选择地对边缘和光带 (Bar of light) 作出响应。这些细胞对方位, 并且经常对运动的方向也是有选择性的。这就是说, 要激发这样的单元, 刺激物必须具有这种单元所选择的方位, 并且必须沿所选择的方向运动。因此对这种简单细胞的结构和功能的分析揭示了视觉运动的早期检测和测量的机理。这些机理可被用于计算机视觉系统。为了理解简单细胞的作用, 我们需要粗略地描述视网膜对图象的处理。

(3) 视网膜对图象的处理

视网膜结构主要完成两个功能。第一个是把进入的光线记录在对光敏感的光接收器阵列上。第二，所记录的图象完成初始的转换。所转换的图象从最后的视网膜细胞层（神经节细胞层）沿着大约上百万的神经纤维送到 LGN（外侧膝状体），并从 LGN 送到视觉皮

层。如前所述，视网膜对图象 I 进行 $\frac{d^2}{dx^2}(G * I)$ 运算， G 是高斯函数。运算后结果中的过零点对应于原始图象中灰度急剧变化处，利用附加的子单元可确定过零点的运动。设 Z 表示过零点当前的位置。从图 7.2 可看到，如果过零点向右运动，在位置 Z 的卷积值就增加；如果过零点向左运动，这个值就减小。因此，通过检查卷积对时间的导数的符号就可以明确地确定运动的方向。

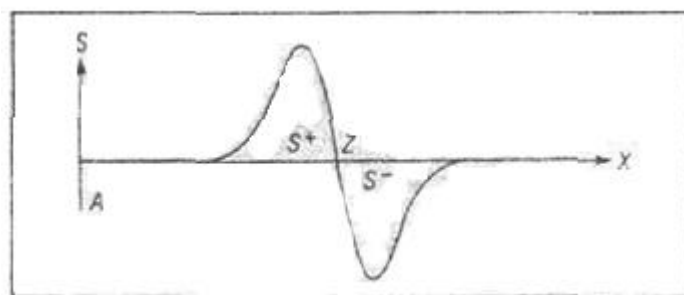


图 7.2

按这种方案，运动可由过零点的斜率和时间导数来确定，因此，这是以灰度为基础的方案，它不要求在时间域上作元素的匹配。在相反方向上的运动是由另外的单元来检测的。这样的关于人类视觉的观点得到心理物理学证据的支持。

利用这种方案，可以将过零点轮廓用于运动量测。然而，由于图 7.3 中所示的孔径问题，以纯粹的局部测量为基础既不可能完全地确定轮廓的运动，也不可能完全地确定任何其他线性特性。如果是由跟轮廓相比较小的单元来检测运动，可以抽取的唯一信息是垂直于局部边缘方向的运动分量。沿边缘的运动是感觉不到的。要完全地确定运动，在第二阶段必须组合局部量测，或者在局部邻域中沿着轮廓检测。

(4) 计算机视觉中的以灰度为基础的技术

已经提出了许多用于计算机视觉系统的以灰度为基础的运动检测和测量的方案。其中一个方案是用于测量图象对中运动的灰度交叉相关技术。例如，这种技术可用于根据卫星图象数据测量云层的运动，以及用于交通控制和用于 SLR 图象的比较。已提出一种基于比较小的图象区域上灰度分布的方案。图象相减的方法已被用于运动和变化探测以及用于运动测量。

大多数交叉相关法和相减法的基本问题是这些方法假设图象（或大部分图象）是作为整体在两帧图象之间运动。包含独立运动物体的图象以及由物体在空间无限制地运动引入的图象畸变造成了这些技术的困难。

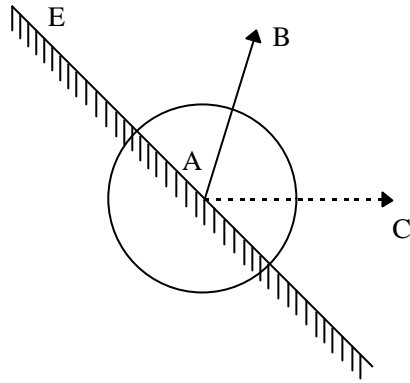


图 7.3 孔径问题

孔径比运动的轮廓小时，不能唯一地确定运动方法，从对孔径 A 观察不可能确定边缘是朝 B 方向还是朝 C 方向运动。

(5) 计算机视觉系统中的梯度方法

这些方法都基于在给定点上的灰度梯度和在这些点当灰度图案运动时引入的灰度随时间变化之间关系。如果 $E(x, y)$ 表示图象中的光强，那么

$$-\frac{dE}{dt} = E_x u + E_y v$$

其中 dE/dt 是在位置 (x, y) 处灰度随时间的变化； E_x 和 E_y 表示在这个图象点的灰度梯度； u, v 是沿 x 和 y 方向的局部速度。

梯度方法和过零点方法在几个方面是相似的。这两种方法都利用时间变化和图象梯度去测量在梯度方向上的局部运动。

以灰度为基础的图象速度场计算方法还有一些困难问题有待解决。因为在一般情况下（无限制的运动，几个物体）图象的不同部分有不同的运动。所以，原先的初始量测是局部的。这些量测不足以完全确定运动。因此，在稍后的阶段必须将局部测量组合起来。组合阶段是理解生理系统和构成计算机视觉系统中主要的未被解决的问题。

3. 标记匹配方案

按测量运动的标记匹配方法，首先要确定可识别的基元一标记，然后匹配不同时间的标记。以上所讨论的表观运动现象（apparent motion phenomena）说明了人类视觉系统有能力通过匹配在空间和时间上相距相当间隔的标记来建立运动感觉。在检测相继出现的基元之间的连续运动时，视觉系统必须建立依次出现的基元之间的对应关系。这就是说，在第二帧图象中必须找到跟第一帧中的基元相对应的基元的位置。一个简单的对应问题如图 7-4 中所示。图中实心的小圆表示在第一帧中的，空心小圆表示在第二帧中的。在这两帧图中有两种可能的一对一的匹配，这导致两种可能的可感觉的运动：水平的和对角线的。

在图 7-4 中只有两种方式的多义性，在实际中，每帧图中可包含许多排列成复杂图形的基元，因此必须在它们之间建立对应关系。对控制人类视觉系统中对应过程的规律已进行了若干研究，但离完全理解这个问题还很远。

当参加运动的基元是一些孤立的点，它们的对应关系主要由点之间的距离来控制。在其他参数相同的条件下，每个点优先跟下一帧中相距最近的点匹配。

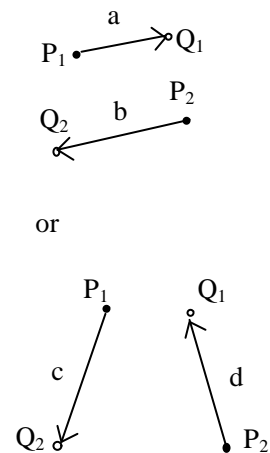


图 7-4 对应性问题， P_1 和 P_2 在第一帧， Q_1 和 Q_2 在下一帧。

如果每帧图中的基元是短线段，那么可应用附加的规则，对短线段来说，对应关系不仅取决于基元之间的距离，而且跟它们的相对长度和方向有关。在其他参数相同时，给定的线段优先跟另一根有相似长度和方向的线段相匹配，这就要规定线段之间的相似性矩阵。相似性矩阵是相似性、方位、长度和对比度的函数。每个基元倾向于跟最相似的基元匹配。这种选择性的简单和著名的例子是运动图象中马车车轮现象，在这样的图中，带辐条的车轮的表观运动跟车轮的实际方向相反。这种现象是视觉系统从两个相竞争的运动中优先选择其中距离和角度变化量小的那个的结果。

形体（figure）和基元组之间的对应关系不是只以相似性为基础。还存在着在一对一的基础上匹配基元组这样的倾向，一个基元可能不跟最相似的基元匹配，因为这个基元已经找到它的对应物，而通常二对一的匹配是要避免的。

标记匹配方案的两个常见的问题跟生理系统和跟计算机视觉系统都有关系。

第一个问题涉及在怎样的水平上建立对应关系。这就是处理的程度和分解标记的复杂程度。可以在象点、斑点（blobs）、边缘段或短线段这样的简单标记之间建立匹配，也可以在具有结构形式的甚至已识别物体的图象这样的复杂标记之间建立匹配。应用复杂的标记可以简化对应过程，因此通常复杂的标记在后续的那帧图象中只有唯一的对应物。与此相反，原始的标记象短的边缘段通常存在许多相互竞争的可能匹配。但是应用原始标记具有两个突出的优点。首先是降低了对预处理的要求。这在运动感觉中是非常重要的，因为这时对计算时间的限制很严格。第二，原始标记允许在各种物体作复杂外形变化的运动时进行对应关系的匹配。这是因为复杂形体的对应关系是通过构成形体的基元组之间的匹配来建立的。因此，似乎对应关系应该建立在相当原始的水平上，可能在 Marr 的完全初始简图水平。

第二个问题涉及在视觉运动分析的集成系统中基于灰度的方案和基于标记匹配方案可能起的作用。基于灰度的方案一般比较快和灵敏。例如，人类视觉系统可以探测低到每秒一视觉弧分（one minute of visual arc）的速度。这个速度仅为太阳径过天空速度的 4 倍。猫的视觉皮层中的方向选择单元能可靠地对小到 0.87 弧分的偏离作出响应（大约 4 米距离上 1 毫米）。人类甚至可探测更小的偏离。但另一方面，局部测量的不确定性使得单独地以灰度为基础的方案来准确求得速度场很困难。从原理上讲，基于标记匹配的方案可以在长距离范围内跟踪明确定位的标记（象线段的端点），因此可达到高的准确性，这样做的代价是需要比较费时的标记定位和解决对应问题的过程。

根据它们的基本特性，这两种运动测量方案可以完成不同的视觉任务。基于灰度的方案一般用作为外围的，吸引注意力的和早期警报系统和用于把运动的物体跟背景相分割的系统。在根据运动发现物体三维结构的任务中标记匹配方案可能起主要作用。在发现三维结构的任务中要进行在相当长的距离范围内准确跟踪。某个最新的心理物理学的证据说明在人的视觉系统中长范围过程对根据运动发现结构来说是关键的。第二种可能性是这两种方案相互作用以便相互补充。例如，长范围对应关系的计算可以由通过短范围方案提供的附加约束来导引的。

7.1.2 视觉运动的解释

生理的和基于计算机的运动分析系统有三种用途。这些用途按其复杂性排列如下：

- (1) 周围物体的分割
- (2) 发现三维形状
- (3) 基于运动的识别

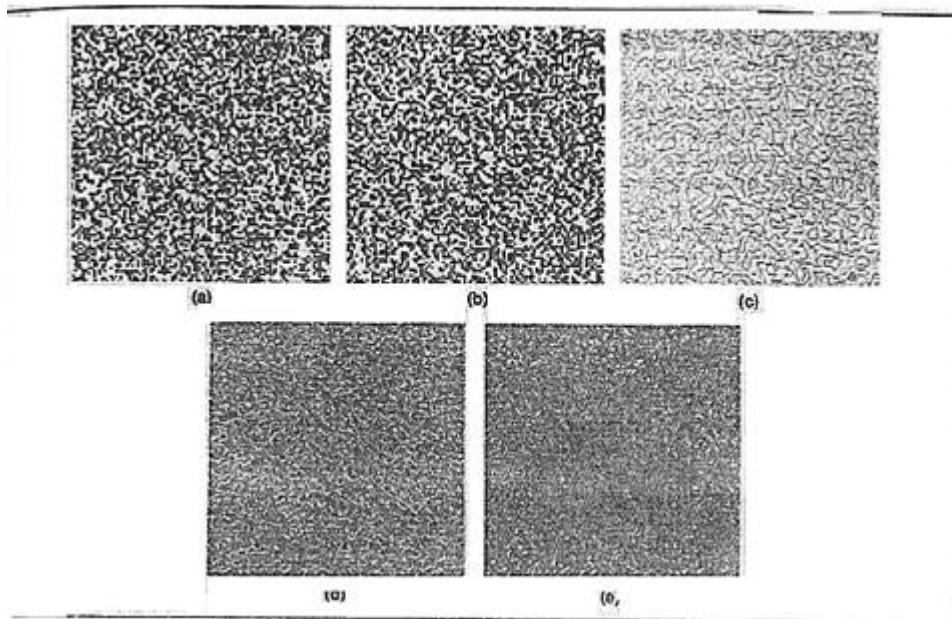


图 7-5 (a) 中的中心方块比(b)中的中心方块略微右移。这两幅图象中的背景是不相关的。当把这两幅图顺序地让观察者观察时，就可感觉到用线条勾划出方块在噪声背景上移动。(c)是(a)用 $\nabla^2 G$ 算子滤波后的过零点轮廓。(d)中过零点的运动是沿亮点的方向。(e)根据运动信息就可以发现中心方块，在此区域中的亮点被删去。

1. 周围物体的分割

人类视觉系统可以只根据运动信息把运动物体跟它的背景分割开。图 7-5 表示可从一对随机点模式中探测一个运动的子形体的例子。图 7-5a 中的中心方块比 5b 中的中心方块略微向右位移。这两幅图象中的背景是不相关的。当把这两幅图顺序地让观察者观察时，就可感觉到用线条勾划出的方块在噪声背景上移动。这个子形体在这两个静止图象中都探测不出来，因为这些子形体仅仅由帧对帧的偏离来定义的。这种实验证明视觉系统利用运动甚至在无灰度边缘或在边界处没有纹理变化时，能描述运动物体的边界。

研究表明视觉运动可对周围物体的分离提供有用的线索。然而，可靠的基于运动的分割比预想的要困难。从以上叙述的基于灰度的方法可理解某些这样的困难。在把局部测量综合起来时，经常要假设速度场是连续的。因此在物体边界处就会产生显著的误差，因为在边界处连续性假设不成立。合乎理想的是在测量运动以前就探测出速度场的不连续性，而不是根据速度测量。

克服这种困难的一种方法是利用探测运动不连续性不要求准确的速度场测量，粗略估计方向和速度就足够了。所以，分两阶段进行局部的基于灰度的运动测量被证明是有利的。在第一阶段，对运动作粗略估计，并用于定位速度场中的不连续点。第二阶段，在建立速度场时不把平滑性约束传播过不连续点的边界。

2. 根据运动发现三维结构

人类视觉系统甚至能从不熟悉物体的景物图象序列发现运动物体的三维形状。而其中的每个静图象并不包含三维信息，在实验中不熟悉的物体在透明的屏幕后面旋转，从屏幕的另一边看到物体垂直投影的阴影。在大多数情况下，观察者可以正确地描述被挡住物体的三维结构和在空间的运动，甚至当每个静止观察是不能识别和不包含三维信息时也可以做到这点。原始运动深度效应主要利用线框物体，这些物体投影成一组相连的线。最近的研究显示根据包含在运动中的不连接的基元可建立三维结构。

对根据运动发现三维结构的计算问题已进行了许多研究。其中所探索的主要问题是什么条件下从运动得到结构的问题有唯一解，以及研究根据运动发现结构的算法。这些研究的主要结果列出在表 1 中。如表中所示，对问题的离散形式和连续形式都进行了研究。第一个形式在空间和时间上都是离散的。计算是以若干离散的帧，每帧包含许多孤立的点。已经证明为了从运动中得到结构的唯一解释只要少数帧和点数就足够了：对包含至少四个不共面点的运动刚体而言，用三帧就可以确定空间运动和点的三维坐标。

唯一性的证明是很有价值的，因为这将导致根据运动发现结构的可能方案。景物被分成大约每组包含 4 个基元的组。每组结构是独立地求取的，最后局部的结果在一个附加的步骤中被组合起来。这样的方案是基于垂直投影，通过用发现方案的局部化来处理透视投影。这就是说，对四邻域点，两种类型的投影是相似的。可也能直接利用透视投影。有迹象表明在透视情况下，对唯一的解释来说，可能需要两帧图和 5 个点就足够了，但这种情况的证明还不知道。

也可以不是只考虑在若干离散观察中点的位置，而是把点的位置和速度同时作为输入（表中第二栏）。这可能被考虑为两帧问题，帧间的时间间隔趋于零时的极限情况。这时问题就取以下形式：给出图象中 N 点的位置和速度，确定这些点是否属于单个运动物体和寻找在空间中的运动和这些点的三维坐标。

先期的理论问题是确定能发现问题的唯一解时的最少点数 N，从数学观点来看，这个问题仍然没解决。计算方程式的变量和未知数的数目揭示 5 个点就足够了，一个证明是由 progeny 执行的计算机程序。因为计算机程序对输入中的误差很敏感，特别是当被察的物体很小时，似乎强有力的发现算法应需要多于 5 个点。

表 1

离散点和离散观察帧	离散点及其速度	速度场和速度场的空间导数
3 个垂直投影中 4 个点 (ullman 1979)	单个透视观察投影中 5 点和它们的速度	对纯平移的运动有唯一解 (clock SIN 1980)
两个透视投影中的 5 个点 (ullman 1979)		对一般的运动最多可有三个解 (longuet-higgins and prageny 1980)

问题的第三种形式是利用速度场和速度场的空间导数。这可被看作是前面形式当点之间的距离趋于零时的特殊情况。在这种问题的简化类型中，空间的运动被限于纯粹平移。在这样的假设下，解的唯一性很容易建立。根据所在点的速度和速度场的空间导数可求得每个图象点的表面方位。

在包含旋转分量的更为一般的情况则要复杂得多。有一种分析方法认为某一点的速度最多时有三种解释。更准确地说，已被证明对非平面的表面（在点上给出速度和它的一阶和二阶导数）。对这个点上表面方位最多存在三个解。

上述的方案是为了根据运动到结构问题的完全解。这就是说，这些方法试图发现所有的运动参数和可见环境的三维结构。与此相反，部分描述要有选择地抽取出某些运动参数。例如若干数学和心理研究试图在图象速度场中孤立出以下这样的变量：这些变量允许分离和独立地发现运动的旋转和平移分量中的每一个。一般来说，这样地分解问题是不可能的。因为图象中的速度场是由所有的运动参数跟可见的物体三维形状的内在相互作用所决定的。因此，似乎不可能在运动场中孤立出一个变量。譬如说跟平移分量无关的旋转运动分量，然而在象纯粹平移这样的特定情况下的部分解变成有吸引力的。

3. 基于运动的识别

人类观察者可单独根据特性运动模式识别某些物体。由 Johansson 进行的研究为这种能力提供了强有力的证明。通过对在黑暗中运动的，在其主要关节上装有小光源的演员摄影可提供这样的证明。每个演员由多达 13 个运动光点来表示。所得到的点的动态模式为运动着的演员造成一种生动的印象。以后的研究已经证明，经常可以把男演员跟女演员区分，在某些情况下，熟人可以单独根据运动的光点被识出。

研究这个问题的策略可把问题分成两步：第一步组织和描述运动模式；第二步把得到的表达跟存在内存中的相似描述比较。第一步被认为是自动的。也就是由通用结构成份

(principle)、而不是由关于特定物体的知识所导引。不少心理学研究已证明在人类视觉系统中存在这样的组织成份,但对其细节还远远不清楚。最近已有人提出一种方法,根据平面性假设,四肢的运动可被用于求得对运动着的光的显式的三维解释。当这些原理用于 Johansson 类型的结果结构时,组织成份将在主要关节之间建立连接,产生正在运动的棒状图形的表达形式。这些棒状图形在以后的步骤中可能用 Marr 和 Nishihara 所提出的方案进行分析。

经验证明在这个领域中的进步取决于计算研究跟实验心理学研究的结合。

7.2 运动的检测和测量

7.2.1 以灰度为基础的方法

1. 速度场和光流 (motion field and optical flow)

当物体从摄像机前面移过或摄像机移过固定的环境时,图象中会产生相应的变化,这些变化可用于发现相对运动以及物体的形状。

图象中各点的速度矢量构成了速度场。在某一特定时刻,图象中的 p_i 点与物体表面上的某点 p_0 对应。这两点之间的关系符合投影方程。在透视投影的情况下,从图象点出发经过透镜中心的射线一直伸展到与不透明物体相交(图 7.6)。

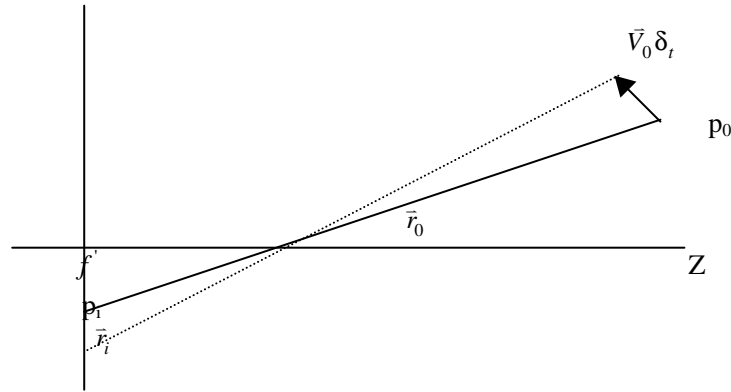


图 7.6 景物中点的位移造成图象中相应点的位移

设物体上的点 p_0 相对于摄像机的速度为 \vec{V}_0 , p_0 点的运动造成图象中的对应点 p_i 以 \vec{V}_i 的速度运动。 dt 的时间间隔里 p_0 点移动 $\vec{V}_0 dt$, 对应的图象点 p_i 移动 $\vec{V}_i dt$ 。速度分别为:

$$\vec{V}_0 = \frac{d\vec{r}_0}{dt} \quad \text{和} \quad \vec{V}_i = \frac{d\vec{r}_i}{dt}$$

其中 \vec{r}_0 和 \vec{r}_i 的关系是

$$\frac{1}{f'} \vec{r}_i = \frac{1}{\vec{r}_0 \cdot \vec{Z}} \cdot \vec{r}_0$$

其中 \vec{Z} 是 Z 轴方向的单位矢量。对此透视投影方程作微分可得

$$\frac{1}{f'} \vec{V}_i = \frac{(\vec{r}_0 \cdot \vec{Z}) \vec{V}_0 - (\vec{V}_0 \cdot \vec{Z}) \vec{r}_0}{(\vec{r}_0 \cdot \vec{Z})^2} = \frac{(\vec{r}_0 \times \vec{V}_0) \times \vec{Z}}{(\vec{r}_0 \cdot \vec{Z})^2}$$

在这里并不想去求解这个方程,重要的是说明可用这种方式对图象中各点赋予一个速度矢量。这些矢量构成速度场。

物体上相邻点的速度相近。因此,可以认为在多数场合下图象中的速度场也是连续点。在物体成象的外轮廓处,由于那里的运动场是不连续的将作为例外。

2. 光流 (optical flow)

当成象物体运动时，图象中的亮度图案也随之移动。光流是可看得到的亮度图案的运动或称为表观运动 (apparent motion)。希望光流能相应于运动场，但以下将会看到不一定是这样的。

先研究一下在成象系统前旋转的绝对均匀的圆球。由于球面是曲面，所以球的图象中会有亮度的空间分布即影调。但这样的影调不随表面运动，所以图象不随时间变化。这时各处的光流都为零，尽管这时运动不为零。其次，看一下由运动的光源照明的固定球面。随光源的运动，图象的影调将随之变化。这时光流显然不为零，而运动场则到处都是零。虚象和阴影是使光流和运动场不相同的其它的原因。

我们可以观察得到的是光流。此外除了上述的些特殊情况外，一般情况下光流与运动场相差不大。这使我们能利用光流信息来估计相对运动。

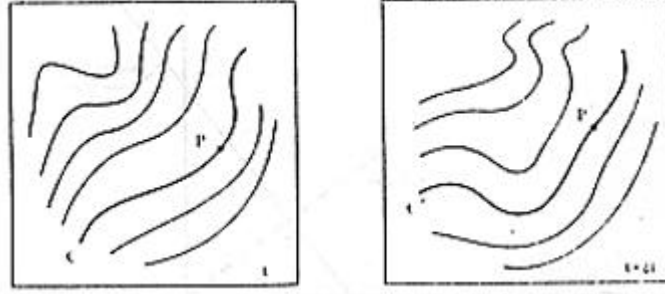


图 7.7 亮度图案的表观运动是个不便于使用的概念

亮度图案的表观运动是什么含义呢？试考虑图象中在时刻 t 时亮度为 E 的 p 点 (图 7.7)。在 $t + dt$ 时刻图象中与 p 点对应的点是 p' 。在此时间间隔中亮度图案是如何运动的？通常在 p 点附近有许多亮度都为 E 的点。如果亮度连续变化，那么 p 点将位于等亮度线 C 上。在 $t + dt$ 时刻在附近将会有一条亮度相同的等亮度线 C' 。但 C 与 C' 上的点是如何对应的？因为通常这两条线的形状都不相同，所以这个问题难以回答。由此可知，根据时变图象 (changing image) 中的局部信息不能唯一地确定光流。

设，在 t 时刻图象点 (x, y) 处的辐照度是 $E(x, y, t)$ 。如该点光流在 x 和 y 轴的分量分别为 $u(x, y)$ 和 $v(x, y)$ ，那么在 $t + dt$ 时刻相同亮度的点将会在 $(x + dx, y + dy)$ 。其中 $dx = udt$ 和 $dy = vdt$ ，即对小的时间间隔可有：

$$E(x + udt, y + vdt, t + dt) = E(x, y, t)$$

单有这一个约束不足以唯一地确定 u 和 v 。这里显然我们也可以利用各处的运动场是连续的这个事实。

如果亮度随 x, y 和 t 平滑变化，可把上式按泰勒级数展开得到：

$$E(x, y, t) + dx \frac{\partial E}{\partial x} + dy \frac{\partial E}{\partial y} + dt \frac{\partial E}{\partial t} + e = E(x, y, t)$$

其中 e 包括在 dx ， dy 和 dt 中的二次以上的项。上式中约去 $E(x, y, t)$ ，并用 dt 除等式两端和取 $dt \rightarrow 0$ 的极限后可求得：

$$\frac{\partial E}{\partial x} \frac{dx}{dt} + \frac{\partial E}{\partial y} \frac{dy}{dt} + \frac{\partial E}{\partial t} = 0 \quad (7-1)$$

此式实际上是等式 $\frac{dE}{dt} = 0$ 的展开形式。用以下形式简写：

$$u = \frac{dx}{dt}, \quad v = \frac{dy}{dt}$$

$$E_x = \frac{\nabla E}{\nabla x}, \quad E_y = \frac{\nabla E}{\nabla y}, \quad E_t = \frac{\nabla E}{\nabla t}$$

上式可写成:

$$E_x u + E_y v + E_t = 0 \quad (7-2)$$

根据图象可估算 E_x , E_y 和 E_t 等项导数。上述方程被称为光流的约束方程 (optical constraint equation), 因为此方程表示了对光流的 u 和 v 分量的约束。

我们可把以 u 和 v 为轴的二维空间称为速度空间 (velocity space) (图 7.8)。满足约

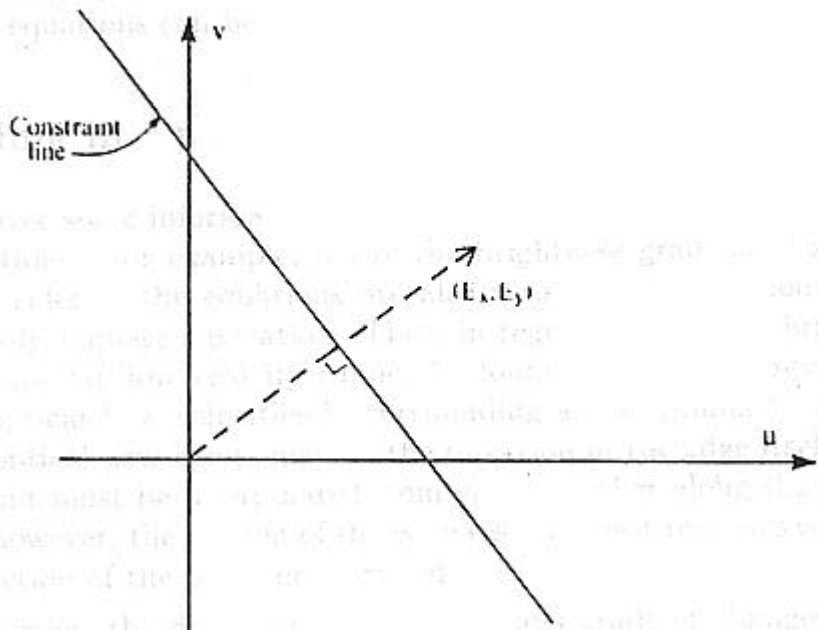


图 7.8 光流速度与亮度梯度方向垂直

束方程的 (u, v) 值都在速度空间的直线上。但局部的测量无法辨别实际的 (u, v) 是在约束线上的那一点。约束方程可改写为以下形式:

$$(E_x, E_y)(u, v) = -E_t$$

光流在亮度梯度 $(E_x, E_y)^T$ 方向上的分量为

$$\frac{E_t}{\sqrt{E_x^2 + E_y^2}}$$

但我们不能确定光流在与梯度垂直方向 (即沿等亮度线) 上的分量。这样的多义性问题被称为孔径问题 (aperture problem)。

为理解光流约束方程不能对速度向量 (u, v) 提供足够约束的物理意义, 我们可作以下进一步的讨论。

如果 E_x 和 E_y 是零 (或接近于零), 那么约束方程不能给我们任何关于 u 和 v 的信息。这相当于直观的观察, 在灰度是常数或变化缓慢的地方, 像素的运动不能精确地确定。如果只是其中的一个, 例如 E_x 是零, 那么约束方程变为 $E_y u + E_t = 0$, 所以可求得 $v = -E_t/E_y$, 而 u 不知道。这相应于在边缘处的直观观察, 譬如说边缘沿水平方向 ($E_x \approx 0, E_y \neq 0$), 这时我们可以确定沿垂直于边缘方向的运动, 但不能确定沿边缘方向的运动。一般情况下, 设 g 和 l 分别表示在约定像素处的亮度梯度和亮度不变的方向, 即 $g = \tan^{-1}(E_y/E_x)$ 是灰度变化最快的方向; l 是与此垂直的方向, 在 l 方向灰度不变。

然后，如果我们用 g 和 l 代替 x 和 y 作为约束方程中的空间坐标，我们得到

$$E_t = -\frac{\nabla E}{\nabla g} \cdot \frac{dg}{dt}, \text{ 所以我们能确定运动在梯度方向 } \left(-\frac{\nabla E}{\nabla t} / \frac{\nabla E}{\nabla g} \right) \text{ 的分量，但不能确定}$$

灰度不变方向的运动。也即可确定运动垂直边缘的分量，但不能确定沿边缘方向的运动。

因为约束方程只对速度向量 (u, v) ，给出一个约束，如果我们要根据它来估计运动，我们必须施加附加的约束。一个方法是假设速度从一个像素到另一个像素的变化是平滑的。所以变化的速率（用梯度的幅度或拉普拉斯运算来测量）很小。因此我们可以用迭代的方法发现满足约束方程和使变化速率最小的速度场。这个关于变化率最小的假设在一个物体内部是合理的，因为在物体内部旋转和比例变化使像素间的速度变化很小，但这样的假设在物体边界处是不正确的。

另一种利用约束方程估计速度的方法是观察拐角的速度（一个突然转弯的边缘），可提供关于速度在两个方向分量的信息（横过两个边缘相交在角的方向），所以在拐角处的速度被完全地确定。然后，我们可以沿边缘传播速度估计（再次应用速度从一个像素到另一个像素的变化很慢的假设），以便在整个物体边界得到跟角的速度相一致的速度场。如果希望的话，我们可把这些速度传播到物体内部，请注意，这个方法要求明确地检测边缘，所以可称为基于边缘的方法。

如果我们已知同一刚体的一小组像素的速度向量，就可以完全地确定物体的三维运动。已经证明 [Tai 84] 根据在给定时间 5 个像素的速度或两个连续时间四个像素的速度就足以确定物体的运动。当然，如果是某种特殊类型的运动，只要更少的像素就足够了。例如，如果运动是二维平移，物体上每个像素的速度向量都相同，我们只需知道一个像素的速度就够了。

当用邻域匹配，或用空间和时间的导数估计在个别像素的速度时，通常估计值的噪声很大，因为这估计值是基于少量的信息。在某个特定物体的运动是二维平移，属于同一物体像素将有相同的速度。因此，聚类检测可用来检测是否存在这样的物体，并能用平均方法更精确地估计速度。

3. 光流的估算^[Hor 81]

如前所述为估算光流除了光流约束方程以外还需要附加的约束。如果假设视场内的物体是刚体，那么要求解的关键参数是摄像机相对物体的平移和旋转，刚体运动假设是很严格的，它可对求解提供很有用的约束。但在这里我们希望较为有限度的假设，以使我们能处理较为一般的情况，如可变形的塑料物体。

通常图象中的大部分区域中光流是平滑变化的。因此，我们可以利用平滑性约束。在离散的情况下，可利用在网格点 (i, j) 及其领域的光流值来测量对平滑性的偏离

$$s_{ij} = \frac{1}{4} \left[\left(u_{i+1, j} - u_{ij} \right)^2 + \left(u_{i, j+1} - u_{ij} \right)^2 + \left(v_{i+1, j} - v_{ij} \right)^2 + \left(v_{i, j+1} - v_{ij} \right)^2 \right] \quad (7-3)$$

在光流约束方程中的误差 e_{ij} 等于

$$e_{ij} = \left(E_x u_{ij} + E_y v_{ij} + E_t \right)^2$$

其中 E_x, E_y 和 E_t 分别是点 (i, j) 处，对亮度随 x, y 和 t 变化速率的估算值。我要寻找一组使以下分式中的 e 最小的 $\{u_{ij}\}$ 和 $\{v_{ij}\}$ 的值。

$$e = \sum_i \sum_j \left(s_{ij} + \lambda e_{ij} \right) \quad (7-4)$$

使 e 对 u_{kl} 和 v_{kl} 求导，得

$$\frac{\partial e}{\partial u_{kl}} = 2(u_{kl} - \bar{u}_{kl}) + 2\lambda(E_x u_{kl} + E_y v_{kl} + E_t)E_x$$

$$\frac{\partial e}{\partial v_{kl}} = 2(v_{kl} - \bar{v}_{kl}) + 2\lambda(E_x v_{kl} + E_y v_{kl} + E_t)E_y$$

其中 \bar{u} 和 \bar{v} 是 u 和 v 的局部平均值。上述 e 的导数为零的地方出现极值。所得方程经过整理后的形式为

$$\begin{aligned} (1 + \lambda E_x^2)u_{kl} + \lambda E_x E_y v_{kl} &= \bar{u}_{kl} - \lambda E_x E_t \\ \lambda E_x E_y u_{kl} + (1 + \lambda E_y^2)v_{kl} &= \bar{v}_{kl} - \lambda E_y E_t \end{aligned} \quad (7-5)$$

我们可把上述方程看作是 u_{kl} 和 v_{kl} 的方程对， 2×2 的系数矩阵的行列式为

$$1 + \mathbf{I}(E_x^2 + E_y^2)$$

所以

$$\begin{aligned} [1 + \lambda(E_x^2 + E_y^2)]u_{kl} &= (1 + \lambda E_y^2)\bar{u}_{kl} - \lambda E_x E_y \bar{v}_{kl} - \lambda E_x E_t \\ [1 + \mathbf{I}(E_x^2 + E_y^2)]v_{kl} &= -\mathbf{I}E_x E_y \bar{u}_{kl} + (1 + \mathbf{I}E_x^2)\bar{v}_{kl} - \mathbf{I}E_y E_t \end{aligned}$$

现在可从上述方程组求解 u_{kl} 和 v_{kl} ，与前述章中求解表面方向的松弛法相似，所得结果可理解成一个迭代的过程

$$\begin{aligned} v_{kl}^{n+1} &= \bar{v}_{kl}^n - \frac{E_x \bar{u}_{kl}^n + E_y \bar{v}_{kl}^n + E_t}{1 + \mathbf{I}(E_x^2 + E_y^2)} \mathbf{I}E_y \\ u_{kl}^{n+1} &= \bar{u}_{kl}^n - \frac{E_x \bar{u}_{kl}^n + E_y \bar{v}_{kl}^n + E_t}{1 + \mathbf{I}(E_x^2 + E_y^2)} \mathbf{I}E_x \end{aligned} \quad (7-6)$$

对上式可作以下几何解释：在点 (u, v) 处的新的值被置成等于周围值的平均值减去一个速度空间的调整量，调整是沿着梯度的方向（见图 7.9），这是因为

$$\frac{u_{kl}^{n+1} - \bar{u}_{kl}^n}{v_{kl}^{n+1} - \bar{v}_{kl}^n} = \frac{E_x}{E_y}$$

所以是沿斜率为 E_x/E_y 的直线修正。

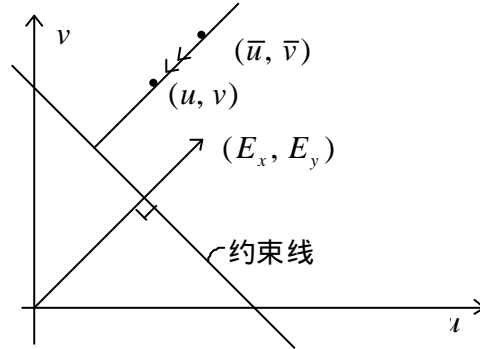


图 7.9 按迭代方法估算光流

光流速度位于与亮度梯度方向垂直的直线上，由此只能确定在亮度梯度方向上的速度分量，也就是说沿斜率为 E_x/E_y 的直线进行。调整使新的计算值不断靠近约束线。

为实现上述方法还需要估算亮度的空间和时间导数。利用网格上的一阶差分可容易地求得这些导数（图 7.10）。如果标号 i, j 和 k 相应于 x, y 和 t ，那么三个一阶偏导数的估算分别如下所示：

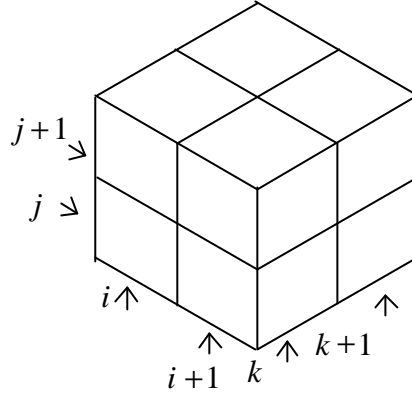


图 7.10 一阶导数估计

$$\begin{aligned}
 E_x &\approx \frac{1}{4\delta x} (E_{i+1,j,k} + E_{i+1,j,k+1} + E_{i+1,j+1,k} + E_{i+1,j+1,k+1}) \\
 &\quad - \frac{1}{4\delta x} (E_{i,j,k} + E_{i,j,k+1} + E_{i,j+1,k} + E_{i,j+1,k+1}) \\
 E_y &\approx \frac{1}{4\delta y} (E_{i,j+1,k} + E_{i,j+1,k+1} + E_{i+1,j+1,k} + E_{i+1,j+1,k+1}) \\
 &\quad - \frac{1}{4\delta y} (E_{i,j,k} + E_{i,j,k+1} + E_{i+1,j,k} + E_{i+1,j,k+1}) \\
 E_t &\approx \frac{1}{4\delta t} (E_{i,j,k+1} + E_{i,j+1,k+1} + E_{i+1,j,k+1} + E_{i+1,j+1,k+1}) \\
 &\quad - \frac{1}{4\delta t} (E_{i,j,k} + E_{i,j+1,k} + E_{i+1,j,k} + E_{i+1,j+1,k}) \quad (7-7)
 \end{aligned}$$

上述公式表示所估算的是相邻两幅图象之间，并位于像素中间点的光流速度。

图 7.11 所示为表面上具有平滑变化的亮度图案的旋转圆球的图象，四幅图象是相互接续的。根据这些图象估算出的亮度的空间和时间导数作为上述光流估算算法的输入。

经过 1, 4, 16 和 64 次迭代后估算的光流如图 7.12 所示。第一次估算受球面上的亮度图案影响很大。经过几次迭代以后，除了边缘处以外，估算的光流向正确的解收敛。

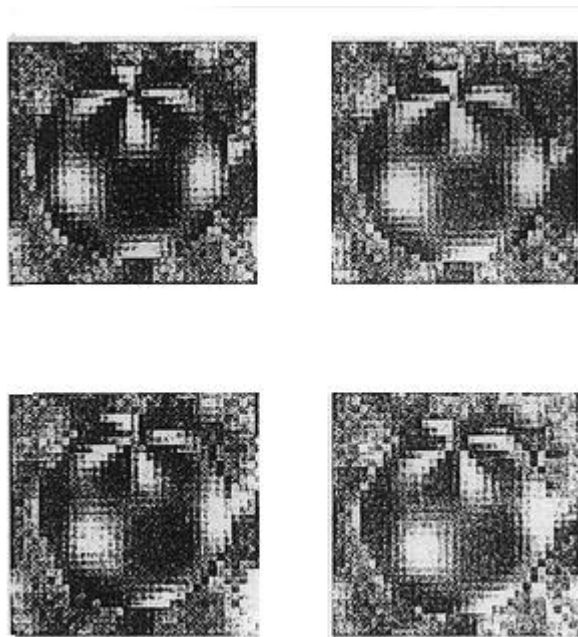


图 7.11 四帧合成图象序列，表示在随机图象背景前有一个球面在缓慢旋转
上面两幅为(a)和(b)，下面两幅为(a)和(b)

光流中的不连续点

当一个物体遮挡另一个物体时，在物体的边缘处光流将出现不连续点。如果我们想要避免在采用上述方法时试图把解从一个区域平滑地延伸到另一个区域的话必须检测出这些不连续点。这似乎是一个先有鸡还是先有蛋的问题：如果我们已对光流作了准确的估算，就可寻找光流急剧变化的地方，以便分割图象；另一方面，如果我们能正确的分割图象，就能更准确地估算光流。解决这个难题的方法是把分割与光流的迭代求解结合起来。这就是说，在每次迭代以后，我们寻找光流急剧变化的地方。在这些地方设立标记以禁止下次迭代时把解平滑的越过不连续点进行连接。为不至于过早地把图象切切割开，开始时把作出这样决策的阈值定得很高。当对光流可作较为准确的估算时逐渐降低阈值。

以下我们研究一下在实际情况下平滑性假设满足的程度。在物体作平移和旋转刚体运动的一般情况下。为简化起见我们讨论垂直投影的情况。平移导致与运动物体对应的图象区域中光流是均匀的。在这样区域中的光流是平滑的，而在区域的边界处（边界处物体遮挡背景）将会有不连续点。

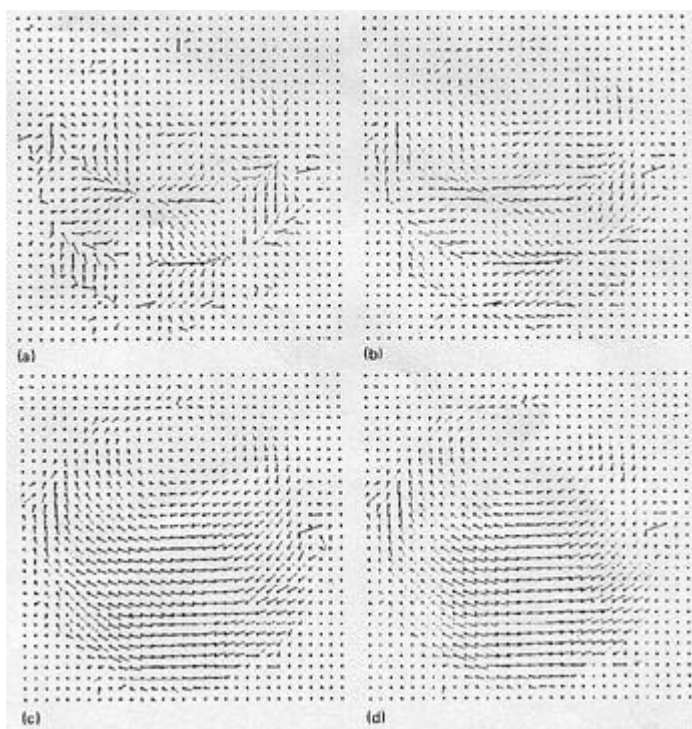


图 7.12 在算法经过 1,4,16 和 64 次迭代后，以指针图表示的光流估计

7.2.2. 基于过零点的方法

以上介绍的梯度法是运动测定的基本方法。从原则上讲，只要有亮度，运动就可以测量。然而在人的视觉系统中固有运动测量只有在强度有明显变化的位置才有可能。为了检测这些变化，Marr 和 Hildreth^[Hil 84]提出在与 $\nabla^2 G$ 作卷积的图象中检测过零点（见第三章第一节），利用滤波图象的空间、时间梯度机理可知过零点位置处的运动。

在二维的情况下，运动测量面临孔径问题。对边界而言，局部运动测量只检测出垂直于边界方向上的运动分量，而沿着边界方向的运动分量仍然没有被检测出来。沿边界的速度场可以用向量函数 $V(s)$ 表示，其中 s 表示弧长。如图所示 $V(s)$ 可被分解为切线方向和垂直方向两个分量， $u^t(s)$ 和 $u^\perp(s)$ 分别是切线方向和垂直于曲线方向的单位向量， $v^t(s)$ 和 $v^\perp(s)$ 分别代表两个分量：

$$v(s) = v^t(s)u^t(s) + v^\perp(s)u^\perp(s) \quad (7-8)$$

分量 $v^\perp(s)$ 及方向向量 $u^t(s)$ 和 $u^\perp(s)$ 可直接从变化的图象由初始测量得到。而分量 $v^t(s)$ 则不能直接得到，必须通过计算 $V(s)$ 来获得。可见，通过垂直分量 $v^\perp(s)$ 沿整个边界的测量集必然给运动边界提供相当多的约束。然而为唯一地确定运动，还要求有其它约束。可能的附加约束有以下几种：

(1) 速度在整个图象区域中是常数。在正交投影的情况下，只有当物体作简单位移的情况下，匀速的约束才是严格满足的。在其它的情况下都是近似的。

(2) 在平面中的刚体运动。

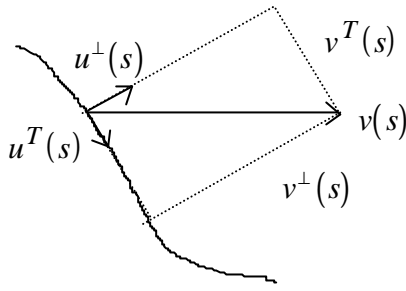


图 7.13 速度的分解， $u^\perp(s)$ 和 $u^\tau(s)$ 分别为与曲线垂直和相切的单位方向向量， $v^\perp(s)$ 和 v^τ 为相应的速率分量

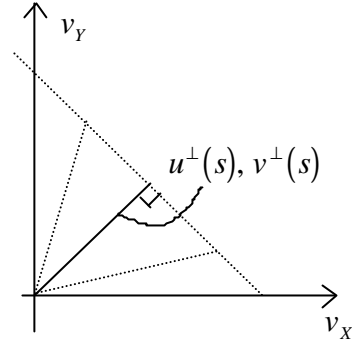


图 7.14 速率空间中的速率约束 $v^\perp(s)$ 和 $u^\perp(s)$ 分别为曲线上 p 点处的速率垂直分量和单位向量

假设刚体曲线在空间作一般运动。它的瞬时运动可看成是以下两种运动的结合：(1) 在空间绕某一轴以角速度 \mathbf{w} 旋转。空间轴的方向可用单位向量 $\bar{n} = [n_x, n_y, n_z]^\tau$ 表示；(2) 用向量 $\bar{t} = [t_x, t_y, t_z]^t$ 表示的位移。曲线 $c = [x(s), y(s), z(s)]$ ，其中 s 是弧长。点在曲线上的位置可用位置向量 $\bar{r} = [x(s), y(s), z(s)]^t$ 给出。如果设旋转轴通过坐标系统的原点。由位置向量 \bar{r} 给出的某点速度就等于 \bar{r} 与向量 $\bar{\mathbf{w}}_n$ 的叉积。如果光轴沿 z 轴方向（ z 轴的正方向指向观察者），同时令曲线在图象平面（ x, y 平面）上作垂直投影。那么沿边界的二维速度场 $v(s)$ 为：

$$v(s) = M(\bar{r} \times \bar{\omega} \bar{n} + \bar{t}) = -\omega_z(s) \begin{bmatrix} n_y \\ -n_x \end{bmatrix} - \omega_n z \begin{bmatrix} -y(s) \\ x(s) \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (7-9)$$

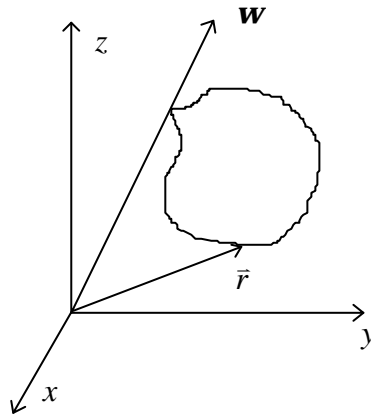


图 7.15

其中 M 表示作垂直投影的变换矩阵，后面第一项说明由于绕平行于图象的轴（ $\bar{n} = [n_x, n_y, 0]^t$ ）旋转的速度分量；第二项表示在图象平面内的分量（绕轴 $\bar{n} = [0, 0, n_z]^t$ 旋转）；第三项为平移。考虑到在图象平面内仅限于刚体运动。对应于平移和绕轴 $\bar{n} = [0, 0, 1]^t$ 旋转的速度场 $v(s)$ 可用下式表示：

$$v(s) = -\omega \begin{bmatrix} -y(s) \\ x(s) \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (7-15)$$

注:

$$\begin{aligned} \vec{r} \times \vec{\omega}_n &= \omega(\vec{r} \times \vec{n}) \\ &= \omega \left\{ (x(s)\vec{i} + y(s)\vec{j} + z(s)\vec{k}) \times (n_x\vec{i} + n_y\vec{j} + n_z\vec{k}) \right\} \\ &= \omega \left\{ (y(s)n_z - z(s)n_y)\vec{i} + (z(s)n_x - x(s)n_z)\vec{j} + (x(s)n_y - y(s)n_x)\vec{k} \right\} \end{aligned}$$

如只取 $\vec{r} \times \vec{\omega}_n$ 在 $x-y$ 平面上的垂直投影, 即只取上式中 \vec{i}, \vec{j} 方向的分量, 则可得

$$= -\omega \begin{bmatrix} n_y \\ -n_x \end{bmatrix} - \omega n_z \begin{bmatrix} -y(s) \\ x(s) \end{bmatrix}$$

$v(s)$ 是简单的平移加旋转。图象曲线 $x(s), y(s)$ 如图 7.16 所示。图 7.16(a) 中表示曲线 C_1 在图象平面作小的旋转和平移得到曲线 C_2 , 箭头所指为沿曲线的局部速度向量的取样。图 7.16(b) 中这些速度向量已经转换到速度空间的公共原点。在速度空间的曲线形状与图象曲

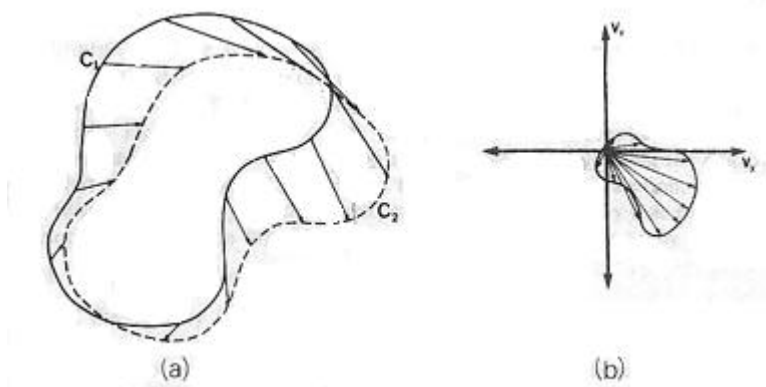


图 7.16 图象平面内刚体运动的速度场

- (a) 图象曲线 C_1 在图象内作小的旋转和平移产生曲线 C_2 ,
(b) 速度空间中的速度向量

线 C_1 相同, 其大小与角速度成正比并相对于 C_1 而言旋转了 90° 。曲线在图象中附加位移, 会在速度空间产生一个向量。在曲线离散运动的情况下, 图象曲线形状与其位移场存在类似关系, 这个位移场是描述曲线上的点作离散位移的向量集。用简单的几何方法可以计算图象平面中刚体运动的速度场。如果已知边界上两点的速度方向, 则可计算出整个速度场, 方法如下:

- ① 在两个已知点作与速度方向垂直的线;
- ② 计算这两点的交点;
- ③ 沿边界上各点向交点连线, 该点速度方向垂直于此连线。

如图 7.17 所示, 已知 p_1 和 p_2 处的速度的方向, 则可求出 p_3 处的速度方向。这个方法就是简单地确定这样一个点的位置, 在该点的运动可看成是单纯的旋转。对于简单位移来说, 从已知速度方向的点出发的两条线是平行的。所以各处的运动方向都与已知点相同。当然, 如果已知边界上两点的速度大小和方向, 则整体运动参数和各处速度的大小和方向都可以计算出来。此外, 如果已知边界上两点的速度方向, 各处的速度方向都可以求出来。此外, 如果速度的垂直分量已知, 那么沿边界速度的大小和方向都可计算出来。

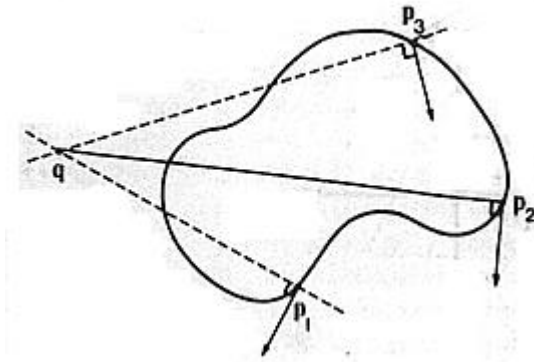


图 7.17 速度场的建立

上述方法仅适用于图象中平面内刚体运动的速度场计算。对一般运动分析是不够的。

(3) 平滑性约束

平滑性约束可适用于表面在三维空间自由运动的情况。假设现实世界主要由固态物体组成，其表面相对于它们与观察者之间的距离而定是平滑的。平滑表面的运动通常产生平滑变化的速度场。因此，我们要寻求一个与运动测量相一致的速度场，而这个运动测量是从平滑变化的图象中得到的。遗憾的是有无数速度场可满足这两种特性。Horn 和 Schunck 在光流计算中提出通过寻找变化尽可能小的速度场可以求得唯一解。为此先要解决测量速度沿边界变化的方法问题。有许多测量方法，可沿边界测试速度方向的变化，或速度大小的变化，或综合速度的方向和大小，测量速度向量的变化，从数学上检查这些测量，可得出以下结论：为计算速度场，所有的速度向量的变化都应该考虑进去，使用只考虑速度大小或方向的函数通常不能得到速度场的唯一解。但是可以证明，如果在从图象中导出的约束上给定一个简单的条件，就存在着一个唯一解满足这些约束，并使特定的速度变化测量

具有最小值。这样的速度变化测量是 $\int \left| \frac{\nabla v}{\nabla s} \right|^2 ds$ ，积分域是沿着封闭的边界。

7.2.3 基于特征的运动检测方法

如前所述在根据运动物体的图象序列恢复物体的三维结构和空间运动时通常采取基于特征的方法。这时首先要从图象中检测出由物体上的顶点和边缘通过投影在图象平面产生的点和线。然后在图象序列中确定各帧图象中特征之间的对应性，这样根据这些特征的对应关系以及它们的二维图象中的位置就可以计算出这些特征点的三维结构和空间运动。如果这些特征点属于刚体，那么也就可以确定刚体的空间运动，这种根据二维图象序列恢复三维运动的方法在自主车导航、机器人视觉导引、利用视频图象的目标跟踪和生产过程的动态监控中广泛应用。

在立体视觉中，人们根据左、右图对应点视差计算物体点的深度。当在左、右图中搜索对应点时可以利用外极线约束。因此，与外极线成较大角度的边缘线段可以用作匹配特征。与此不同，在运动图象序列中搜索特征的对应时，由于物体可能在三维空间中作任意运动，所以没有外极线的约束。这时短的边缘线段就不适合作为特征，而要应用具有特殊性质的点^{[Moc 77, 84][Ban 84][Tho 84]}。例如角点^[Fan 82]，一定长度的线条^[Kor 78]，或小块^[lolob]。

以下简要介绍 Moravec 的利益算子 (Interest Operator)

Moravec 的利益算子通过在指定点周围小区域内 (约为 5×5 象素) 沿四个方向 (水平、垂直、两个对角线) 计算灰度差异，并选择四个方向中的最小差异值的方法来测量这个点的利益值。各个方向上的差异是通过计算指定点周围沿此方向上象素灰度值差的平方和来求得的。例如，用下列一小段程序可计算沿水平方向的差异。

```
for (r=minrow to maxrow)
  for (c=mincol to maxcol - 1)
    Hvar += (img[r, c]-img[r, c+1])**2
```

用相似的程序可求得垂直方向和两个对角方向上的差异。这里利用差异的原因是在无特点的区域一般灰度变化小，所以利益值就低；选用沿四个方向中差异最小值作为利益值的原因是因为对位于直线上的点，虽然在垂直于直线的方向上灰度差异大，但沿直线的方向上差异很小，所以它的利益值仍然较小，而只有位于角上的点或在纹理区域中的点才能在所有的方向都有大的差异值。在一个高度纹理区域内有可能得到一堆有高利益值的点，这是我们不希望的。为此可用一个局部极大值滤波器来把每个指定点与其 24 个最近邻点作比较，然后选择其中的极大点。

一般利益算子只应用于分辨率金字塔中某一分辨率已降低了的图象，通常是次最高的分辨图象，而不是就用于原图。这样做的部分原因是为了降低计算负担，加速计算速度，同时也是为了可通过降低分辨率时的平均运算来平滑高频噪声。

Moravec 利益算子由于其良好性能被广泛地采用，但它也还存在着一些缺陷。例如，在某些情况下仍会偶然地选择沿某个不在上述四个方向上的边缘点等。针对这些缺点 Thorpe^[Tho 84]提出了新的利益算子。这个算子中包括梯度算子、像素内插器和差异计算器等三部分，取得较好效果。

除了以上利益算子以外还有 Kitchen-Rosenfeld^[Kit 82]的角点检测算子，它通过测量拟合表面的转弯来检测角点。Nagel^[Nag 82]提出基于二阶灰度差异的利益算子，此外还有基于平面拟合和自相关的利益算子等。

7.3 运动理解

与基于灰度和基于特征的运动检测方法相对应，也存在着基于光流和基于特征的运动图象理解方法。这里的理解是指从二维的图象序列获得关于物体的三维结构和三维空间运动信息。

7.3.1 光流的理解

这里以物体的平移运动为例来说明光流的理解，在物体作平移运动时存在着所谓的膨胀中心（Focus of Expansion）现象。

(1) 膨胀中心 FOE

当人们经过静止景物时，景物在视网膜上的投影似乎在向后流动。事实上，对给定的平移运动方向和给定的注视方向来说，景物的视觉投影似乎是从一个特定的视网膜点流出来的，这个点就是膨胀中心（FOE）。对每个运动方向和注视方向要引入一个单独的 FOE。如果运动与视网膜（图象）平面平行，FOE 点可能是无穷远点。图 7.18 所示为 FOE 的一些例子，在图 7.18(c)中有第二个 FOE 这是由于在视场中含有自身也在运动的物体所造成的。为简化起见设观察点在原点，观察方向沿 z 轴的正方向，并设焦距 $f = 1$ ，因此透

视变换为 $x' = \frac{x}{z}$ ， $y' = \frac{y}{z}$ 。点在三维空间的坐标为 (x, y, z) ，相应的速度分量为 (u, v, w) 。在某个初始时刻某点位于 (x_0, y_0, z_0) ，经过时间间隔 t 以后，其像点将位于 (x', y') ，其中

$$(x', y') = \left[\frac{x_0 + ut}{z_0 + wt}, \frac{y_0 + vt}{z_0 + wt} \right] \quad (7-11)$$

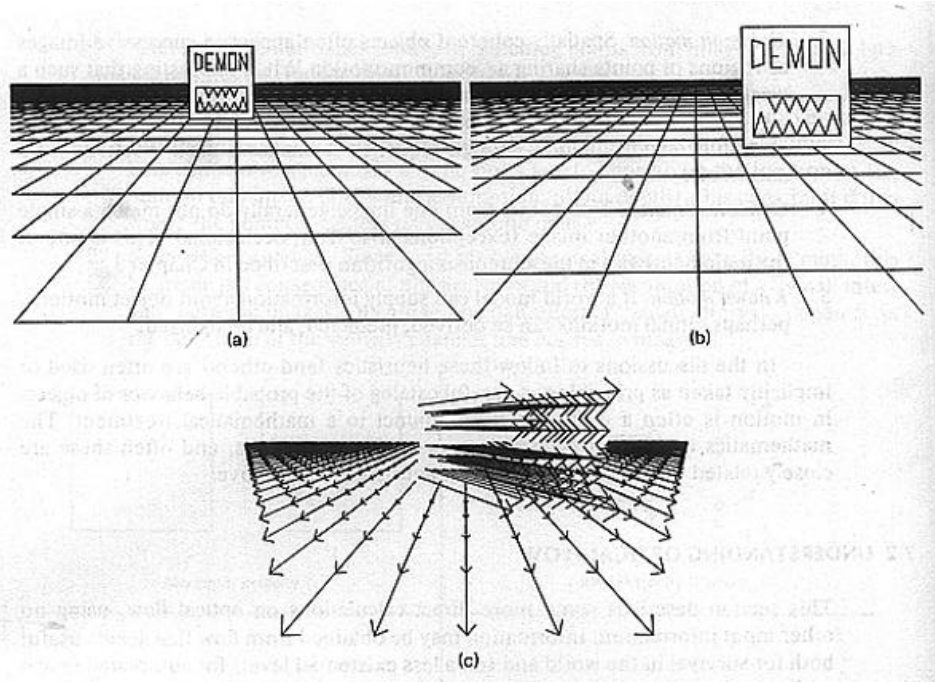


图 7.18 直线平称观察者运动的 FOE

(a)前一时刻的图象,(b)后一时刻的图象,(c)静止的地面和运动物体各有 FOE
 当时间 t 变化时, 光流路径方程是直线 (flow-path)

$$\Delta x' = x' - \frac{x_0}{z_0} = \frac{x_0 + ut}{z_0 + wt} - \frac{x_0}{z_0} = \frac{t(z_0 u - x_0 w)}{z_0(z_0 - wt)}$$

$$\Delta y' = y' - \frac{y_0}{z_0} = \frac{y_0 + vt}{z_0 + wt} - \frac{y_0}{z_0} = \frac{t(z_0 v - y_0 w)}{z_0(z_0 - wt)}$$

$$\frac{\Delta y'}{\Delta x'} = \frac{z_0 v - y_0 w}{z_0 u - x_0 w} = \text{Constan } t$$

当 t 趋向无穷大时各点的图象沿此直线朝图象中的特定点缩回去, 就个特定点就是 FOE

$$FOE = \left[\frac{u}{w}, \frac{v}{w} \right]$$

这个膨胀中心就是图象上光流的出发处。如果观察方向改变 (或物体的运动方向改变) FOE 也随之改变。

(2) 基于 FOE 的深度估计和碰撞检测

以固定速度运动的点的光流路径可提供该点在 Z 方向上的深度信息, 但深度信息并不是直接提供的, 因为在给定深度上各点的光流路径看起并不相同, 但存在以下关系:

$$\frac{D(t)}{V(t)} = \frac{Z(t)}{W(t)} \quad (7-12)$$

其中 $W = \frac{dZ}{dt}$, $V = \frac{dD}{dt}$, D 是从 FOE 沿光流路径到空间点对应像点的距离, 此式的含意是图象点的距离 / 速度比与空间点的距离 / 速度比相等。这个结论很重要, 但它的意义似乎还不明显。此式的证明为:

$$D = \sqrt{\left[\left(\frac{x_0 + ut}{z_0 + wt} \right) - \frac{u}{w} \right]^2 + \left[\left(\frac{y_0 + vt}{z_0 + wt} \right) - \frac{v}{w} \right]^2}$$

$$\frac{dD}{dt} = \frac{1}{2} \frac{2 \left(\frac{x_0 + ut}{z_0 + wt} - \frac{u}{w} \right) \cdot \frac{d}{dt} \left(\frac{x_0 + ut}{z_0 + wt} \right) + 2 \left(\frac{y_0 + vt}{z_0 + wt} - \frac{v}{w} \right) \cdot \frac{d}{dt} \left(\frac{y_0 + vt}{z_0 + wt} \right)}{\sqrt{\left[\left(\frac{x_0 + ut}{z_0 + wt} - \frac{u}{w} \right)^2 + \left(\frac{y_0 + vt}{z_0 + wt} - \frac{v}{w} \right)^2 \right]}}$$

因为 $\frac{d}{dt} \left(\frac{x_0 + ut}{z_0 + wt} \right) = \frac{u}{z_0 + wt} - \frac{(x_0 + ut)w}{(z_0 + wt)^2} = \frac{uz_0 - x_0 w}{(z_0 + wt)^2}$

$$\frac{x_0 + ut}{z_0 + wt} - \frac{u}{w} = \frac{(x_0 + ut)w}{(z_0 + wt)^2}$$

所以 $\frac{dD}{dt} = -\frac{1}{D} \left[\frac{(x_0 w + z_0 u)^2}{w(z_0 + wt)^3} + \frac{(y_0 w + z_0 v)^2}{w(z_0 + wt)^3} \right]$

$$\begin{aligned} \frac{D(t)}{\frac{dD}{dt}} &= \frac{D(t)}{V(t)} = -\frac{\left(\frac{x_0 + ut}{z_0 + wt} - \frac{u}{w} \right)^2 + \left(\frac{y_0 + vt}{z_0 + wt} - \frac{v}{w} \right)^2}{\frac{1}{w(z_0 + wt)^3} [(x_0 w - z_0 u)^2 + (y_0 w - z_0 v)^2]} \\ &= \frac{z_0 + wt}{w(t)} = \frac{z(t)}{w(t)} \end{aligned}$$

上式被称为“接近时间”关系。因为此式的右边是 z/w ，也就是空间点离图象平面的距离 z 被它朝图象平面运动的速度除。这个关系一直保持到此点穿过图象平面为止。当涉及空间中的物体时，这个基本的时间间隔显然是很有用的。当空间点的速度改变时，这个时间间隔也随之改变。

若已知任何点的深度就可确定所有其它的速度为 w 的点的深度，这结论可从两个点接近方程（adjacency equations）得出

$$Z_2(t) = \frac{Z_1(t)D_2(t)V_1(t)}{V_2(t)D_1(t)}$$

利用接近时间方程可方便地确定点在空间中的坐标，但这些坐标与点沿 z 轴的移动速度成正比。如果观察者本身是可移动的，并自行控制运动速度，以及如果景物是静止的，那么这样带比例系数的坐标是有用的。利用透视投影方程：

$$\begin{aligned} Z(t) &= \frac{w(t)D(t)}{V(t)} \\ Y(t) &= \frac{Y'(t)w(t)D(t)}{V(t)} \\ X(t) &= \frac{X'(t)w(t)D(t)}{V(t)} \end{aligned}$$

最后一个例子是把光流用于探测观察者与景物中物体可能即将出现的碰撞。成象系统的焦点（focal point），或坐标的原点在任何瞬间都对着膨胀点的朝向，膨胀点的坐标是 $(u/w, v/w)$ 。因此，观察者是沿以下方向运动

$$O = \left(\frac{u}{w}, \frac{v}{w}, 1 \right)$$

并且在任何时刻都沿着场景中以下述参数方程定义的路径

$$(x, y, z) = tO = t\left(\frac{u}{w}, \frac{v}{w}, 1\right)$$

其中 t 的作用就好像是时间是量度。已知观察者路径的矢量表达式，我们可根据立体解析几何中大家熟知的公式得到有关观察者的路径与场景中空间点关系的信息，空间点的位置也可用矢量表示。

例如，在观察者路径上与指定空间点最靠近的点 p 的位置是

$$\bar{P} = \bar{O}_n \frac{(\bar{O} \cdot \bar{X})}{|\bar{O}|} = \frac{\bar{O}(\bar{O} \cdot \bar{X})}{|\bar{O}|^2} P = \frac{O(O \cdot X)}{(O \cdot O)}$$

其中 O 是观察者运动的方向， \bar{O}_n 是该方向上的单位向量， X 是空间点的位置，观察者在最靠近处与空间点的距离平方 Q^2 等于

$$Q^2 = (X \cdot X) - (X \cdot O)^2 / (O \cdot O)$$

7.3.2 基于特征的运动理解

基于特征的运动理解就是根据图象序列中二维特征的对应关系恢复物体的三维结构和空间运动。这在目标跟踪、机器人的视觉导引和动态过程监控等方面有广泛应用。例如：

(1) 对象跟踪 (tracking)。有两种跟踪：背景跟踪和目标跟踪。

背景跟踪在景物的背景中识别物体，例如房子，并在摄像机移动时监视这个物体。Landsat 卫星利用地面特征来精确地确定卫星的位置。

目标跟踪的例子是天气预报。在天气预报中根据云层的运动预测风的速度。另一个例子是工业零件处理，这时需要识别和定位传送带上的零件。

(2) 分析不同时刻拍摄的航空照片可发现城市发展的模式。为了得到人类机体中血液循环的定量信息可从一序列 X-ray 或 CAT 图象中抽取信息。

(3) 立体分析。运动可提供关于世界的立体本质。

以下介绍一种根据二维量测得到三维点结构的算法：

假设摄像机坐标系的原点在投影中心，并且 y 轴跟光轴对齐（图 7.19）。此外假设 (u, v, w) 坐标系跟物体运动固定。设 $\bar{A}m = (Amu, Amv, Amw)$ 是运动物体上的固定点。例如出租汽车顶部的标志在物体坐标系中的坐标矢量。

固定点在摄像机坐标系中的坐标 $\bar{C}m$ 可由下述转换得到：

$$\bar{C}m = (\bar{A}m + \bar{T})D \quad (7-16)$$

其中 $-\bar{T}$ 表示摄像机坐标系原点在物体坐标系中的坐标值，

D 表示为使二个坐标系的轴对齐所要求的旋转矩阵。

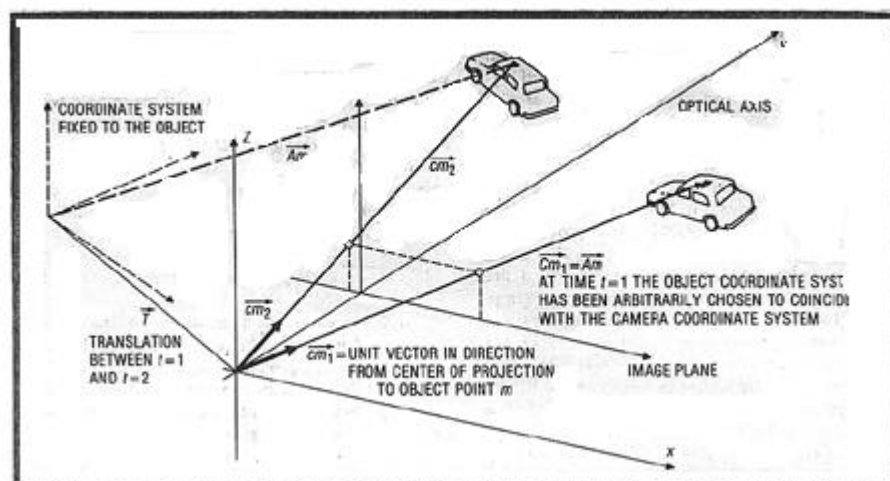


图 7.19 为从二维测量序列得到三维点结构时的摄像机坐标

我们可以把 $\bar{C}m$ 写成它的幅度 Cm 跟在摄象机坐标系的原点跟运动物体上的固定点 m 连线方向上的单位矢量 $\bar{C}mu$ 的乘积。因为物体坐标系随物体运动，位移矢量以及旋转矩阵 D 应是时间的函数，也就是图象帧数的函数。相似地，运动物体上每个固定点 m 的摄象机坐标系统的坐标也跟帧数有关。这可用加上帧标志 n 来表示：

$$\bar{C}mn = Cmn \cdot \bar{C}mun = (\bar{A}m + \bar{T}n)D_n \quad (7-17)$$

我们只能测量从投影中心（摄象机系统的原点）到固定点射线的方向 $\bar{C}mun$ ，例如，用把投影中心跟点 m 在第 n 帧中的成象相连接的方法来测量。然而，我们不知道在摄象机坐标系中到点 m 的矢量的长度 Cm 。因为我们假设物体是固体，固定点相互之间的位置，从而对所有帧的 $\bar{A}m$ 矢量在物体运动过程中是不变的。

无论我们选什么样的点作为物体坐标系的原点，都存在一个合适的矢量 \bar{T}_n 使(7-17)式成立，相似的，无论我们如何假设物体坐标的轴相对摄象机坐标系统的方向，都将存在一个合适的 D_n 使(7-17)成立。我们利用这样的选择自由，使对第 1 帧来说，这两个坐标系重合，也即

$$Cm_1 \cdot \bar{C}mu_1 = \bar{A}m \quad (7-18)$$

我们将首先集中精力研究两帧图象的情况。因此，所有未知的，但固定的 $\bar{A}m$ 将限于一个位移 \bar{T}_2 和一个旋转 D_2 ，这两个量反映第一帧与第二帧之间的位移和旋转。在这种特殊情况下，我们可以消掉 \bar{T} 和 D 的帧标志，写成

$$Cm_2 \cdot \bar{C}mu_2 = (\bar{A}m + \bar{T})D \quad (7-19)$$

在等式(7-19)的二边乘以旋转矩阵 D 的逆，得到

$$Cm_2 \cdot \bar{C}mu_2 D^{-1} = \bar{A}m + \bar{T}$$

或用(7-18)置换式中的 $\bar{A}m$ 得

$$\bar{T} = Cm_2 \bar{C}mu_2 D^{-1} - Cm_1 \bar{C}mu_1 \quad (7-20)$$

等式(7-20)的左边表示在二帧图象之间物体所做的位移。位移 T 跟固定点的标志 m 无关。图象帧间旋转的逆转矩 D^{-1} 使每个 m 点的单位矢量从第二帧中的位置旋转某个角度，旋转角度的大小是使帧间位移 \bar{T} 是反转单位矢量 $\bar{C}mu_2 D^{-1}$ 和 $\bar{C}mu_1$ 的线性组合。这些线性组合的系数 Cm_1 和 Cm_2 是不知道的，然而我们确实知道 \bar{T} 在由 $\bar{C}mu_1$ 和 $\bar{C}mu_2$ 两个矢量决定的平面上，因而与该平面的法线正交。由这二个矢量的叉积可求得这个平面的法线。这就是说对所有 m 都有下式

$$(\bar{C}m_1 \times \bar{C}m_2 D^{-1}) * \bar{T} = 0 \quad (7-21)$$

式中 \times 表示叉积， $*$ 表示点积。

这是因为位移矢量是在这个平面之内，因此必须垂直于平面的法线。

试考虑由二个不同的固定点，譬如说 $m=1$ 和 $m=2$ 的测量所定义的叉积。位移矢量被包含在由法线

$$(\bar{C}1_1 \times \bar{C}1_2 D^{-1}) * \bar{T} = 0 \quad (7-21a)$$

和同样地由法线

$$(\bar{C}2_1 \times \bar{C}2_2 D^{-1}) * \bar{T} = 0 \quad (7-21b)$$

定义的平面内。

因此，位移矢量 \bar{T} 平行于这二个平面的交线。因为交线必须同时垂直于这二条法线。这就是说

$$\bar{T} = S(\bar{C}1_1 \times \bar{C}1_2 D^{-1}) \times (\bar{C}2_1 \times \bar{C}2_2 D^{-1}) \quad (7-22)$$

式中 S 为某个未知的比例因子。把 m 等于 3 或大于 3 的等式(7-22)跟等式(7-21)组合起来，我们得到

$$(\bar{C}m_1 \times (\bar{C}m_2 D^{-1})) \times (\bar{C}1_1 \times (\bar{C}1_2 D^{-1})) \times (\bar{C}2_1 \times (\bar{C}2_2 D^{-1})) = 0 \quad (7-23)$$

方程(7-23)包含从三个固定点得到的量测和未知的旋转矩阵 D^{-1} 。三个参数可以完全规定一个旋转矩阵：这三个参数是旋转轴的方向（二个角）和绕旋转轴的转角。如果我们在两帧图中有 5 个点的量测。我们就能对 $m = 3, 4, 5$ 写出方程(7-23)，并对第一和第二帧之间物体旋转的未知数得到三个非线性方程。

假设已对这三个方程求解，我们可以把 $m = 1$ 时的方程式(7-20)代入方程(7-21)（7-23）并得到：

$$(\bar{C}_2 \times \bar{C}_2 D^{-1}) \cdot (C_1 \bar{C}_1 D^{-1} - C_1 \bar{C}_1) = 0$$

上式表示二个未知的幅度值 C_1 和 C_1 的标量方程。以下我们可以把 C_1 表示成正比于 C_1 ，也即把 C_1 作为由 $m = 1$ 时的方程(7.20)给出的位移矢量 \bar{T} 的比例因子。一旦确定了，所有其他的幅度因子 C_{m_1} 和 C_{m_2} 就可以根据 $m = 2, 3 \dots$ 时的方程(7-20) 计算得到。当知道由任意比例因子 C_1 决定的所有 C_{m_1} 时，我们就知道五个固定点相对于物体坐标系统的三维坐标矢量 \bar{A}_m ，这就如方程式(7-18)所示。

如果测量的点多于 5 个，我们可以用下述方法确定矩阵 D ：把方程式(7-23)的左边平方，并把 m 大于 2 的上述平方后的方程相加，要求所得和最小。如果可以在不止两帧中得到量测，这个方法可以对每对至少可得到 5 对相应描述的图象对重复上述过程。

这里所介绍的方法不同于 Roach 和 Aggarwal 所报告的方法。我们所做的是通过消除未知变量减小搜索空间的维数。此外，这里所描述的封闭形式的表达式便于识别量测的特殊组合，这种组合可导致按最小途径的局部极值。这种根据两帧图象得到点的三维结构的方法可用作整体最小化的初始值。这将试图同时确定所有固定点相对物体坐标系统的三维位置，然后确定从帧到帧的位移，也就可确定物体在整个帧序列中的空间轨迹。

按最小化方法可进一步减小搜索空间。假设所研究物体在平面中运动，因此，物体的运动被限制在平移和绕垂直于运动平面的轴旋转。这些约束由误差表达式公式的最小化来实现。

7.3.3 根据图象流动力学恢复表面结构和三维运动

当有纹理的刚体在空间运动时，由单个观察者记录的图象序列包含以图象流场形式表示的信息。图象流不仅由观察者可见的物体表面的结构，而且由物体与观察者之间的相对刚体运动所决定。Waxman 和 Ullman 导出了这个流场（flow field）^[Wax 85]。

我们认为相对刚体运动是由用空间坐标系(X,Y,Z)（图 7.20）表示的观察者产生的。系统的原点位于透视投影的顶点，并且 Z 轴的方向是沿观察瞬场的中心。在这样的坐标系之下，瞬时刚体运动可用原点的位移矢量 $V=(V_x, V_y, V_z)$ 以及系统的旋转矢量 $\Omega=(\Omega_x, \Omega_y, \Omega_z)$ 来表示。屏幕上图象坐标系的点(x,y)位于空间坐标(x,y,1)处。

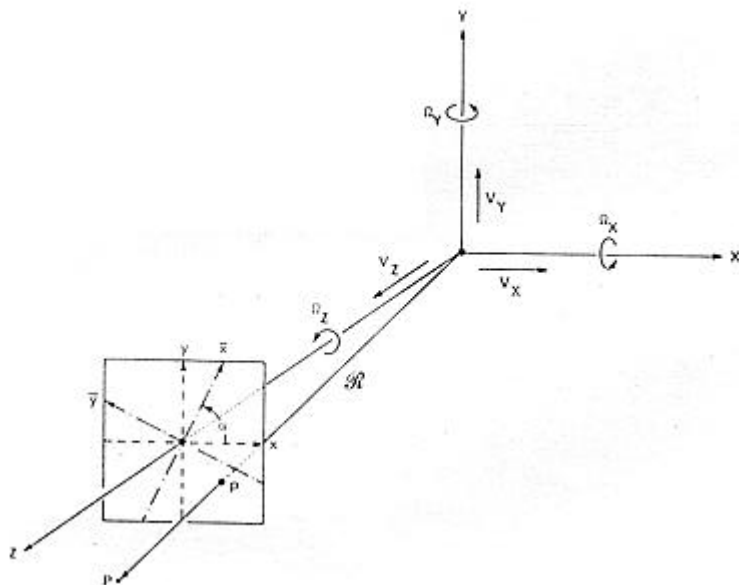


图 7.20 随观察者运动的 (x, y, z) 空间坐标系，在原始坐标系中的图象坐标 (x, y) 和在被转换坐标系中的图象坐标 (\bar{x}, \bar{y})

也即图象是重新转化过并调整比例到单位焦距。

由于观察者的运动，空间中的 P 点（用位置矢量 R 定位）以相对速度 $U = -(V + \Omega \times R)$ 运动。在每一瞬间 P 点在屏幕上的投影是 p ，坐标 $(x, y) = (X/Z, Y/Z)$ 。 p 点速度为 $(u_x, u_y) = (\dot{x}, \dot{y})$ ，这可通过把图象坐标相对时间求微分和应用 P 点空间坐标时间导数 U 的分量求得。新得结果为

$$\begin{aligned} u_x &= \left(x \frac{V_Z}{Z} - \frac{V_x}{Z} \right) + \left[xy\Omega_x - (1 + x^2)\Omega_y + y\Omega_z \right] \\ u_y &= \left(y \frac{V_Z}{Z} - \frac{V_y}{Z} \right) + \left[(1 + y^2)\Omega_x - xy\Omega_y - x\Omega_z \right] \end{aligned} \quad (7-24)$$

(7-24)式的推导

$$\begin{aligned} \because x &= X/Z \\ \therefore \frac{\partial x}{\partial t} &= \frac{\partial}{\partial t} (X/Z) = \frac{1}{Z} \left(\frac{\partial X}{\partial t} \right) + X \frac{\partial}{\partial t} \left(\frac{1}{Z} \right) \\ &= \frac{1}{Z} \frac{\partial X}{\partial t} - \frac{X}{Z^2} \frac{\partial Z}{\partial t} \\ \therefore \frac{\partial X}{\partial t} &= U_X = -(V + \Omega \times R)_X \quad \text{即 } U \text{ 的 } X \text{ 分量} \\ \frac{\partial Z}{\partial t} &= U_Z = -(V + \Omega \times R)_Z \quad \text{即 } U \text{ 的 } Z \text{ 分量} \\ (V + \Omega \times R)_X &= V_X + (\Omega \times R)_X \\ &= V_X + \left[(\bar{\Omega}_x + \bar{\Omega}_y + \bar{\Omega}_z)(X\bar{i} + y\bar{j} + Z\bar{k}) \right]_X \\ &= V_X + Z\Omega_y - Y\Omega_z \end{aligned}$$

上式用到了下面的向量叉积公式

$$\begin{aligned}\bar{A} \times \bar{B} &= (a_1 \bar{i} + a_2 \bar{j} + a_3 \bar{k})(b_1 \bar{i} + b_2 \bar{j} + b_3 \bar{k}) \\ &= (a_2 b_3 - a_3 b_2) \bar{i} + (a_3 b_1 - a_1 b_3) \bar{j} + (a_1 b_2 - a_2 b_1) \bar{k}\end{aligned}$$

同理有

$$\begin{aligned}(V + \Omega \times R)_Z &= V_Z + (\Omega \times R)_Z \\ &= V_Z + (\Omega_X Y - \Omega_Y X) \\ \therefore \frac{\partial x}{\partial t} &= \frac{1}{Z} \frac{\partial X}{\partial t} - \frac{X}{Z^2} \frac{\partial Z}{\partial t} \\ &= -\frac{1}{Z} (V + \Omega \times R)_X + \frac{X}{Z^2} (V + \Omega \times R)_Z \\ &= -\frac{1}{Z} (V_X + Z\Omega_Y - Y\Omega_X) + \frac{X}{Z^2} (V_Z + Y\Omega_X - X\Omega_Y) \\ &= \left(\frac{X}{Z^2} V_Z - \frac{V_X}{Z} \right) + \left[\frac{XY}{Z^2} \Omega_Z - \left(\frac{X^2}{Z^2} + 1 \right) \Omega_Y + \frac{Y}{Z} \Omega_Z \right]\end{aligned}$$

在单位集焦距的假设下 $x = X/Z$, $y = Y/Z$

$$\therefore \frac{\partial x}{\partial t} = \left(x \frac{V_Z}{Z} - \frac{V_X}{Z} \right) + [xy\Omega_X - (1+x^2)\Omega_Y + y\Omega_Z]$$

总结上面的推导，我们得到以下结论：

设透视投影的顶点位于空间坐标 (X, Y, Z) 的原点， Z 轴沿视线指向物体；坐标系平移的速度 $V \equiv (V_X, V_Y, V_Z)$ ，旋转的速度 $\Omega \equiv (\Omega_X, \Omega_Y, \Omega_Z)$

图象坐标系 (x, y) 的原点在空间的位置为 $(X, Y, Z) = (0, 0, 1)$ ，设空间点 P 的像点为 p 。由于观察者的运动，空间中 P 点的相对运动 R 为

$$-(V + \Omega \times R)$$

P 点运动的分量可表示为

$$\begin{aligned}\dot{X} &= -V_X - \Omega_Y Z + \Omega_Z Y \\ \dot{Y} &= -V_Y - \Omega_Z X + \Omega_X Z \\ \dot{Z} &= -V_Z - \Omega_X Y + \Omega_Y X\end{aligned} \quad (7-25)$$

P 在屏幕上的投影坐标为（屏幕位于 $Z=1$ 处）

$$x = X/Z, \quad y = Y/Z$$

图象点 p 的相应运动速度为 $(u_x, u_y) = (\dot{x}, \dot{y})$

$$\begin{aligned}u_x &= \left(x \frac{V_Z}{Z} - \frac{V_X}{Z} \right) + [xy\Omega_X - (1+x^2)\Omega_Y + y\Omega_Z] \\ u_y &= \left(y \frac{V_Z}{Z} - \frac{V_Y}{Z} \right) + [(1+y^2)\Omega_X - xy\Omega_Y - x\Omega_Z]\end{aligned} \quad (7-26)$$

从上式可见图象流（image flow）既不反映到物体上点的绝对距离，也不反映通过空间的绝对平移速度；因为这两者都与到物体的距离成比例。

为了解视线附近物体的局部结构，我们将做图象原点 $(x, y) = (0, 0)$ 附近图象流的运动分析（Kinematics analysis），为此，我们必须取图象流（方程 7-24）对图象坐标的各种导数，这将引入视线上物体表面的斜率和曲率（slopes and curvatures）。因此，我们必须能以所希望的分辨率把视线周围的表面区域描述为平滑的，即二阶可微。给定局部的唯一表面 $Z = h(X, Y)$ 。我们可以对 Z 轴的台劳级数来描述这个领域

$$\begin{aligned}
Z = Z_0 &+ \left(\frac{\partial Z}{\partial X} \right)_0 X + \left(\frac{\partial Z}{\partial Y} \right)_0 Y + \frac{1}{2} \left(\frac{\partial^2 Z}{\partial Y^2} \right)_0 Y^2 \\
&+ \left(\frac{\partial^2 Z}{\partial X \partial Y} \right)_0 XY + \frac{1}{2} \left(\frac{\partial^2 Z}{\partial X^2} \right)_0 X^2 + \dots
\end{aligned} \quad (7-27)$$

我们可以利用方程(7-26)，用图象坐标系来表示(7-27)。因此蕴含方程 $Z = h(xZ, yZ)$ 。这可被局部地转换到显式表面关系 $Z = Z(x, y)$

$$\begin{aligned}
Z = Z_0 &+ \left(\frac{\partial Z}{\partial x} \right)_0 x + \left(\frac{\partial Z}{\partial y} \right)_0 y + \frac{1}{2} \left(\frac{\partial^2 Z}{\partial x^2} \right)_0 x^2 \\
&+ \frac{1}{2} \left(\frac{\partial^2 Z}{\partial y^2} \right)_0 y^2 + \left(\frac{\partial^2 Z}{\partial x \partial y} \right)_0 xy + \dots
\end{aligned} \quad (7-28)$$

通过用(7-26)式替代(7-27)中的 X 和 Y ，以及每当出现 Z 时就置换成(7-27)的右边。经过项的合并以及与级数(7-28)比较得到

$$\begin{aligned}
\frac{1}{Z_0} \left(\frac{\partial Z}{\partial x} \right)_0 &= \left(\frac{\partial Z}{\partial X} \right)_0 \\
\frac{1}{Z_0} \left(\frac{\partial Z}{\partial y} \right)_0 &= \left(\frac{\partial Z}{\partial Y} \right)_0 \\
\frac{1}{Z_0} \left(\frac{\partial^2 Z}{\partial x^2} \right)_0 &= \left(\frac{\partial^2 Z}{\partial X^2} \right)_0 + 2 \left(\frac{\partial^2 Z}{\partial X} \right)_0^2 \\
\frac{1}{Z_0} \left(\frac{\partial^2 Z}{\partial y^2} \right)_0 &= \left(\frac{\partial^2 Z}{\partial Y^2} \right)_0 + 2 \left(\frac{\partial^2 Z}{\partial Y} \right)_0^2 \\
\frac{1}{Z_0} \left(\frac{\partial^2 Z}{\partial x \partial y} \right)_0 &= \left(\frac{\partial^2 Z}{\partial X \partial Y} \right)_0 + 2 \left(\frac{\partial Z}{\partial X} \right)_0 \left(\frac{\partial Z}{\partial Y} \right)_0
\end{aligned} \quad (7-29)$$

方程式(7-29)把空间中的表面描述的无量纲的斜率和曲率与图象坐标系的斜率和曲率相联系，式(7-29)中所有的量是沿物体的视线求值的，所以有下标 0。

7.3.4 基于光流的代数法恢复运动参数和景物深度^[Tek 96]

(7-26) 式表示了物体在空间的运动参数 $V_x, V_y, V_z, \Omega_x, \Omega_y, \Omega_z$ 与图象平面上的光流 u_x, u_y 的关系，但此关系与深度 Z 有关。不过 Z 的因素可通过在 u_x, u_y 的方程中分别求解 Z ，并使两者相等来消去。这样，就可得到一个光流与运动参数之间的表达式如下：

$$\begin{aligned}
&-u_y e_1 + u_x e_2 - x(\Omega_x + \Omega_z e_1) - y(\Omega_y + \Omega_z e_2) - xy(\Omega_y e_1 + \Omega_x e_2) \\
&+ (x^2 + y^2)\Omega_z + (1 + y^2)\Omega_x e_1 + (1 + x^2)\Omega_y e_2 = u_x y - u_y x
\end{aligned} \quad (7-30)$$

上式中 $e_1 = V_x / V_z$ ，和 $e_2 = V_y / V_z$ ，代表扩展中心。(参见 7.3.1 中对 FOE 的定义)。

通过运动视觉的分析，我们能确定物体的平移速度和深度，但不是准确的值，而还差一个比例常数。在方程(7-30)中的未知数共有 5 个。根据这个方程可以求得运动参数的闭合解 [Zhu 88], [Hei 93]。

如果假设物体表面是平面，那么有可能根据光流，利用代数法估计景物的三维运动和结构。

1. 仿射流(Affine flow)

进行刚体运动的平面在正交投影下生成仿射流。为此我们先给出三维速度场在图象平面上的正交投影表达式。

$$\begin{aligned} u_x &= \dot{x} = \frac{\partial x}{\partial t} = \frac{\partial X}{\partial t} = U_x = -(V + \Omega \times R)_x \\ &= V_x + Z\Omega_y - y\Omega_z \\ u_y &= \dot{y} = \frac{\partial y}{\partial t} = \frac{\partial Y}{\partial t} = U_y = -(V + \Omega \times R)_y \\ &= V_y + x\Omega_z - Z\Omega_x \end{aligned} \quad (7-31)$$

物体的局部表面结构可用以下的平面方程 (7-32) 来近似

$$Z = l_0 + l_1 X + l_2 Y \quad (7-32)$$

把 (7-32) 式代入 (7-31)，并采用正交投影下的关系 $x = X$ 和 $y = Y$ ，可得

$$\begin{aligned} u_x &= a_1 + a_2 x + a_3 y \\ u_y &= a_4 + a_5 x + a_6 y \end{aligned} \quad (7-33)$$

$$\begin{aligned} \text{其中} \quad a_1 &= V_x + l_0 \Omega_y, & a_2 &= l_1 \Omega_y, & a_3 &= l_2 \Omega_y - \Omega_z \\ a_4 &= V_y - l_0 \Omega_x, & a_5 &= \Omega_z - l_1 \Omega_x, & a_6 &= -l_2 \Omega_x \end{aligned}$$

如果已知三个或更多点的光流，就可建立 6 个以上的方程，从而求解 a_1, a_2, \dots, a_6 。然而，由于正交投影的缘故，不可能从 a_1, a_2, \dots, a_6 中唯一地确定所有的 8 个运动和结构参数。例如， l_0 在正交投影下是不可观察的。

2. 二次流 (Quadratic flow)

二次流是最基本的流场模型。因为它是平面表面在透视投影下的准确模型。如果表面不是平面，那么此模型是表面的一阶泰勒展开意义下的近似。在透视投影条件下由平面方程 (7-32) 可得

$$\frac{1}{Z} = \frac{1}{l_0} - \frac{l_1}{l_0} x - \frac{l_2}{l_0} y \quad (7-34)$$

因为 $Z = l_0 + l_1 X + l_2 Y = l_0 + l_1 x Z + l_2 y Z$ ，(设：摄像机的焦距 f 为 1)，二边同除 $Z l_0$ 即得 (7-34)。

把 (7-34) 代入透视速度场模型 (7-26)，即得 8 参数的二次流场

$$\begin{aligned} u_x &= a_1 + a_2 x + a_3 y + a_7 x^2 + a_8 xy \\ u_y &= a_4 + a_5 x + a_6 y + a_7 xy + a_8 y^2 \end{aligned} \quad (7-35)$$

$$\begin{aligned} \text{其中} \quad a_1 &= -\left(\frac{V_x}{l_0} + \Omega_y \right), & a_2 &= \frac{V_x l_1}{l_0} + \frac{V_z}{l_0}, & a_3 &= \frac{V_x l_2}{l_0} + \Omega_z \\ a_4 &= -\left(\frac{V_y}{l_0} - \Omega_x \right), & a_5 &= \left(\frac{V_y l_1}{l_0} - \Omega_z \right), & a_6 &= \frac{V_y l_2}{l_0} + \frac{V_z}{l_0} \\ a_7 &= -\left(\frac{V_z l_1}{l_0} + \Omega_y \right), & a_8 &= -\left(\frac{V_z l_2}{l_0} - \Omega_x \right) \end{aligned}$$

如果已知平面上 4 个以上点的光流, 则可建立 8 个以上的方程, 而未知量是 8 个, 从而可以求解 a_1, a_2, \dots, a_8 . 请注意, l_0 是表面与摄像机之间的距离, 它总是以比例因子的形式出现。

参考文献:

- [Ban 84] Bandyopadhyay, A., Interest Points, disparities and Correspondence, Proc. of Image Understanding Workshop, 1984, PP104-107.
- [Hei 93] Heikkonen, J., Recovering 3D motion from optical flow field, Image Processing :Theory and Applications, Vernazza, Venetsanopoulos, and Braccini, eds., Elvsevier,1993
- [Hil 84] Hildreth, E.C., Computations Understanding the measurement of Visual motion, AI 23, 1984, pp309-354.
- [Hor 81] Horn B.K.P & Sckunck B.G., Determining Optical flow, AI vol 17, No 1-3, Aug, 1981, pp 185-203.
- [Hor 86] Horn, B.K.P., Robot Vision, 1986, MIT Press.
- [Kan 87] Kanatani, K., Structure and motion from optical flow under perspective projection, CUGIP 38, 1987, 122-146.
- [Kit 82] Kitchen, L, and Rosenfeld, A., Gray-level Corner detector, Pattern Recognition letters, Dec, 1982, PP.95-102.
- [Kor 78] Korn, A., Segmentierung und erkennung eines objektes in natuerlicher umgebung, In E. Triendl(ed), Informatik - Fachbericht, Nr. 17 "Bildrevarbeitung und Mustererkennung, pp.265-274, Springer-Verlag, 1928.
- [Mor 77] Moravec, H.P., Towards automatic visual obstacle avoidance, Int. Joint Conf. On Artificial Intelligence, 1977, Cambridge, Mass., P.584.
- [Mor 80] Moravec, H.P., Obstacle avoidance and navigation in real world by a seeing robot rover, Stanford Artificial Intelligence laboratory AIM-340, 1980.
- [Nag 81] Nagel, H.H., Representation of moving rigid objects based on visual observation, Computer, Vol.14, No.8. Aug, 1981, pp29-39.
- [Nag 82] Nagel, H., Displacement vectors derived from second order inder intensity variation in image sequences, Mitteilung IFI-HH-M-97/82,IFI, March, 82.
- [Tai 84] Tai, R.Y., and Huang, T.s., Uniqueness and estimation of 3-D motion parameters of rigid objects with curved surface, IEEE Trans of Pattern Analysis and Machine Intelligence, PAMI-6, No.1, 1984.
- [Tek 96] Tekalp,A. M.,Digital Video Processing, Prentice Hall PTR,1996
- [Tho 84] Thorpe, C.E., An Analysis of interest operators for FIDO, IEEE Workshop on Computer Vision, 1984, PP 135-140.
- [Ull 79] Ullman, S., The interpretation of Visual motion Cambridge, MIT Press, 1979.
- [Ull 81]Ullman, S., Analysis of visual motion by biological and Computer systems, Computer, Vol.14, No8, Aug.1981.PP.57-69.
- [Wax 85] Waxman, A.M., and Ullman, S., Surface structure and three-dimensional motion from image flow Kinematics, Int. J. of Robotics Research, Vol.4, No.3, Fall, 1985, PP72-94.
- [Zhu 88] Zhuang,X., Huang, T.S., Ahuja, N., and Haralick, R. M., A simplified linear optical flow motion algorithm, Comp. Vis. Graph. And Image Proc., Vol. 42, 334-344,1988