

第 11 讲：运动建模

- 摄像机模型

- 透视投影
- 正交投影
- 仿射投影

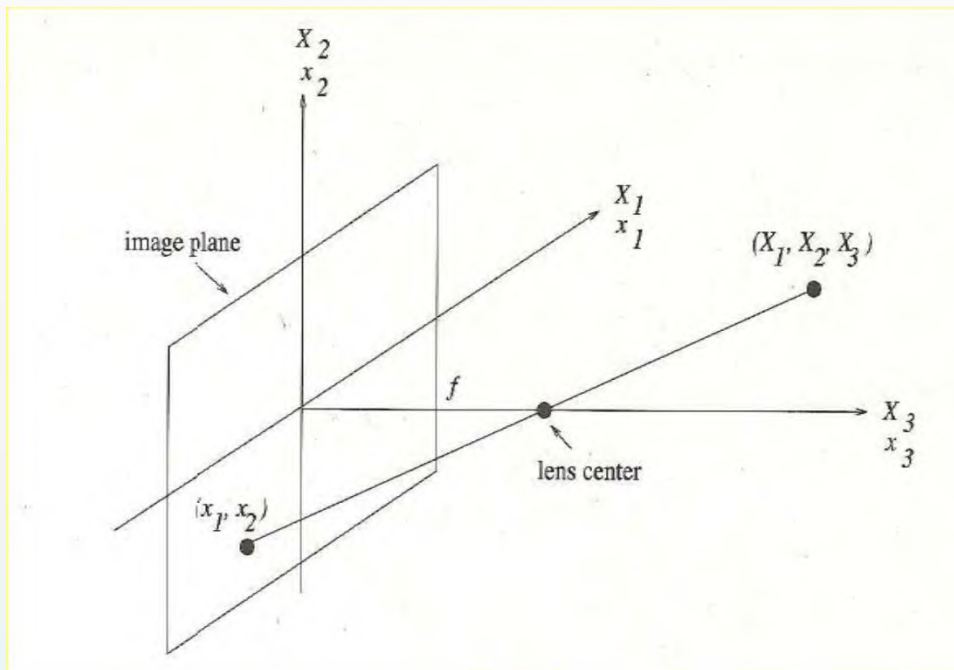
- 运动模型 – 投影运动 vs. 表观运动

- 投影 3D 刚体运动模型
- 表观 2D 运动: 参数运动模型
- 表观 2D 运动: 非参数运动模型

- 二维运动估计问题

- 光流方程
- 位移帧差
- 孔径与遮挡问题
- 分层运动估计
- 运动估计的性能度量

透视投影



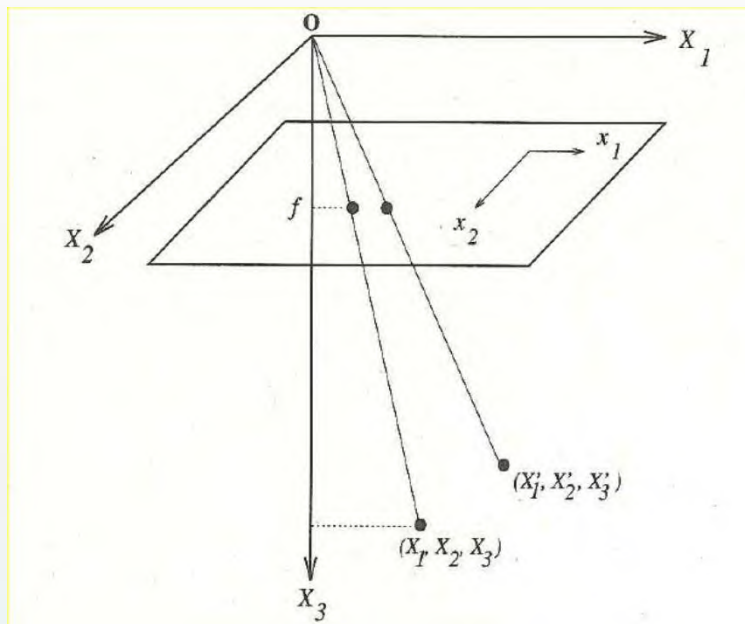
图像平面 (x_1, x_2) 相当于场景（世界）坐标系中的 $(X_1, X_2, 0)$ 平面，投影中心（镜头）位于物体和图像平面之间。

- 在此配置下，描述**透视投影**的关系式为：

$$\frac{x_1}{f} = -\frac{X_1}{X_3 - f} \quad \frac{x_2}{f} = -\frac{X_2}{X_3 - f} \quad \text{或} \quad x_1 = \frac{f X_1}{f - X_3} \quad x_2 = \frac{f X_2}{f - X_3}$$

其中 **f** 表示**镜头的焦距**（从投影中心到图像平面的距离）。

透视投影 (cont'd)



投影中心为世界坐标系的原点。

- 经过简单的变量代换，则有如下**等价表达式**：

$$x_1 = \frac{f X_1}{X_3} \quad x_2 = \frac{f X_2}{X_3}$$

为先前配置模型在 $X_3 \gg f$ 时的近似模型，只是**符号翻转**了，这是由于此时**图像和物体位于相同的坐标方向**上。

齐次坐标

- 透视投影在笛卡尔坐标中是非线性的，因为它需要除以 X_3 坐标。
- 可以表示为齐次坐标中的线性映射，其中3D场景点和2D图像点分别由4分量和3分量表示，由下式给出：

$$X_h = \begin{bmatrix} X_{h,1} \\ X_{h,2} \\ X_{h,3} \\ X_{h,4} \end{bmatrix} \quad x_h = \begin{bmatrix} x_{h,1} \\ x_{h,2} \\ x_{h,3} \end{bmatrix}$$

- 笛卡尔坐标和齐次坐标之间的关系，由下式表示：

$$X_1 = \frac{X_{h,1}}{X_{h,4}}, \quad X_2 = \frac{X_{h,2}}{X_{h,4}}, \quad X_3 = \frac{X_{h,3}}{X_{h,4}} \quad x_1 = \frac{x_{h,1}}{x_{h,3}}, \quad x_2 = \frac{x_{h,2}}{x_{h,3}}$$

称为去均质化。

透视投影 (cont'd)

- 透视投影可以写成**齐次坐标形式**:

$$\lambda \mathbf{x}_h = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{bmatrix}$$

其中 λ 是**尺度参数 (常量)**。在不知道关于场景具体测量数据的情况下, 无法获得尺度参数。

- 这种**线性关系**可以改写为:

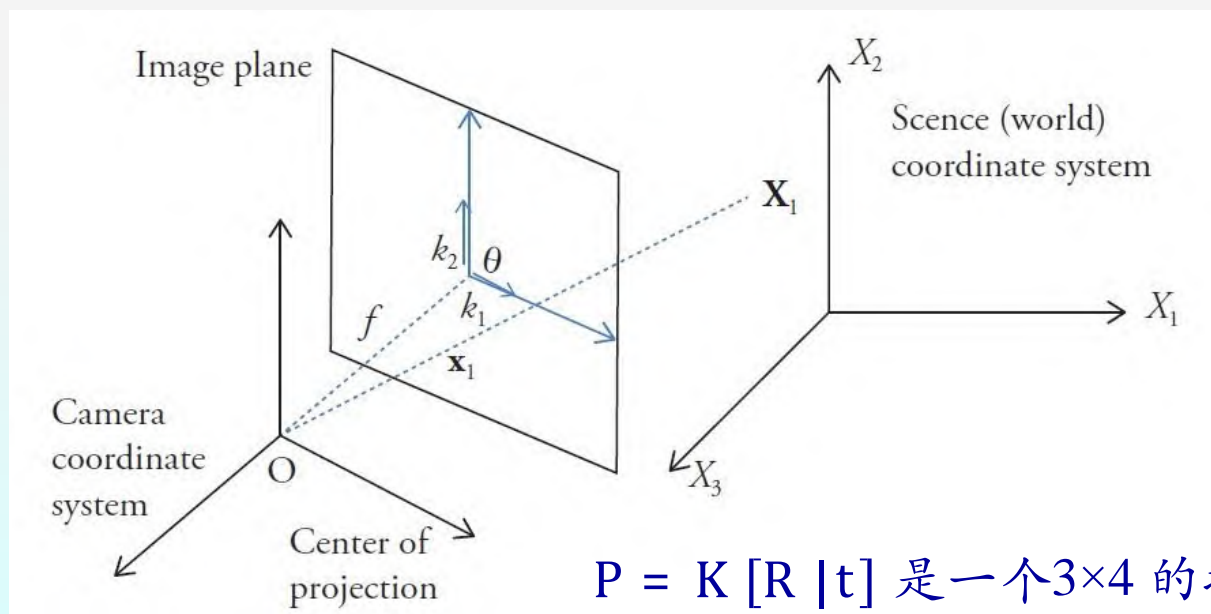
$$\lambda \mathbf{x}_h = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{bmatrix}$$

其中 3×3 矩阵被称为**相机标定 (内部参数) 矩阵**, 而 3×4 矩阵被称为**外部矩阵**, 表示相机和世界坐标系的相对位置。在以上表示中, 相机和世界坐标是对齐的。

透视相机模型

- 包括其它相机标定（内部）参数和外部参数（任意相机姿态）的最一般形式是：

$$\lambda \mathbf{x}_h = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} f & k_1 & s & x_{1,0} \\ 0 & f & k_2 & x_{2,0} \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{bmatrix} = \mathbf{K} [\mathbf{R} \mid \mathbf{t}] \mathbf{X}_h = \mathbf{P} \mathbf{X}_h$$



$\mathbf{P} = \mathbf{K} [\mathbf{R} \mid \mathbf{t}]$ 是一个 3×4 的投影矩阵

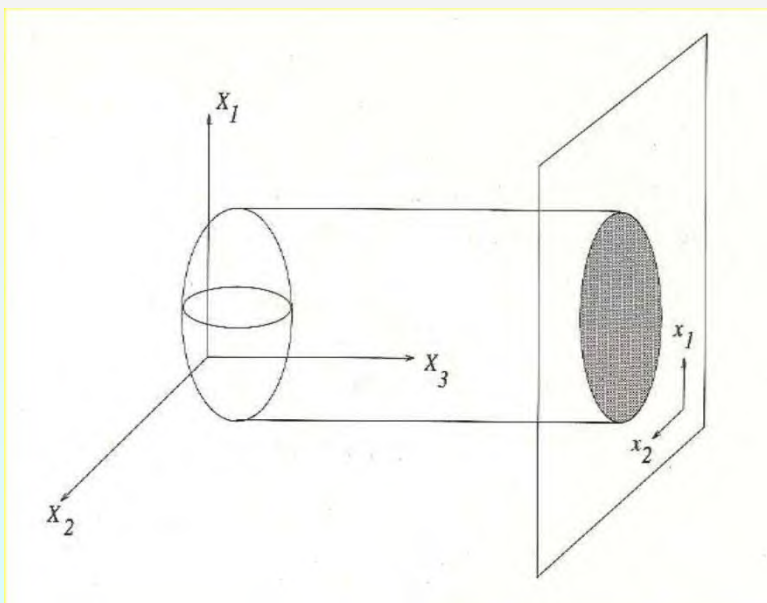
正交投影

- 正交投影可以在笛卡尔坐标中描述为：

$$x_1 = X_1 \quad \text{and} \quad x_2 = X_2$$

或以向量-矩阵形式：

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$



当物体距相机的距离远远大于物体上的点的深度范围时，正交投影提供了一种实际成像的近似表示。

仿射相机模型

- **仿射相机**包括正交投影，弱透视和平行透视投影模型。
- **弱透视投影**是一种正交尺度投影，提供了在正交投影的简单性和透视投影的现实性之间的折中。由下式表示：

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \mathbf{M} \mathbf{X}$$

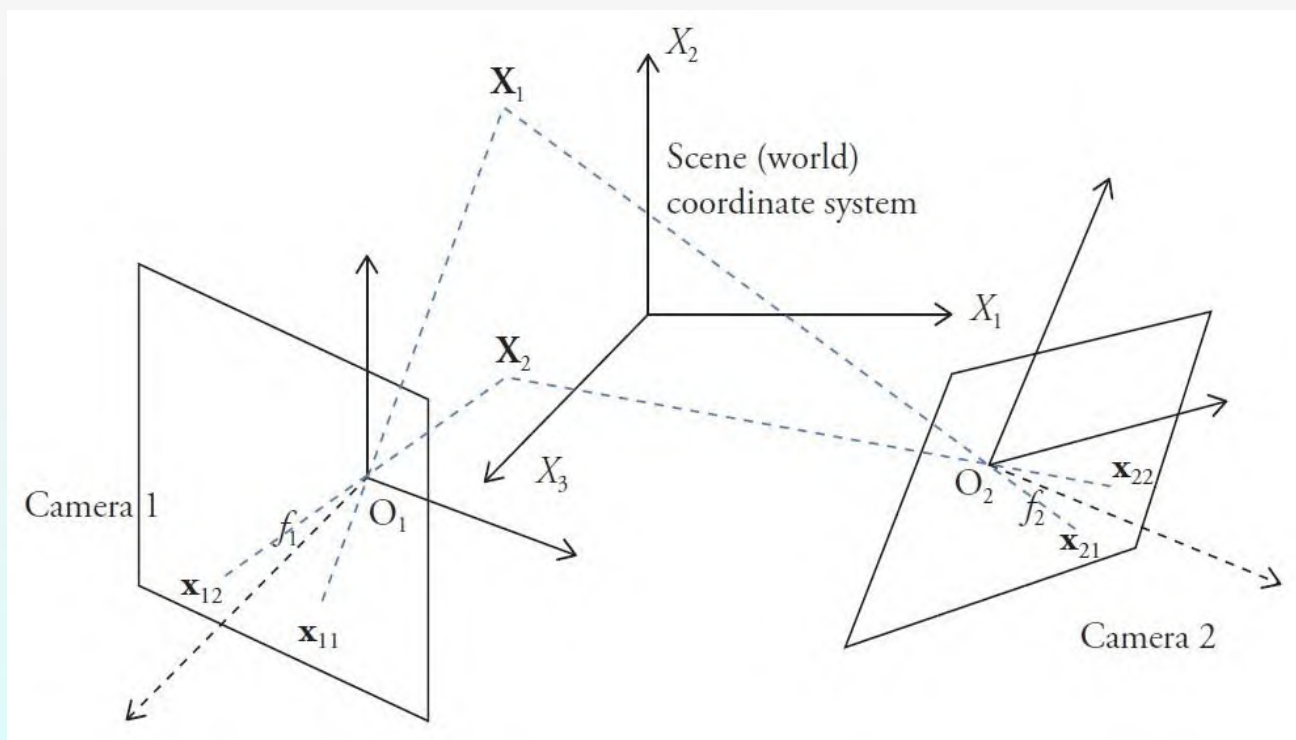
其中 f 是相机的焦距， Z 表示3D对象某个参考点的深度（沿着光轴的 X_3 值），并且 \mathbf{M} 称为**仿射相机矩阵**。

- 通过对物体相对于附近的平行于图像平面的参考平面的透视变形进行建模，**平行透视投影**扩展了正交尺度投影。

投影3D 运动

- 特征点的投影图像平面坐标的**齐次坐标形式**由下式给出：

$$\lambda_{ij} \mathbf{x}_{ij} = \mathbf{P}_i \mathbf{X}_j \quad i = 0, \dots, M-1, \quad j = 0, \dots, N-1$$



\mathbf{X}_j 是场景（世界）坐标中的第 j 个特征点

\mathbf{P}_i 是第 i 个相机的 3×4 投影矩阵， λ_{ij} 是标量。

CASE 1: 相机在静态场景或立体视觉中的运动

投影3D 运动 (cont'd)

- 令**初始相机矩阵**为 $\mathbf{P}_0 = \mathbf{K}[\mathbf{I}_{3 \times 3} | \mathbf{0}_{3 \times 1}]$ ，然后通过**旋转矩阵** \mathbf{R}_i 和**平移向量** \mathbf{T}_i , $i = 1, \dots, M-1$ 来模拟相机 i 和相机 0 （3D 相机坐标）之间的相对姿态（运动）。

- 在“投影运动”到可逆投影变换 \mathbf{H} 之间存在**模糊性**，这是由于：

$$\lambda_{ij} \mathbf{x}_{ij} = (\mathbf{P}_i \mathbf{H})(\mathbf{H}^{-1} \mathbf{X}_j) = \mathbf{P}_i \mathbf{X}_j \quad i = 1, \dots, M, \quad j = 1, \dots, N$$

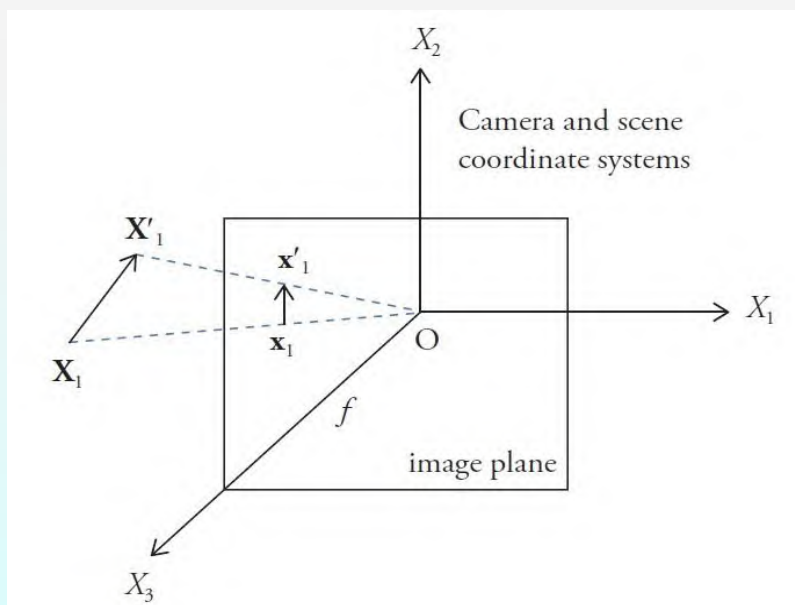
- 因此，对于**任意相机矩阵** $\mathbf{P}_i \mathbf{H}$ 和**形状（结构）矩阵** $\mathbf{H}^{-1} \mathbf{X}_j$ 对，其像平面（像素）投影的坐标集都相同，即其投影是**等效**的 [Har 04]。

投影3D 运动 (cont'd)

- 在笛卡尔坐标中的场景点的**刚体3D运动**（旋转和平移）

$$\begin{bmatrix} X'_1 \\ X'_2 \\ X'_3 \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \mathbf{T} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix}$$

可以用**齐次坐标**表示：



$$\mathbf{X}' = \begin{bmatrix} r_{11} & r_{12} & r_{13} & T_1 \\ r_{21} & r_{22} & r_{23} & T_2 \\ r_{31} & r_{32} & r_{33} & T_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{X}$$

- 3D运动向量 \mathbf{XX}' 被投影到图像平面运动向量 \mathbf{xx}' 。
- 沿**虚线方向的模糊性**。

CASE 2: 静止照相机的3D 场景运动

光度效应

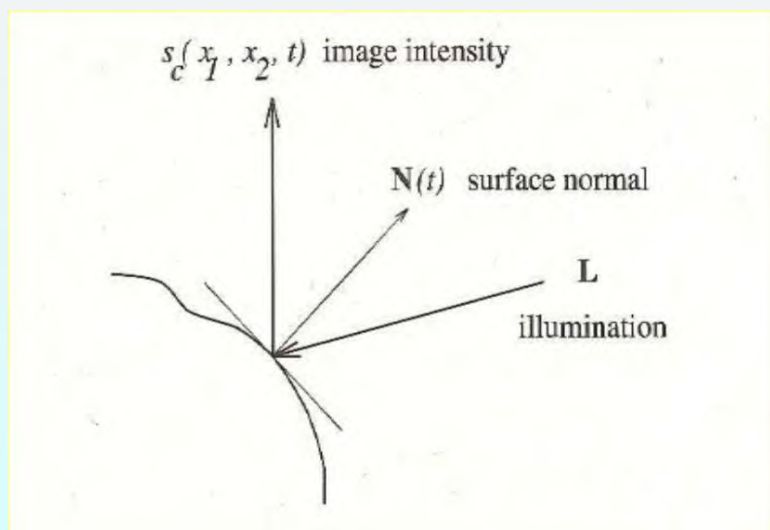
- 朗伯反射模型: 如果朗伯表面被具有均匀强度（时间）的单个远点光源照射，则所得到的图像强度由朗伯余弦定律给出：

$$s_c(x_1, x_2, t) = \rho N(t) \cdot L$$

其中 ρ 表示表面反射率，即表面反射光的部分

$L = (L_1, L_2, L_3)$ 是平均光源方向上的单位向量，

$N(t)$ 是场景在空间位置 (X_1, X_2, X_3) 的单位表面法线。

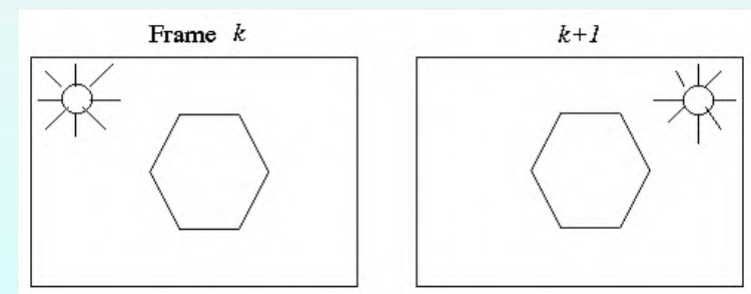
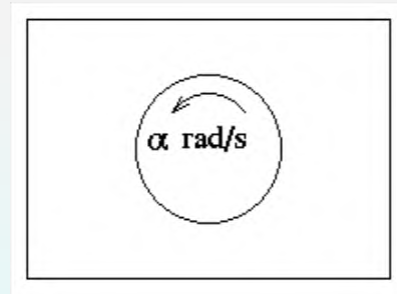


作为场景表面和相机之间的相对3D运动的结果，表面法线随时间变化，表面的光度性质也随之改变。

Pentland [Pen 91]指出，在某些情况下，运动的光度效应可以用来确定对象的几何形状。

表观 2D 运动

- **表观运动**指的是从时变图像强度图案（视频）所能感知（观察）的**对应（位移）**或**光流（速度）**场。
- 通常，“**表观运动**”（对应或光流）与“**投影运动**”（位移或速度）场**不同**，这是由于存在某些**模糊性**：
 - **缺乏足够的空间图像梯度**
 - 所有投影运动不产生光流
 - **外部照明和/或阴影的变化**
 - 可观察的光流可能不总是对应于实际运动。



投影运动模型

- 通用模型（深度图）：如果我们将每个特征点 X_3 的深度作为独立变量，然后将3D特征位置 X' 投影到图像平面中，则有：

$$x'_1 = \frac{f X'_1}{X'_3} = \frac{f (r_{11}X_1 + r_{12}X_2 + r_{13}X_3 + T_1)}{r_{31}X_1 + r_{32}X_2 + r_{33}X_3 + T_3} \quad x'_2 = \frac{f X'_2}{X'_3} = \frac{f (r_{21}X_1 + r_{22}X_2 + r_{23}X_3 + T_2)}{r_{31}X_1 + r_{32}X_2 + r_{33}X_3 + T_3}$$

- 将两个表达式的分子和分母乘以 f 除以 X_3 ，则：

$$x'_1 = f \frac{r_{11}x_1 + r_{12}x_2 + f r_{13} + \frac{f^T}{X_3}}{r_{31}x_1 + r_{32}x_2 + f r_{33} + \frac{f^T}{X_3}} \quad \text{and} \quad x'_2 = f \frac{r_{21}x_1 + r_{22}x_2 + f r_{23} + \frac{f^T}{X_3}}{r_{31}x_1 + r_{32}x_2 + f r_{33} + \frac{f^T}{X_3}}$$

其中每个特征的深度 X_3 作为这些表达式中的参数而出现。

- 存在约束2D图像运动（位移或光流）向量的方向的**6个3D运动参数（三个旋转和三个平移）**，而**深度参数**是确定确切的**2D运动向量**所需要的。

投影运动模型 (cont'd)

- 单应性（透视模型）：如果运动对象的3D结构（形状）由平面表面建模，以使对象点的深度相关，则可以减少自由深度参数的数量。
- 令该组3D特征点位于平面上：

$$aX_1 + bX_2 + cX_3 = 1$$

其中 $[a \ b \ c]^T$ 是平面的法向量。则3D运动模型是：

$$\begin{bmatrix} X'_1 \\ X'_2 \\ X'_3 \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \mathbf{T} = \mathbf{R} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \mathbf{T} \left([a \ b \ c] \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \right) = \mathbf{H} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

其中 $\mathbf{H} = \mathbf{R} + \mathbf{T}[a \ b \ c]$.

- 将3D场景坐标投影到2D图像平面中，可得到单应性关系

$$x'_1 = \frac{h_1 x_1 + h_2 x_2 + h_3}{h_7 x_1 + h_8 x_2 + h_9} \quad x'_2 = \frac{h_4 x_1 + h_5 x_2 + h_6}{h_9 x_7 + h_1 x_8 + h_2}$$

其中 h_9 有时被设置为1以解决尺度模糊性。

投影运动模型 (cont'd)

- 如果 $\mathbf{T} = \mathbf{0}$ ，则 $\mathbf{H} = \mathbf{R}$ ，此时单应性提供图像之间的精确映射（补偿相机标定和旋转），而与场景几何结构无关。
- 总之，两帧通过单应性相关，当且仅当：
 - 它们是来自不同相机位置 (\mathbf{R}, \mathbf{T}) 的相同3D平面表面的视图，
 - 它们由相同相机捕获，其中相机仅允许围绕其光学中心旋转和/或独立于场景几何的缩放（没有任何平移）。
- 残余平面视差运动模型 是以齐次坐标中核点(扩张的焦点) $\mathbf{e} = [e_1 \ e_2 \ e_3]^T$ 为中心的径向矢量场。图像平面的齐次坐标位移可以分解为：
$$\mathbf{x}' - \mathbf{x} = (\mathbf{x}' - \mathbf{x}_w) + (\mathbf{x}_w - \mathbf{x})$$

其中 $\mathbf{x}' - \mathbf{x}_w$ 表示平面运动分量。平面视差运动是：

$$\mathbf{x}_w - \mathbf{x} = -\frac{\gamma}{1 + \gamma e_3} (\mathbf{e}_3 \mathbf{x} - \mathbf{e})$$

其中 $\gamma = H/X_3$, H 表示3D点 \mathbf{X} 与场景中的参考 X_3 平面的垂直距离，该特征与形状无关。

2D 表观运动模型

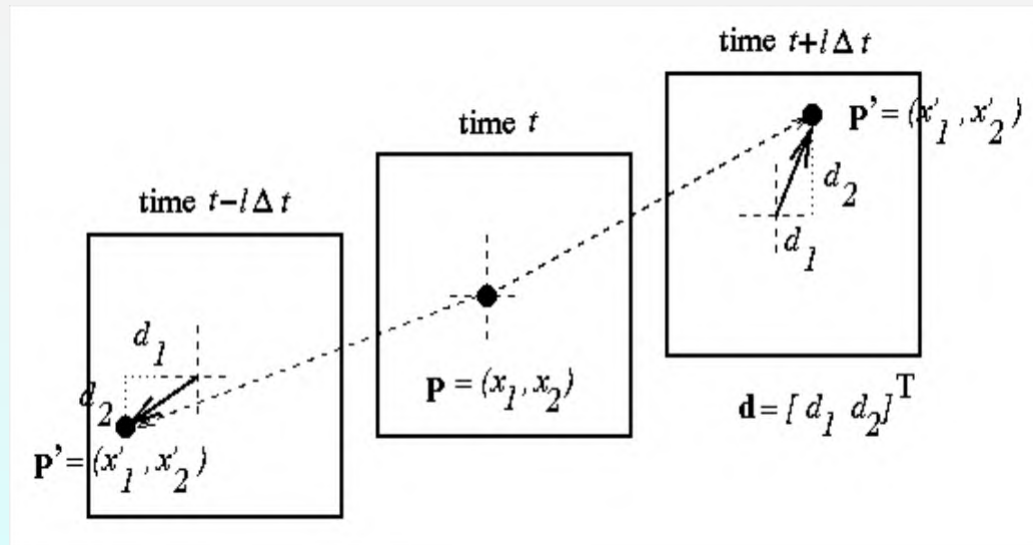
- 块平移模型:

$$x'_1 = x_1 + d_1$$

$$x'_2 = x_2 + d_2$$

其中 (x'_1, x'_2) 表示参考帧中的像素 (x_1, x_2) 的坐标。

- 正向运动估计: 从 t 到 $t + \Delta t$
- 反向运动估计: 从 t 到 $t - \Delta t$



2D 表观运动模型 (cont'd)

- **仿射模型**：平面表面的3D刚体运动的正投影可以通过六参数仿射模型来描述。它可以应用于图像帧或仅仅一个像素块，则有：

$$x_1 = a_1 x_1 + a_2 x_2 + a_3$$

$$x_2 = a_4 x_1 + a_5 x_2 + a_6$$

- 如果**表平面离摄像机足够远**，则可以假设从平面物体到摄像机的所有射线平行，则它可提供对单应性的良好近似。

- **特殊情况**:

- **纯旋转**: $x'_1 = x_1 \cos \theta + x_2 \sin \theta$, $x'_2 = -x_1 \sin \theta + x_2 \cos \theta$

通过角度 θ 对图像平面中坐标轴的旋转进行建模。

- **等距映射**是指具有3个自由度的旋转和平移, θ, a_3, a_6 。

- **各向同性缩放 (变焦)**: 如果 $a_1 = a_5 = k$, 则 $x'_1 = k x_1$, $x'_2 = k x_2$

- **相似性变换**是指旋转, 平移和缩放:

$$x'_1 = k x_1 \cos \theta + x_2 \sin \theta + d_1 \quad x'_2 = -x_1 \sin \theta + k x_2 \cos \theta + d_2$$

2D 表观运动模型 (cont'd)

- **双二次，伪透视和双线性模型** 是对单应性（透视模型）的线性近似。

- **双二次模型**，有12个自由参数：

$$x'_1 = a_1x_1 + a_2x_2 + a_3 + a_4x_1^2 + a_5x_2^2 + a_6x_1x_2$$

$$x'_2 = a_7x_1 + a_8x_2 + a_9 + a_{10}x_1^2 + a_{11}x_2^2 + a_{12}x_1x_2$$

可以通过单应性的泰勒级数展开获得。

- 具有8个自由参数的**伪透视模型**由下式给出：

$$x'_1 = a_1x_1 + a_2x_2 + a_3 + a_7x_1^2 + a_8x_1x_2$$

$$x'_2 = a_4x_1 + a_5x_2 + a_6 + a_7x_1x_2 + a_8x_2^2$$

是瞬时光流的近似。

- 具有8个自由参数的**双线性模型**由下式给出：

$$x'_1 = a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4$$

$$x'_2 = a_5x_1 + a_6x_2 + a_7x_1x_2 + a_8$$

2D 表观运动模型 (cont'd)

非参数模型:

- 确定性平滑度

- 块内为常数: 假设所有像素的MV相同。

- 块匹配, **Lukas-Kanade**, 相位相关

- 全局平滑性: 假设MV的像素到像素变化小。

- **Horn-Schunk**, 像素递归

- 方向平滑性: 假设MV的像素到像素变化仅沿图像边缘较小, 但不限制垂直于边缘方向 (沿图像梯度的方向) 的变化。

- 随机/概率平滑 (贝叶斯运动估计)

- 吉布斯随机场: 位移或分割场的平滑性根据一些能量函数来量化。

- 线性场: 在GRF框架中施加方向平滑约束。

运动估计问题

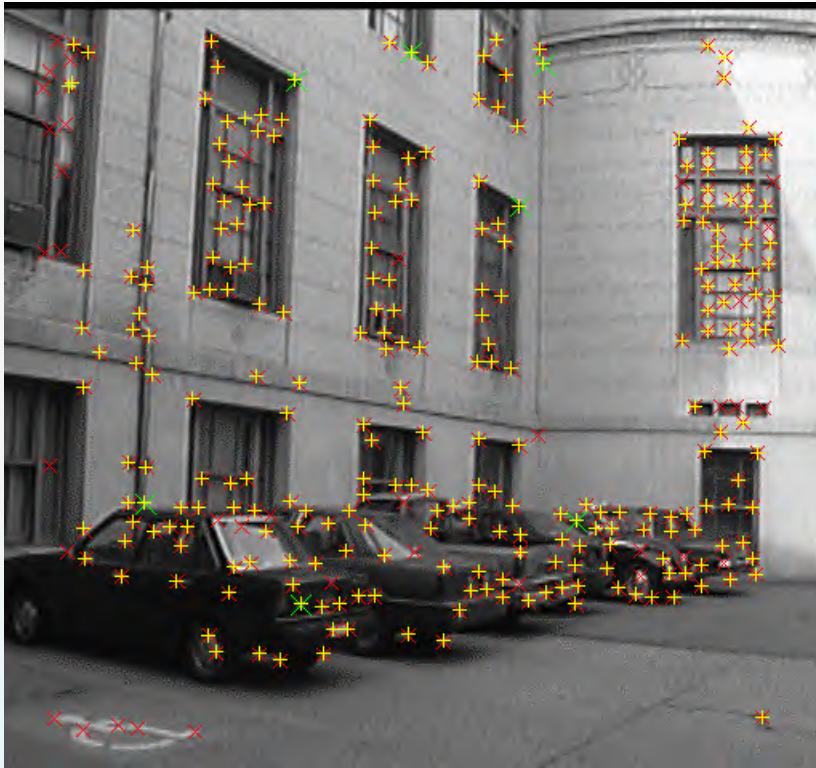
- 2D 运动估计

- 对应/位移 (稀疏)
- 光流估计 (稠密)
 - 匀速运动: $\mathbf{v} = \mathbf{d} / \Delta t$
- 图像配准 (全局或局部, 参数型)

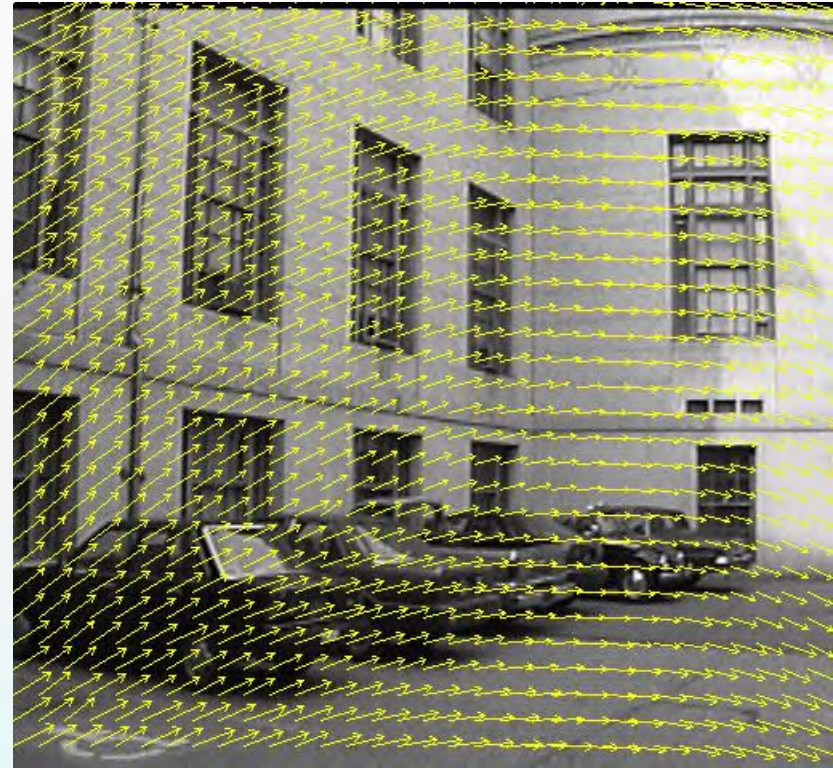
- 3D 运动 (姿态) 与结构估计

- 投影空间
- 仿射空间
- 欧氏空间

示例: 稀疏 vs. 稠密运动



稀疏特征/对应



稠密运动/光流

光流方程

- 沿运动轨迹的图像**亮度（强度）恒定性**
- 假设空间和时间由连续变量表示，亮度恒定性意味着**沿着运动轨迹的亮度变化率**：
$$\frac{d s_c(x_1, x_2, t)}{dt} = 0$$

- 此为**全微分表达式**，因为 x_1 和 x_2 沿着运动轨迹随 t 变化。根据微分的链式法则，有：

$$\frac{\partial s_c(x_1, x_2, t)}{\partial x_1} \frac{dx_1}{dt} + \frac{\partial s_c(x_1, x_2, t)}{\partial x_2} \frac{dx_2}{dt} + \frac{\partial s_c(x_1, x_2, t)}{\partial t} = 0$$

其中 $v_1(x, t) = dx_1/dt$, $v_2(x, t) = dx_2/dt$ 为速度的两个分量

- 光流方程（OFE）** 可以用向量表示法表示：

$$\langle \nabla s_c(\mathbf{x}, t), \mathbf{v}(\mathbf{x}) \rangle + \frac{\partial s_c(\mathbf{x}, t)}{\partial t} = 0$$

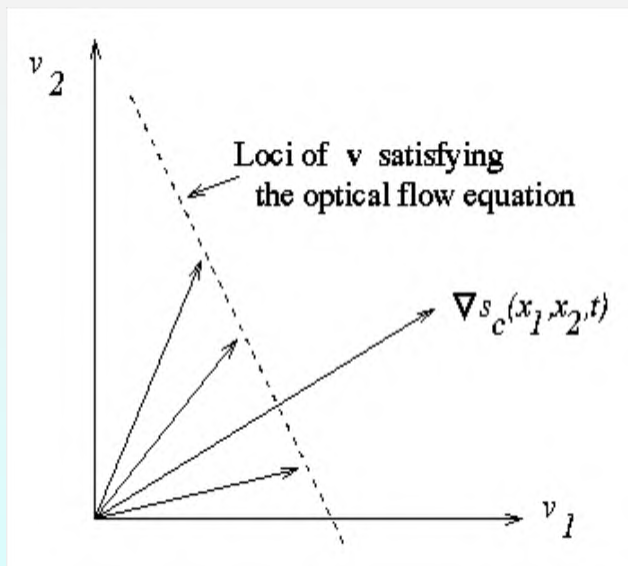
其中 $\nabla s_c(\mathbf{x}, t) = \left[\frac{\partial s_c(\mathbf{x}, t)}{\partial x_1}, \frac{\partial s_c(\mathbf{x}, t)}{\partial x_2} \right]^T$, $\langle \cdot, \cdot \rangle$ 表示向量内积。

法向流

- **光流方程 (OFE)** 对每个像素来讲是一个具有两个未知数的公式。因此，光流估计是一个**欠定 (病态)** 的问题。
- **法向流**: OFE还可以写成:

$$\|\nabla s_c(\mathbf{x}, t)\| \|\mathbf{v}(\mathbf{x}, t)\| \cos \beta + \frac{\partial s_c(\mathbf{x}, t)}{\partial t} = 0$$

其中 β 为向量 $\nabla s_c(\mathbf{x}, t)$ 和 $\mathbf{v}(\mathbf{x}, t)$ 之间的夹角。



- 然后，可以通过设置角度 $\beta = 0$ 计算每个位置处的法向流 $\|\mathbf{v}(\mathbf{x}, t)\|$ 的幅值，即：

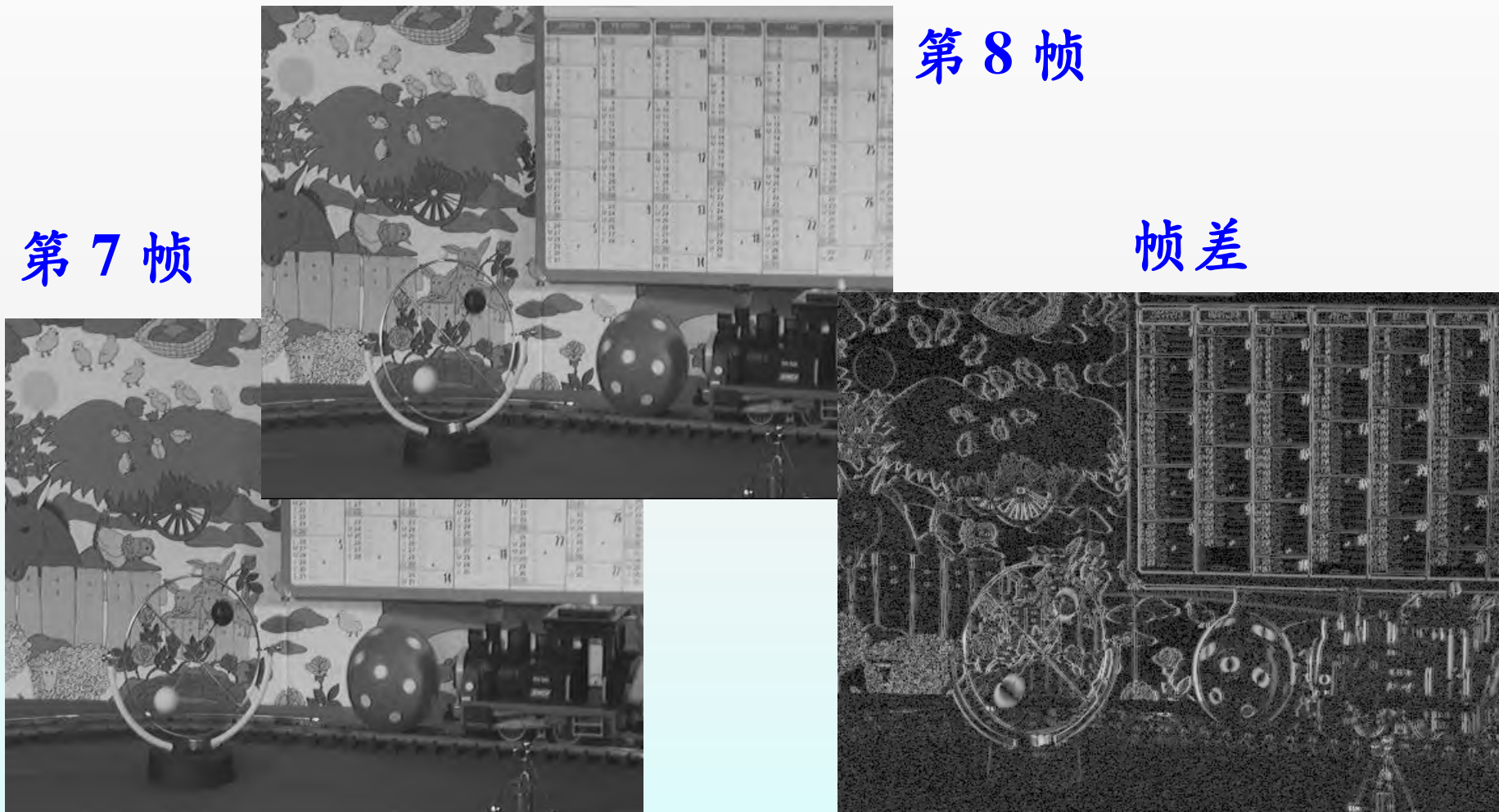
$$\|\mathbf{v}_{\perp}(\mathbf{x}, t)\| = \frac{-\frac{\partial s_c(\mathbf{x}, t)}{\partial t}}{\|\nabla s_c(\mathbf{x}, t)\|}$$

示例: 帧差

第 8 帧

第 7 帧

帧差



位移帧差 (DFD)

- 时间 t 和 $t' = t + \Delta t$ 之间的**DFD**由下式定义:

$$\text{DFD}(\mathbf{x}, \mathbf{d}) = s_c(\mathbf{x} + \mathbf{d}(\mathbf{x}), t + \Delta t) - s_c(\mathbf{x}, t)$$

其中 $\mathbf{d}(\mathbf{x}) = [d_1(\mathbf{x}), d_2(\mathbf{x})]^T$ 表示运动矢量 (MV) 场。

- 可以观察到:
 - i) 由于允许 $\mathbf{d}(\mathbf{x})$ 的分量取非整数值, 需要内插来计算DFD;
 - ii) 如果 $\mathbf{d}(\mathbf{x})$ 等于**真实MV**并且**没有插值误差**, 则**DFD可达到零**。
- 对于**小的 $\mathbf{d}\mathbf{x}$ 和 Δt** , 将 $s_c(\mathbf{x} + \mathbf{d}\mathbf{x}, t + \Delta t)$ 在 (\mathbf{x}, t) 处进行泰勒级数展开:

$$\begin{aligned} & s_c(x_1 + d_1(\mathbf{x}), x_2 + d_2(\mathbf{x}), t + \Delta t) \\ &= s_c(\mathbf{x}, t) + d_1(\mathbf{x}) \frac{\partial s_c(\mathbf{x}, t)}{\partial x_1} + d_2(\mathbf{x}) \frac{\partial s_c(\mathbf{x}, t)}{\partial x_2} + \Delta t \frac{\partial s_c(\mathbf{x}, t)}{\partial t} + \text{h.o.t.} \end{aligned}$$

- 代入DFD, **忽略高阶项** (h.o.t.) ,

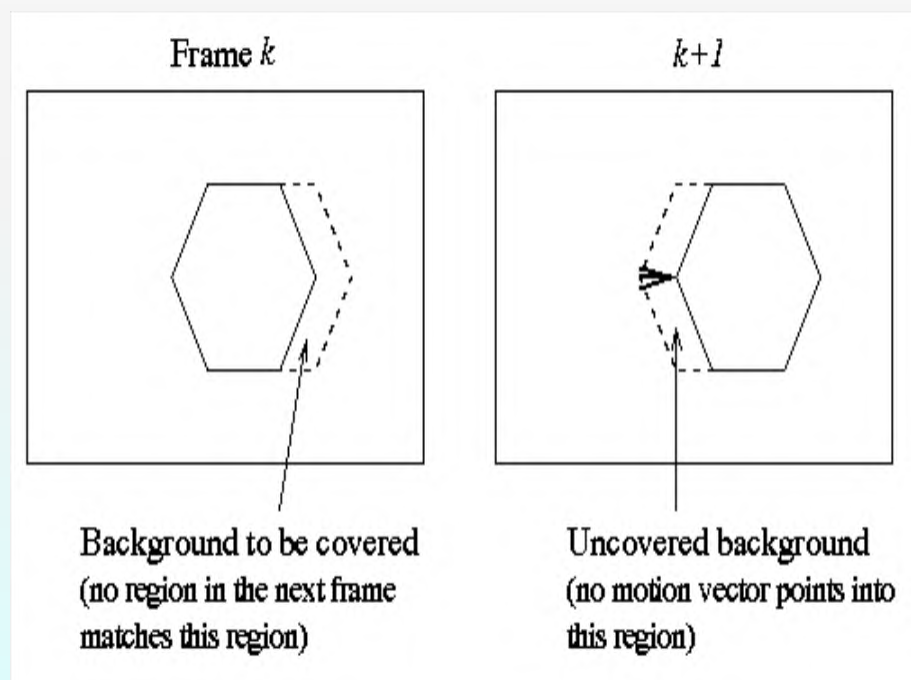
$$\text{DFD}(\mathbf{x}, \mathbf{d}) = \frac{\partial s_c(x_1, x_2, t)}{\partial x_1} d_1(\mathbf{x}) + \frac{\partial s_c(x_1, x_2, t)}{\partial x_2} d_2(\mathbf{x}) + \frac{\partial s_c(x_1, x_2, t)}{\partial t} \Delta t$$

位移帧差 (cont'd)

- **Case 1)** 极限 $\Delta t \rightarrow 0$: 令 $DFD(x, d) = 0$, 将两边除以 Δt , 取极限为 $\Delta t \rightarrow 0$, 得到OFE。可知让DFD等于零等价于在极限 $\Delta t \rightarrow 0$ 时得到OFE。
 - **Case 2)** 对于 Δt 有限情况: 从DFD获得间隔为 Δt 的任意两帧之间的位移矢量 $d(x)$ 的估计可以使用多种方式:
 - 搜索 $\hat{d}(x)$, 使一个像素块上将DFD的绝对值和等于零 (**块匹配策略**)。
 - 计算 $\hat{d}(x)$, 利用基于梯度的优化方案 (**像素递归策略**) 逐像素地令DFD等于零。
 - 令 $\Delta t = 1$, $DFD(x, \hat{d}(x)) = 0$, 使用从像素块获得的一组线性方程对 $\hat{d}(x)$ 进行求解。
 - 实际上, 对于任何 d 值, $DFD(x, \hat{d})$ **几乎都不会为零**, 因为:
 - i) **存在观察噪声**,
 - ii) 存在**遮挡** (覆盖/未覆盖区域),
 - iii) 对于**非整数MV**存在**内插误差**,
 - iv) 场景**照明**可能**逐帧地变化**。
- 因此, 我们需要对**DFD函数进行最小化**。

遮挡问题

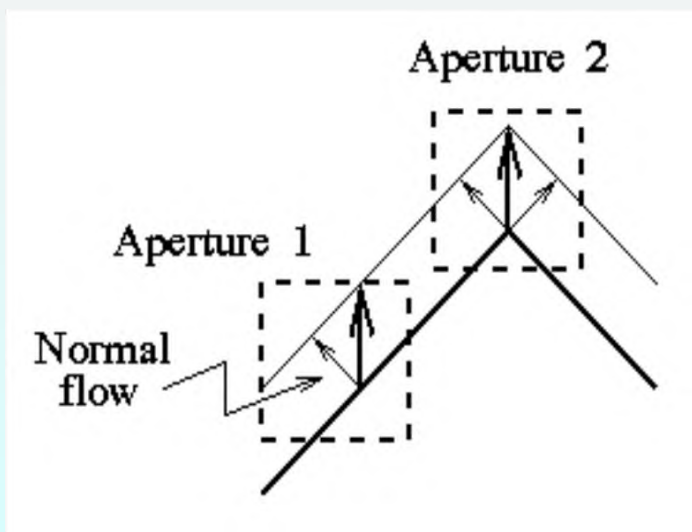
- **遮挡**是指由于对象相对于相机的运动而逐帧地**覆盖/露出**对象或背景的部分，使得当前帧中的一些像素在参考帧中不具有对应性（也称为覆盖/未覆盖的对象/背景区域）。



- 有**两种遮挡源**:
 - **相互遮挡**，其中移动物体覆盖另一物体或背景的一部分；
 - **自遮挡**，例如，物体顺时针旋转出平面，左边缘被覆盖，并且一些新的纹理从右边缘露出。

孔径问题

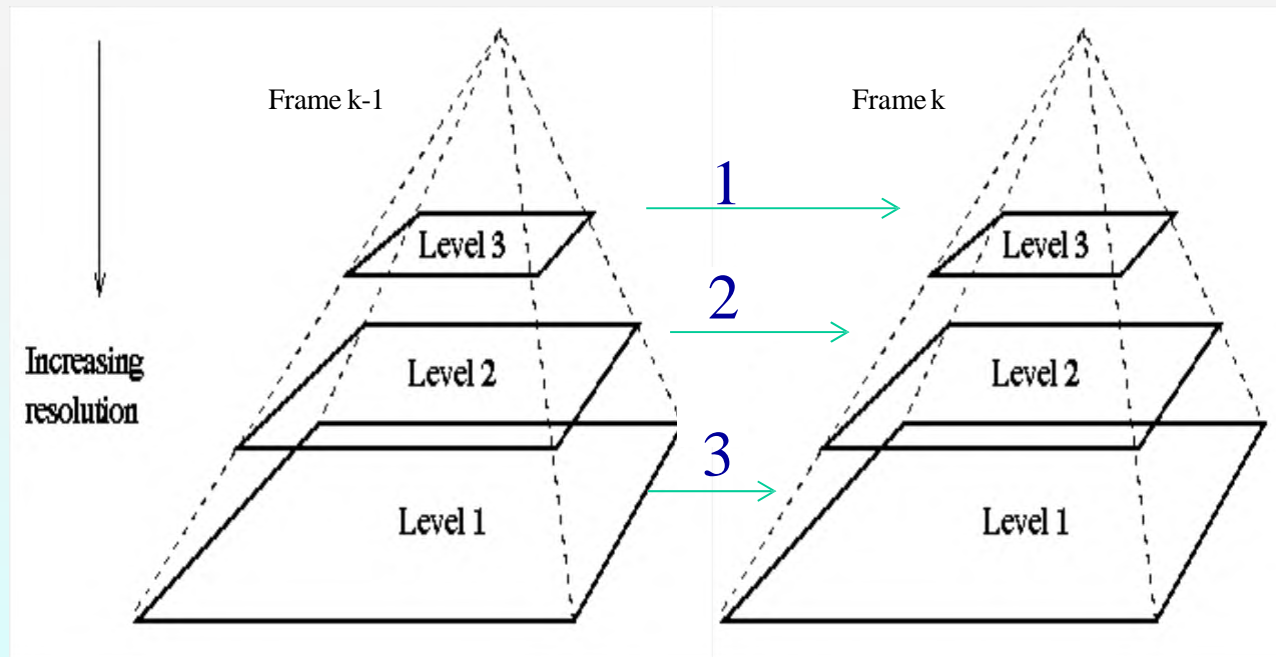
- 给定用于估计MV的块（**孔径**），存在**三种可能**的情况：
 - **CASE 1**:在孔径内存在**两个线性独立的梯度向量**。 v_1 和 v_2 都可以唯一地被估计。
 - **CASE 2**:在孔径内**只有一个梯度方向**。仅可以对块唯一地估计其法向流（运动）向量。



- **CASE 3**:在孔径内**没有梯度**。运动不会导致可观察的强度变化，因此有无穷多解。

分层运动估计

- 孔径问题可以通过**分层（多分辨率或多尺度）**运动估计来克服。
- 以**自上而下**的方式。



性能度量

具有真实参考时的度量:

- 误差范数 (NE) : 给定像素处的估计运动矢量 (d_{i1}, d_{i2}) 和对应的GT矢量 $(d_{i1}^{GT}, d_{i2}^{GT})$, NE测量值可通过下式计算:

$$\|e_i\| = \sqrt{(d_{i1} - d_{i1}^{GT})^2 + (d_{i2} - d_{i2}^{GT})^2}$$

- 像素处的估计运动矢量 (d_{i1}, d_{i2}) 和对应的GT矢量 $(d_{i1}^{GT}, d_{i2}^{GT})$ 之间的角度误差 (AE) 是3D矢量 $d_i = (d_1, d_2, 1.0)$ 和 $d_i^{GT} = (d_1^{GT}, d_2^{GT}, 1.0)$ 之间的夹角 θ_i , 可通过点 (内) 积规则来计算:

$$\cos \theta_i = \frac{\langle d_i, d_i^{GT} \rangle}{\|d_i\| \|d_i^{GT}\|}$$

无真实参考时的度量:

- 运动补偿的均方误差(MSE)或误差和(SAD).