本章目录

目标识别、跟踪与光流估计

第八章 多模态感知与智能决策

8.1 背景意义与图像处理基础

8.2 目标识别、跟踪与光流估计

8.3 定位建图与预测规划

8.4 常用感知传感器

8.5 多模态感知坐标转换

8.6 多模融合与控制决策

8.7 关键工具与框架接口

8.8 实验案例与真机迁移

8.9 习题与课堂讨论

本节 5 个教学单元

8.2.1 目标检测与实例分割
教材对应：第8.1.2.2节

8.2.2 在线目标跟踪与关联
教材对应：第8.1.2.2节

8.2.3 多视角感知与姿态估计
教材对应：第8.1.2.2节

8.2.4 跟踪模式与视觉伺服
教材对应：第8.1.2.2节

8.2.5 光流估计与运动分析
教材对应：第8.1.2.3节

8.2.1 目标检测与实例分割：输出层次

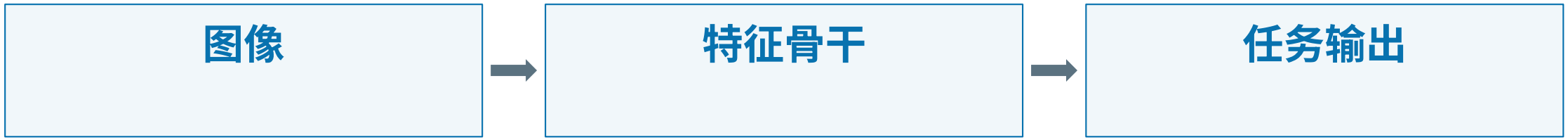
从类别和边界走向身份与姿态

任务	输出	回答的问题
分类	类别	是哪一类对象
检测	类别与二维框	目标在图像哪里
实例分割	每个目标的像素掩膜	哪些像素属于这个实例
跟踪	身份、轨迹与状态	是否仍是同一个对象
姿态估计	关节点或刚体位姿	构型或朝向是什么

连续身份来自时间关联，不能由单帧类别给出。

8.2.1 目标检测与实例分割：卷积神经网络

卷积提取特征，检测头定位目标



- **卷积网络（CNN）**：共享卷积核提取局部模式，多层结构形成层次特征。
- **VGG**：用连续小卷积核堆叠；骨干输出交给分类头或检测头。
- **残差网络（ResNet）**：跳跃连接保留输入，使深层块学习残差。

$$\mathbf{y} = \mathbf{x} + F(\mathbf{x})$$

8.2.1 目标检测与实例分割：区域与单阶段检测

两阶段候选与单阶段YOLO各有路径

模型	候选／检测路径	特征与结果
R-CNN	外部候选区域算法	各区域分别提取特征
Fast R-CNN	外部候选区域算法	整图特征共享，再分类和修框
Faster R-CNN	区域提议网络（RPN）	候选与检测共享骨干
YOLO	整图一次前向的单阶段检测	直接预测类别和框；再筛选输出

YOLO＝“只看一次”；框和类别可继续送入跟踪，单帧检测本身不给连续ID。

8.2.1 目标检测与实例分割：自注意力与集合预测

图像块聚合上下文，查询输出目标集合

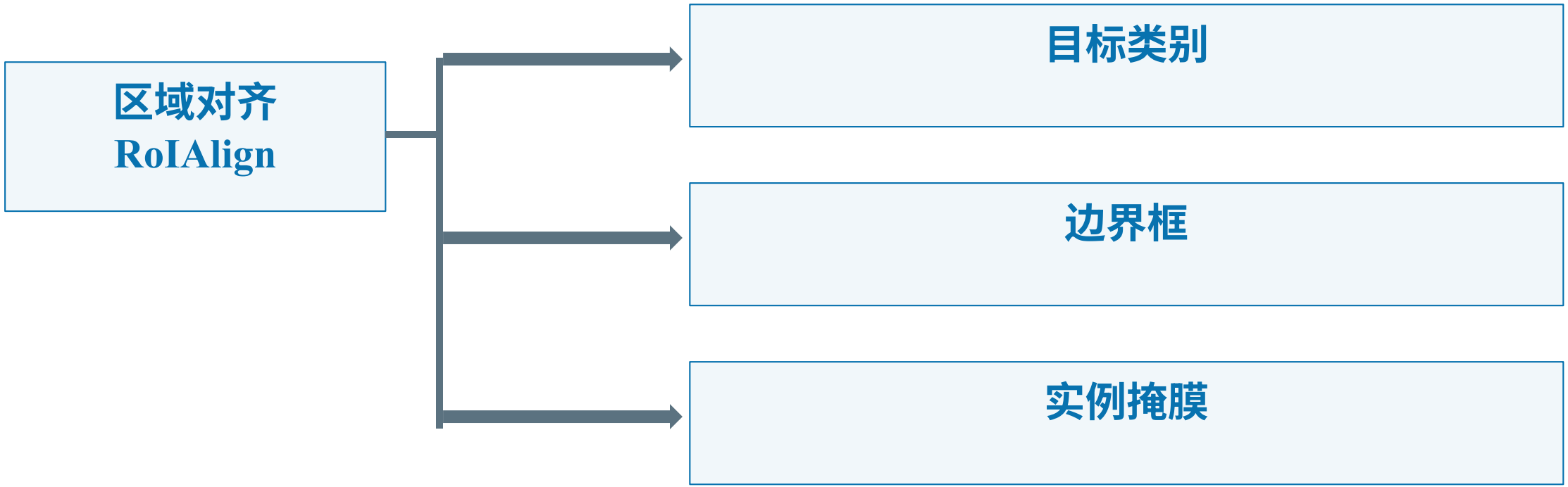
ViT（视觉Transformer）：图像块序列聚合全局上下文。
DETR（检测Transformer）：对象查询直接预测目标集合。



DETR训练采用一对一集合匹配；多余查询输出“无目标”。

8.2.1 目标检测与实例分割：掩膜分支

Mask R-CNN为目标增加掩膜分支



Mask Scoring R-CNN估计掩膜质量；类别分数与边界质量是不同指标。

8.2.1 目标检测与实例分割：原型掩膜

YOLACT共享原型，再组合目标掩膜

共享原型 → 目标系数组合 → 按框裁剪

$$M_i = \mathrm{crop}_{B_i}\left(\sigma\left(\sum_{j=1}^{k} c_{ij} P_j\right)\right)$$

符号： M_i：目标掩膜；P_j：共享原型；c_{ij}：组合系数。
B_i：目标框；σ：压到0–1；k：原型数。

- **YOLACT++：** 可变形卷积等改进；共享计算减少重复开销，边界仍受分辨率影响。

8.2.1 目标检测与实例分割：中心与帧间偏移

用当前中心回指历史目标

CenterTrack：中心式检测与帧间关联；MOT表示多目标跟踪。

两帧＋历史热图 → 中心＋框＋偏移 → 回指并贪心关联

$$\boldsymbol{\delta}_{t\to t-1} = \mathbf{c}_{t-1} - \mathbf{c}_t$$

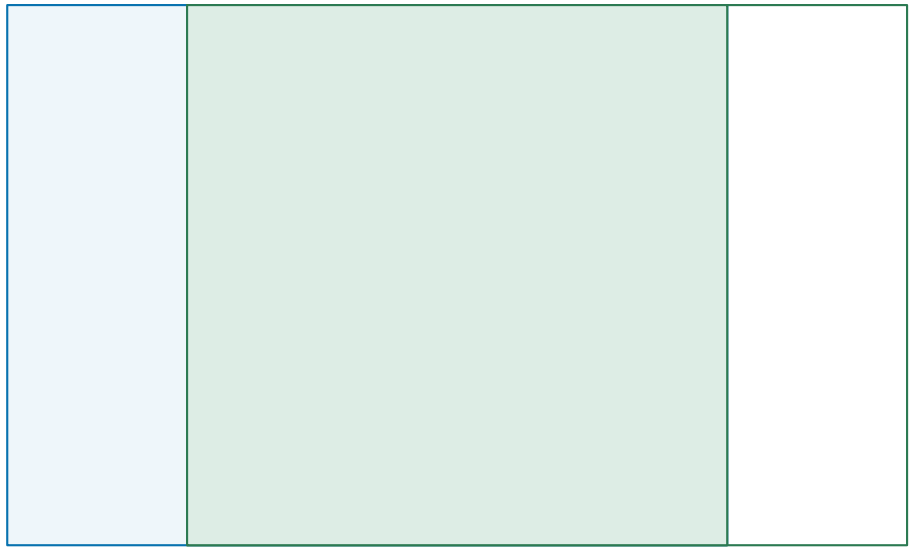
$$\hat{\mathbf{c}}_{t-1} = \mathbf{c}_t + \boldsymbol{\delta}_{t\to t-1}$$

先定义偏移方向，再回指历史中心；短期轨迹保留不等于长期外观重识别。

8.2.1 目标检测与实例分割：重叠算例

框的交并比把位置偏差量化

参考框蓝色；预测框绿色



两个框均为80×60像素
水平偏移20像素

$$\mathrm{IoU} = \frac{|B \cap G|}{|B \cup G|}$$

$$\mathrm{IoU} = \frac{60 \times 60}{2 \times 80 \times 60 - 60 \times 60} = 0.60$$

交并比（IoU）衡量两区域的重叠。
本算例阈值0.5，0.60通过重叠条件。

8.2.1 目标检测与实例分割：输出核验

逐帧看框的位置，也看错检和漏检



图8.3(b) 圆环目标检测

偏移／px	交并比	阈值0.5
0	1.00	通过
20	0.60	通过
40	0.33	未通过

类别、位置与漏检分别核验；
单帧图不给连续身份。

本章目录

目标识别、跟踪与光流估计

第八章 多模态感知与智能决策

8.1 背景意义与图像处理基础

8.2 目标识别、跟踪与光流估计

8.3 定位建图与预测规划

8.4 常用感知传感器

8.5 多模态感知坐标转换

8.6 多模融合与控制决策

8.7 关键工具与框架接口

8.8 实验案例与真机迁移

8.9 习题与课堂讨论

本节 5 个教学单元

8.2.1 目标检测与实例分割
教材对应：第8.1.2.2节

8.2.2 在线目标跟踪与关联
教材对应：第8.1.2.2节

8.2.3 多视角感知与姿态估计
教材对应：第8.1.2.2节

8.2.4 跟踪模式与视觉伺服
教材对应：第8.1.2.2节

8.2.5 光流估计与运动分析
教材对应：第8.1.2.3节

8.2.2 在线目标跟踪与关联：预测、关联与更新

连续身份来自状态递推和数据关联

运动预测 → 检测与关联 → 状态更新 → 轨迹管理

- **SORT：**用卡尔曼滤波和匈牙利分配保持多个目标的连续轨迹。
- **轨迹状态：**新建、确认、暂失与删除，决定身份保留的时间。
- **困难：**交叉、遮挡与相机运动使预测范围重叠，可能引发身份切换。

8.2.2 在线目标跟踪与关联：关联算例

一次分配同时决定两条轨迹的身份

$$\hat{u}_t = u_{t-1} + \dot{u}_{t-1}\Delta t$$

构造输入：A预测140px，B预测120px；检测D1为121px，D2为139px。

中心距离／px	D1：121	D2：139
A：140	19	1
B：120	1	19

A→D2、B→D1总成本2px；相反分配为38px。此例门限取25px。

8.2.2 在线目标跟踪与关联：外观与级联匹配

先用运动排除，再用外观区分

运动门控 → 外观距离 → 级联匹配

$$d_{\text{app}} = 1 - \frac{\mathbf{r}_i^T \mathbf{r}_j}{\|\mathbf{r}_i\| \|\mathbf{r}_j\|}$$

- **外观嵌入：**把目标图像映射成特征向量；余弦距离越小，外观越相似。
- **级联顺序：**最近被观测到的轨迹先关联，随后处理更久未更新的轨迹。

8.2.2 在线目标跟踪与关联：联合检测与嵌入

共享骨干减少重复特征提取

框架	关联依据	计算组织
Deep SORT	运动门控＋外观嵌入	检测后再提取外观
JDE	运动＋共同训练的嵌入	检测与嵌入共享骨干
CenterTrack	中心与帧间偏移	两帧与历史热图联合输入

JDE：联合检测与特征嵌入；一次特征提取不自动保证身份正确。

8.2.2 在线目标跟踪与关联：核相关滤波

模板的相关响应峰给出目标位置

初始化模板 → 搜索图像块 → 相关响应峰 → 更新目标位置

- **核相关滤波（KCF）：** 通过模板与搜索区域的相关响应寻找单个目标。
- **快速傅里叶变换（FFT）：** 把循环相关计算转到频域，利用样本结构减少开销。
- **失效与恢复：** 尺度变化、遮挡和相似背景易漂移；检测可重新初始化模板。

8.2.2 在线目标跟踪与关联：连续跟踪结果

交叉和漏检检验身份连续性

帧	A中心／px	B观测／px	身份解释
1	80	180	初始化A、B
2	110	150	分别更新原身份
3	140	漏检；预测120	B保留预测，未观测
4	170	90	重新关联B，未交换身份

预测能填短缺口，连续身份仍要由新观测确认。

本章目录

目标识别、跟踪与光流估计

第八章 多模态感知与智能决策

8.1 背景意义与图像处理基础

8.2 目标识别、跟踪与光流估计

8.3 定位建图与预测规划

8.4 常用感知传感器

8.5 多模态感知坐标转换

8.6 多模融合与控制决策

8.7 关键工具与框架接口

8.8 实验案例与真机迁移

8.9 习题与课堂讨论

本节 5 个教学单元

8.2.1 目标检测与实例分割
教材对应：第8.1.2.2节

8.2.2 在线目标跟踪与关联
教材对应：第8.1.2.2节

8.2.3 多视角感知与姿态估计
教材对应：第8.1.2.2节

8.2.4 跟踪模式与视觉伺服
教材对应：第8.1.2.2节

8.2.5 光流估计与运动分析
教材对应：第8.1.2.3节

8.2.3 多视角感知与姿态估计：姿态输出

人体关节点与刚体六自由度是两种输出

对象	输出	典型方法／条件
人体骨骼	多个关节点及连接关系	Shotton：深度图→三维关节
人体图像	图像中的关节点位置	HRNet：保留高分辨率特征
刚体目标	三维平移＋三维旋转	已知模型、关键点或深度几何

六自由度（6DoF）指刚体的3个平移与3个旋转自由度。

8.2.3 多视角感知与姿态估计：多视角几何

匹配同一点，才能让两条视线交会



PnP：由已知三维点及其二维投影，求相机与目标之间的相对位姿。

8.2.3 多视角感知与姿态估计：三角测量算例

40像素视差对应3米深度

- **几何条件：**校正双目、水平基线、同步观测同一点；深度沿相机光轴。

- **给定输入：**像素焦距600px，基线0.20m，视差40px。

$$Z = \frac{f_x B}{d}, \qquad d = u_L - u_R$$

$$Z = \frac{600 \times 0.20}{40} = 3.00\ \mathrm{m}$$

8.2.3 多视角感知与姿态估计：重投影与束调整

同时修正相机位姿和三维点

$$\mathbf{r}_{ij} = \mathbf{z}_{ij} - \pi(R_i\mathbf{P}_j + \mathbf{t}_i)$$

i：相机号；j：点号；$\mathbf{z}_{ij}$：像素观测；$\mathbf{r}_{ij}$：重投影残差；π：投影函数。

$$\min_{\{R_i,\mathbf{t}_i,\mathbf{P}_j\}} \sum_{i,j} \|\mathbf{r}_{ij}\|^2$$

- **符号：** R为3×3旋转矩阵，t为3×1平移，三维点为3×1；变换到相机系后，再投影成像素观测。
- **求解条件：** 固定参考与尺度；剔除错误匹配；多视角观测共同约束同一三维点。

8.2.3 多视角感知与姿态估计：学习与多模态融合

不同观测共同约束姿态

观测／方法	提供的约束	使用条件
HRNet／深度人体方法	关节点或部位位置	输出语义与人体模型相符
跨视角Transformer	不同视图之间的特征联系	视图对应、训练分布及同步
视觉＋LiDAR＋IMU	像素、距离与短时运动	外参、时标与误差权重一致

LiDAR：激光测距点云；IMU：惯性测量单元。低置信观测权重降低。

8.2.3 多视角感知与姿态估计：误差结果解释

远处目标的深度对视差更敏感

给定视差	测成较小值	测成较大值
40px → 3.00m	38px → 3.16m	42px → 2.86m
10px → 12.00m	8px → 15.00m	12px → 10.00m

- **计算条件：**焦距600px、基线0.20m；只考察视差误差，其他标定量固定。
- **结果含义：**距离更远时视差更小，相同像素误差造成更大的米制误差。

本章目录

目标识别、跟踪与光流估计

第八章 多模态感知与智能决策

8.1 背景意义与图像处理基础

8.2 目标识别、跟踪与光流估计

8.3 定位建图与预测规划

8.4 常用感知传感器

8.5 多模态感知坐标转换

8.6 多模融合与控制决策

8.7 关键工具与框架接口

8.8 实验案例与真机迁移

8.9 习题与课堂讨论

本节 5 个教学单元

8.2.1 目标检测与实例分割
教材对应：第8.1.2.2节

8.2.2 在线目标跟踪与关联
教材对应：第8.1.2.2节

8.2.3 多视角感知与姿态估计
教材对应：第8.1.2.2节

8.2.4 跟踪模式与视觉伺服
教材对应：第8.1.2.2节

8.2.5 光流估计与运动分析
教材对应：第8.1.2.3节

8.2.4 跟踪模式与视觉伺服：跟踪模式

固定相机、云台和机体改变不同对象

模式	主动改变什么	对目标的作用
固定观察	传感器位姿不主动调整	连续检测与标注
云台动态跟踪	相机视轴方向	保持目标在视场或中心
机体视觉反馈	载具位置与姿态	改变视点、朝向与相对距离

“静态”修饰观测平台；目标仍可以运动。

8.2.4 跟踪模式与视觉伺服：视觉伺服信息路径

误差由三维位姿或图像特征构成

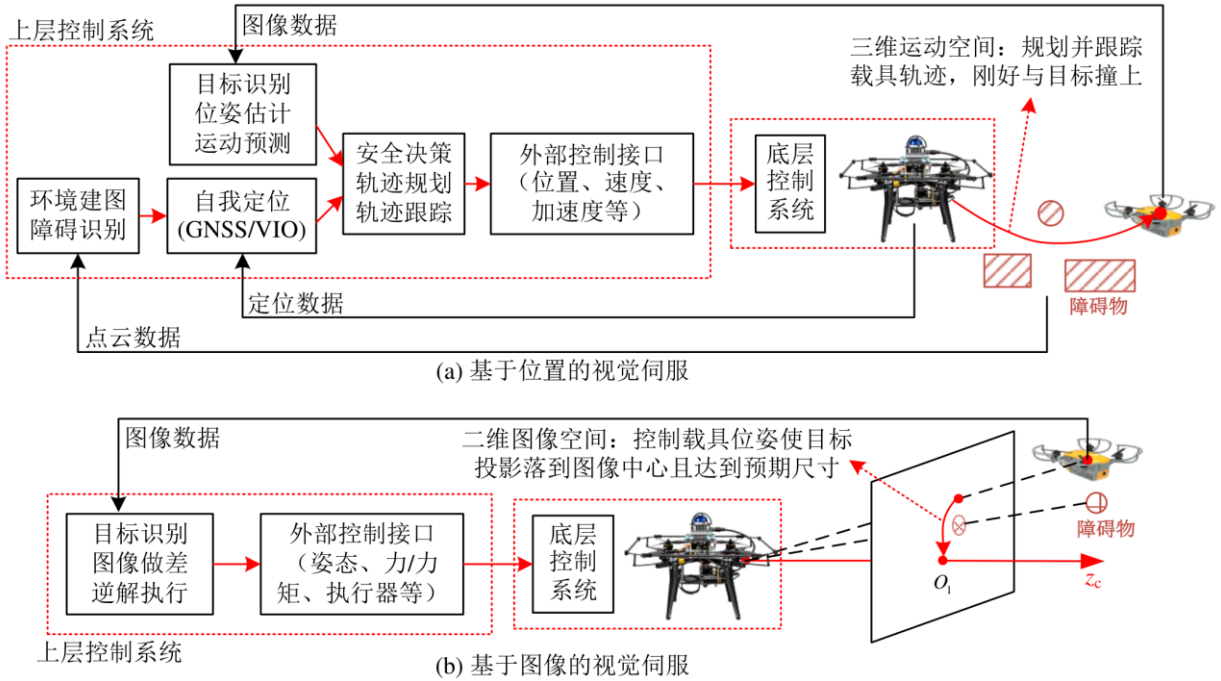
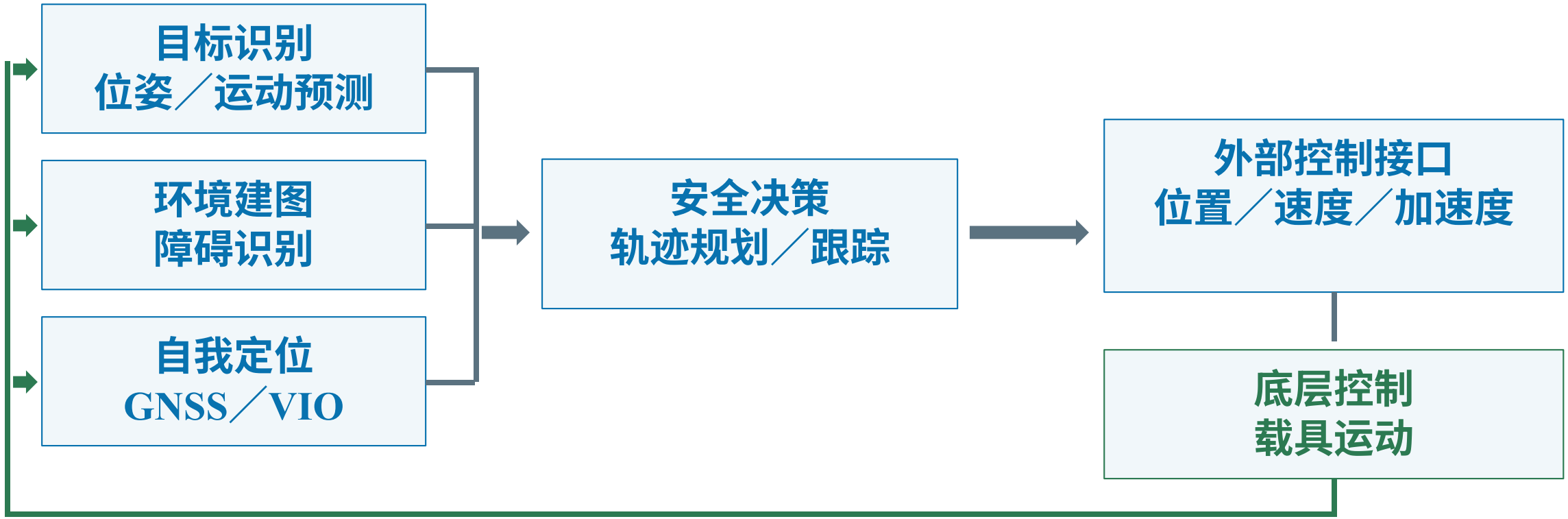


图8.2 视觉伺服控制方法

(a)以三维位姿误差为基础；(b)以图像特征误差为基础。

8.2.4 跟踪模式与视觉伺服：位姿误差控制链

三维状态经过规划后进入位置接口

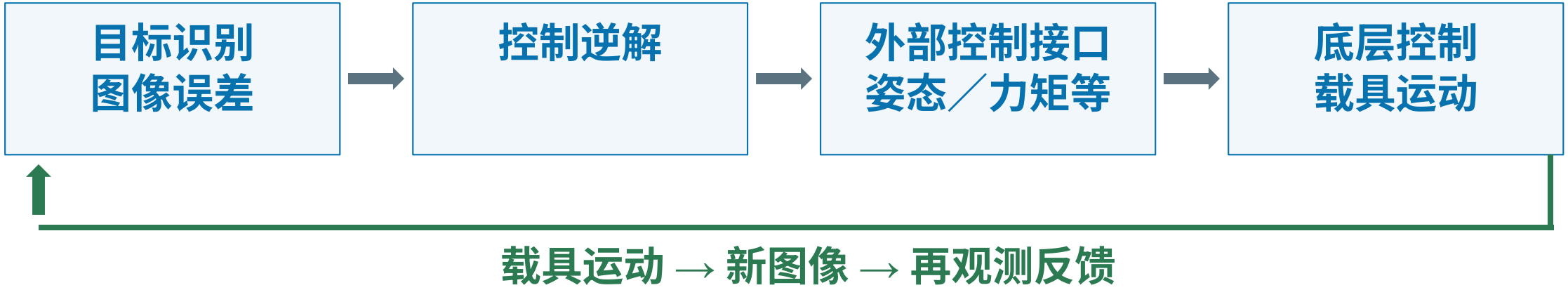


反馈：载具运动改变图像、定位与点云；(a)还集成了环境建图和轨迹规划。

8.2.4 跟踪模式与视觉伺服：图像误差控制链

图像误差经逆解形成执行目标

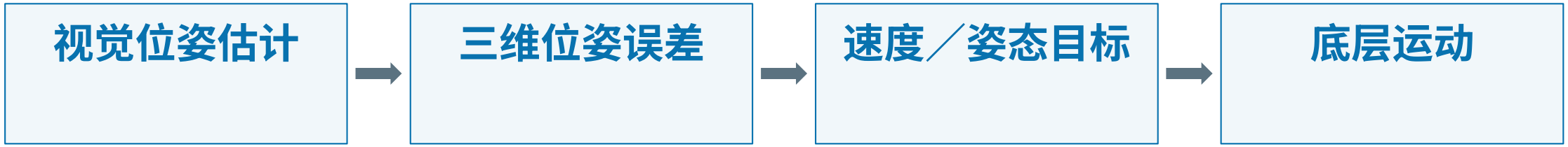
图8.2(b)：当前特征与期望特征做差，不先要求完整三维地图。



- **图像目标：**投影保持在期望位置与尺度；图像误差可包含多个特征。
- **接口边界：**姿态、力矩或速度由系统设计决定；图像型视觉伺服按误差特征定义。

8.2.4 跟踪模式与视觉伺服：基于位置的视觉伺服

相对位姿误差指导三维运动



$$\mathbf{e}_p = \mathbf{s}_p - \mathbf{s}_p^*$$

- **位置型视觉伺服（PBVS）：** 误差特征包含相对目标的三维平移与旋转。
- **规划接入：** 任务需要避障时可附加规划器；基本位姿反馈不以全局建图为前提。

8.2.4 跟踪模式与视觉伺服：基于图像的视觉伺服

图像雅可比连接特征变化与速度

$$\mathbf{e} = \mathbf{s} - \mathbf{s}^*, \qquad \dot{\mathbf{s}} = \mathbf{L}(\mathbf{s}, Z)\mathbf{v}_c$$

$\mathbf{s}, \mathbf{s}^*$：当前、期望特征；$\mathbf{e}$：误差；Z：深度；$\mathbf{v}_c$：相机速度。

$$\mathbf{v}_c = -\lambda \hat{\mathbf{L}}^{+}\mathbf{e}$$

$\hat{\mathbf{L}}^{+}$：估计交互矩阵的伪逆；λ：正增益。

- **图像型视觉伺服（IBVS）：**图像特征构成误差，交互矩阵把特征变化率连接到相机速度。
- **符号与条件：**特征单位由选择决定；相机速度含平移与旋转，增益为正，局部几何有效。

8.2.4 跟踪模式与视觉伺服：横向控制一步

目标偏右40像素，相机向右0.2米每秒

简化条件：目标静止、固定深度2m；相机右下前坐标，只做水平平移。

$$\dot{u} = -\frac{f_x}{Z} v_{c,x}, \qquad v_{c,x} = \frac{\lambda Z}{f_x} e_u$$

$$v_{c,x} = \frac{1 \times 2}{400} \times 40 = 0.20 \ \mathrm{m/s}$$

$$e_1 = 40 - \frac{400}{2} \times 0.20 \times 0.10 = 36 \ \mathrm{px}$$

像素焦距400px；增益1/s；采样0.10s；正速度为相机向右。

8.2.4 跟踪模式与视觉伺服：闭环迭代结果

误差按0.9倍递减，速度同步降低

$$e_{k+1} = (1 - \lambda \Delta t) e_k = 0.9 e_k$$

更新步数	图像误差／px	相机右向速度／m/s
0	40.00	0.200
1	36.00	0.180
10	13.95	0.0697
20	4.86	0.0243

离线迭代：固定深度、静止目标、无延迟、无速度限幅。

8.2.4 跟踪模式与视觉伺服：端到端及混合策略

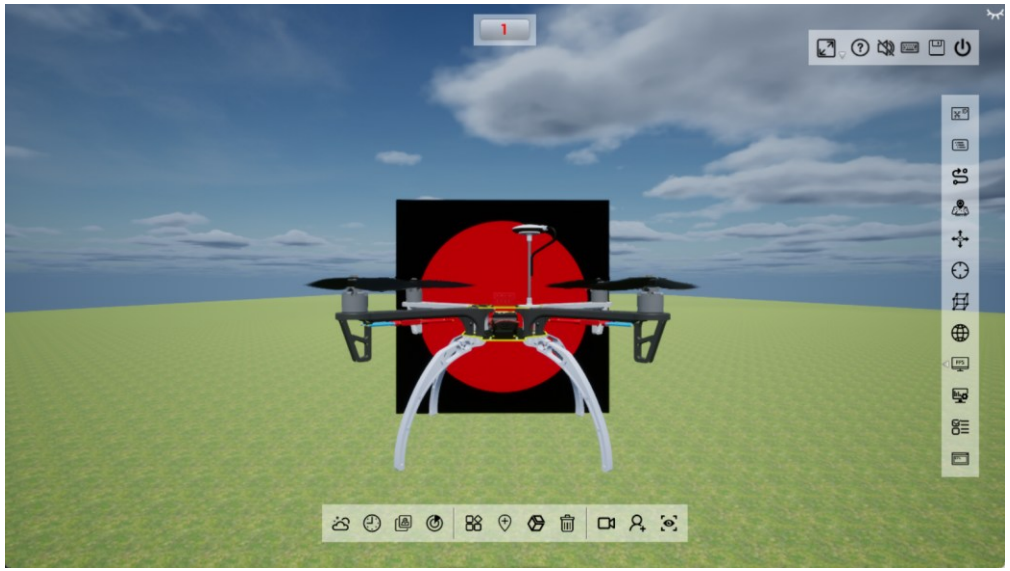
学习映射与几何反馈各有条件

路线	误差或映射	关键条件
PBVS	三维相对位姿误差	目标模型、标定与深度
IBVS	图像特征误差	可见特征、局部交互矩阵
端到端	图像直接映射到动作	训练分布、时延与输出约束

混合策略：到达观察区后再切入图像反馈；切换按置信、距离与限值。

8.2.4 跟踪模式与视觉伺服：圆板跟随源码

中心和尺寸分别生成速度参考



图：圆板跟随例程说明场景

前视对齐、近水平时：
中心偏右／偏下→右／下速度
框高小于初值→前向速度
水平速度随航向；竖直按NED
当前实现为比例项P＋15px死区
丢失目标时跳过发送；Esc结束程序。
退出停流、停车与降落是不同动作。

学生入口：[安装目录]/RflySimAPIs/8.RflySimVision/1.BasicExps/1-VisionCtrlDemos/e6_Circle-follow/circle_follow_autoCtrl.py

本章目录

目标识别、跟踪与光流估计

第八章 多模态感知与智能决策

8.1 背景意义与图像处理基础

8.2 目标识别、跟踪与光流估计

8.3 定位建图与预测规划

8.4 常用感知传感器

8.5 多模态感知坐标转换

8.6 多模融合与控制决策

8.7 关键工具与框架接口

8.8 实验案例与真机迁移

8.9 习题与课堂讨论

本节 5 个教学单元

8.2.1 目标检测与实例分割
教材对应：第8.1.2.2节

8.2.2 在线目标跟踪与关联
教材对应：第8.1.2.2节

8.2.3 多视角感知与姿态估计
教材对应：第8.1.2.2节

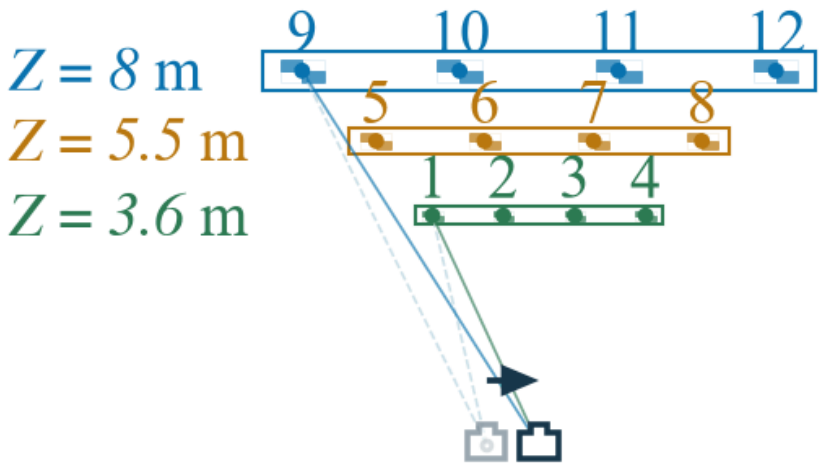
8.2.4 跟踪模式与视觉伺服
教材对应：第8.1.2.2节

8.2.5 光流估计与运动分析
教材对应：第8.1.2.3节

8.2.5 光流估计与运动分析：图像运动

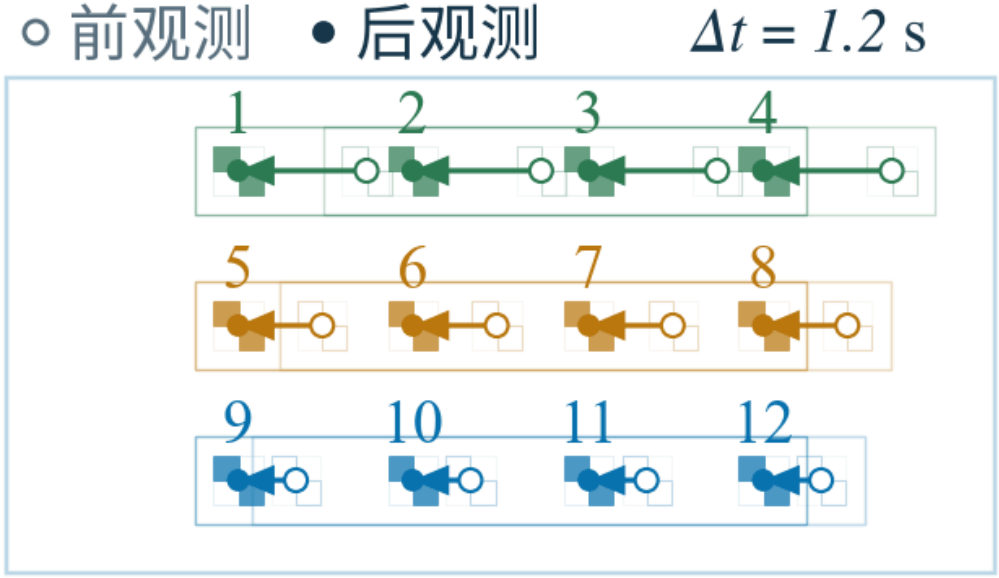
光流描述像素位移，物理速度另需尺度

相机横移：像点反向，近点位移更大



两观测间相机右移

区间均速 $\bar{v}_x = +0.594$ m/s



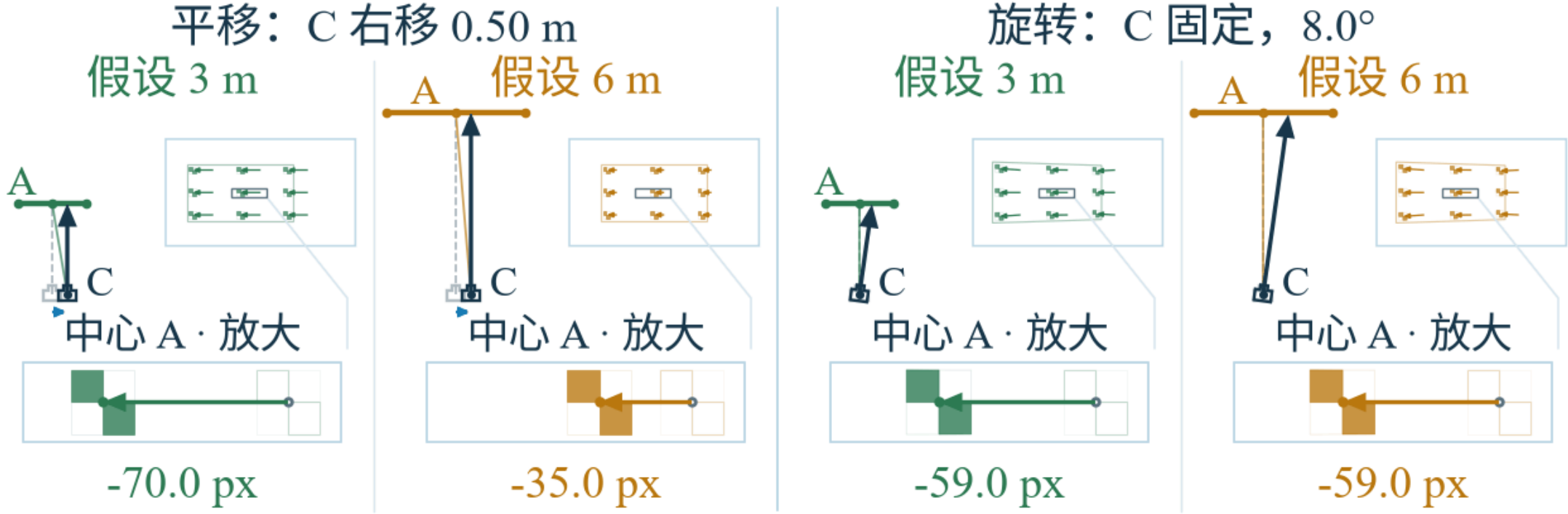
$1: \Delta u = -83.16$ px

图：静止场景中相机纯横移，像点反向移动，近点位移更大。

8.2.5 光流估计与运动分析：相机运动来源

平移与旋转都能产生图像光流

同一初始射线上的近、远点：平移位移不同，纯旋转位移重合。



图：光心右移0.50 m与原地转动8°的投影对照。

8.2.5 光流估计与运动分析：亮度约束与可观测性

一个像素约束不足以确定二维运动

$$I(u, v, t) \approx I(u + \Delta u, v + \Delta v, t + \Delta t)$$

$$I_u \Delta u + I_v \Delta v + I_t \Delta t \approx 0$$

I为亮度；I_u, I_v, I_t分别为水平、垂直、时间偏导数；Δ表示对应增量。

- **亮度一致与小运动：**图像结构可对应，光照变化较小，位移足以用一阶近似。

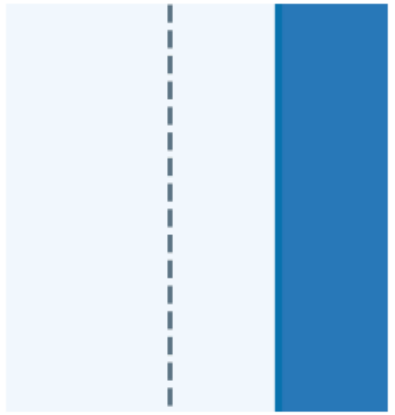
- **孔径问题：**单方向边缘难以确定沿边方向位移；角点和邻域约束补足信息。

8.2.5 光流估计与运动分析：孔径问题

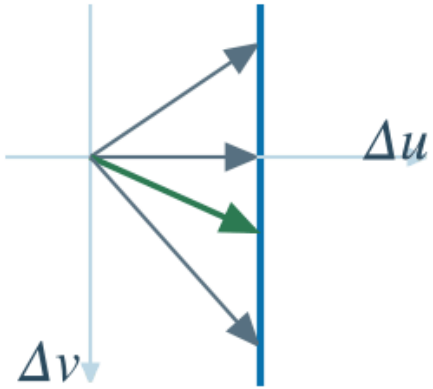
直边的切向歧义与角点的二维约束

垂直边缘只限制水平位移；角点的两个独立方向确定二维位移。

直边：切向不确定



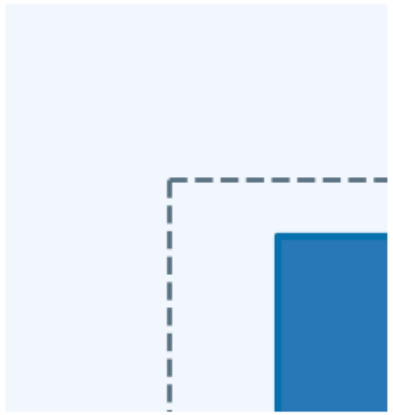
放大邻域



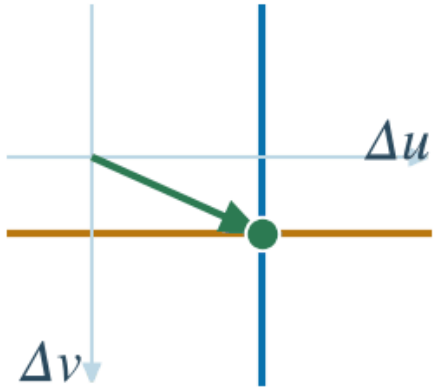
位移约束

同一水平位移，多种切向候选

角点：两方向约束



放大邻域



位移约束

唯一交点 (2.0, 1.0) px

图：与局部算例一致的位移 (2, 1) px。

8.2.5 光流估计与运动分析：稠密与稀疏方法

不同方法用不同约束补足运动信息

方法	补充约束	典型输出／边界
Horn–Schunck	全场亮度一致＋平滑	稠密；运动边界可能平滑
Farneback	局部二次多项式近似	稠密；依赖局部纹理与尺度
Lucas–Kanade	邻域点共享局部运动	常用于稀疏特征点追踪

稠密覆盖像素，稀疏集中于特征；光照、遮挡和大位移共同影响估计。

8.2.5 光流估计与运动分析：局部算例

三个梯度约束解出位移2与1像素

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

$$(A^T A)\boldsymbol{\delta} = A^T \mathbf{b}, \qquad \boldsymbol{\delta} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \text{ px}$$

给定的是构造梯度；两个独立方向才可确定二维位移。

8.2.5 光流估计与运动分析：学习光流

从学习匹配到相关体迭代更新

网络	核心组织	输出与使用条件
FlowNet	学习图像对特征与匹配	监督训练输出二维流场
PWC-Net	金字塔＋扭曲＋代价体	多尺度估计剩余运动
RAFT	全对相关＋循环迭代更新	流场逐步修正；算力依配置

学习方法仍受遮挡、训练分布与分辨率影响；输出不自动具有米制尺度。

8.2.5 光流估计与运动分析：深度与速度算例

四像素地面反向运动对应0.2米每秒

条件：下视平地、深度2m、焦距400px、间隔0.10s；已扣除相机旋转。
图像右向为正；地面点位移−4px。

$$v_{c,x} = -\frac{Z}{f_x}\frac{\Delta u}{\Delta t}$$

$$v_{c,x} = -\frac{2}{400}\frac{-4}{0.10} = 0.20\ \mathrm{m/s}$$

深度恢复尺度，角速度补偿旋转；再经外参转换到机体速度。

8.2.5 光流估计与运动分析：速度与积分结果

短时速度有尺度，位置仍累积误差

区间末端／s	图像位移／px	速度／m/s	累计位移／m
0.10	−4	+0.20	0.02
0.20	−4	+0.20	0.04
0.30	−4	+0.20	0.06

另设：速度有恒定偏差0.01m/s，持续60s。下面计算由此产生的位置偏差。

$$\Delta x_{\text{bias}} = 0.01 \times 60 = 0.60\ \text{m}$$

积分把小速度偏差积成长期位置漂移。

8.2.5 光流估计与运动分析：三维运动恢复

深度反投影后还要扣除相机自运动

像素对应＋深度 → 各自相机三维点 → 共同参考系 → 空间运动

$$\mathbf{P}_t^w = R_{c,t}^w(Z_t K^{-1}\tilde{\mathbf{p}}_t) + \mathbf{t}_{c,t}^w$$

w：固定参考系；R、$\mathbf{t}$：相机系到w的旋转与平移。

$$\mathbf{v}^w = \frac{\mathbf{P}_{t+1}^w - \mathbf{P}_t^w}{\Delta t}$$

R为3×3，t与P为3×1；齐次像素为3×1，内参和光轴深度参与反投影。

练习与讨论

检测、身份和姿态分别缺了什么

- **问题1：**两目标交叉后框仍完整，A、B编号却交换。哪一环节错误？怎样改善？
- **问题2：**HRNet给出一组人体关节点，能否直接当成一个刚体的六自由度位姿？
- **问题3：**固定目标偏右40px，相机向左平移。图像误差会增大还是减小？为什么？

参考要点

先核对输出，再选关联和反馈方法

- **答案1：**身份切换发生在关联环节；加运动门控、外观信息与短时轨迹保留，重新核验ID连续性。

- **答案2：**人体关节点描述骨骼构型；刚体位姿描述三维平移和旋转，需要目标模型与几何约束。

- **答案3：**相机向左使静止目标更偏右，误差增大；本简化例中相机应向右平移。

练习与讨论

把深度和时间代入光流速度

题目：下视静止平地，深度1.5m，焦距300px，间隔0.05s。
旋转补偿后水平位移−3px，求相机右向速度。

$$v_{c,x} = -\frac{1.5}{300}\frac{-3}{0.05} = 0.30\ \mathrm{m/s}$$

- **参考答案：**相机向右0.30m/s；该段位移0.015m，位移和速度单位不同。

- **条件改变：**若光流全由相机旋转造成，先补偿旋转，平移速度应接近零。

本节小结

从目标观测走向可用于决策的状态

- **目标输出：**类别、框、实例掩膜、连续身份和姿态各解决不同问题。
- **几何与反馈：**关节点不等于刚体位姿；位置型／图像型按误差特征区分。
- **运动尺度：**光流加深度、时间与自运动补偿，才可恢复有尺度的速度。
- **易错点：**固定平台仍能观察动态目标；短期预测不是新观测；积分仍可能漂移。