智能无人系统控制实践

第八章 多模态感知与智能决策

作者：戴训华 中南大学

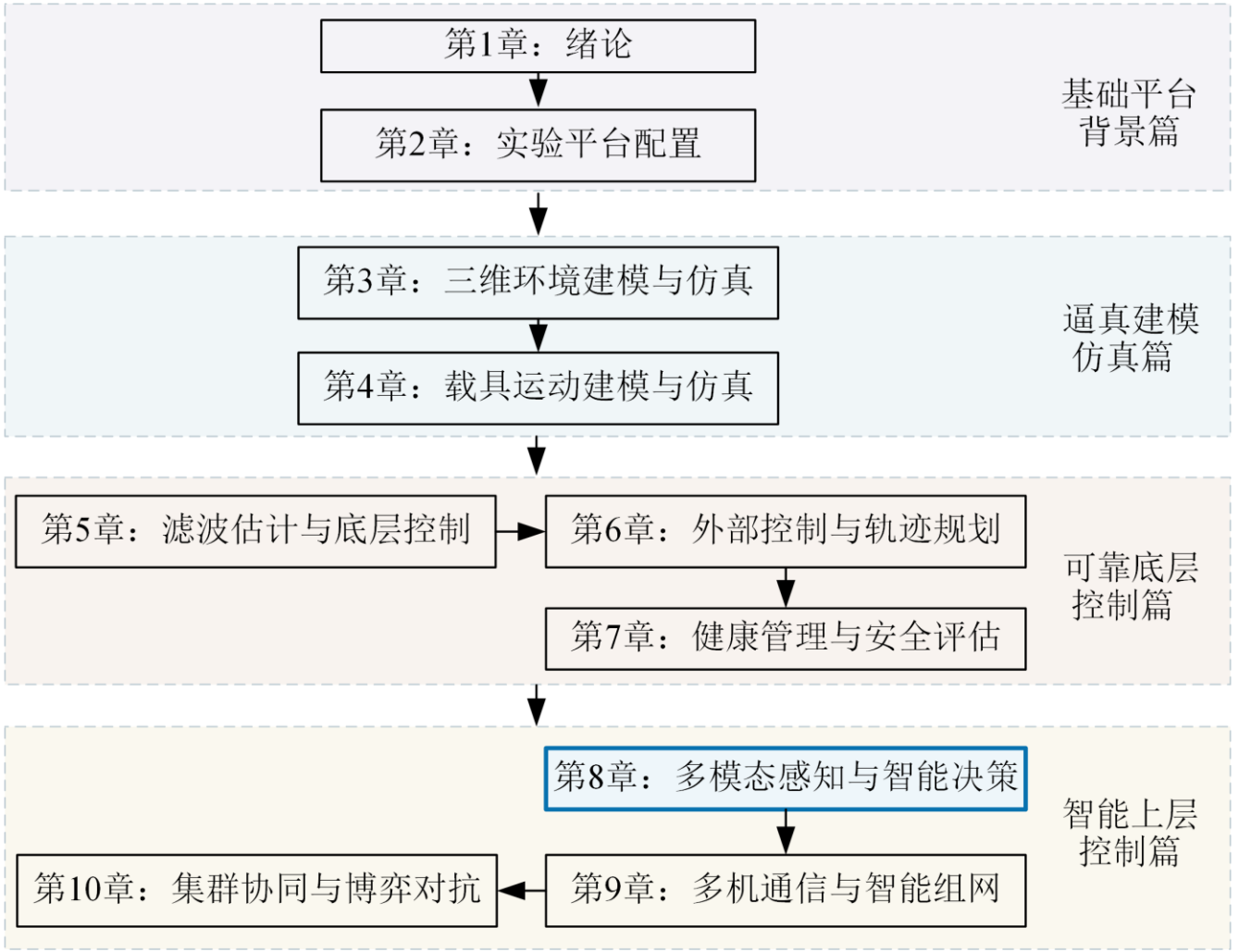
主页：https://faculty.csu.edu.cn/daixunhua

邮箱：dai.xh@csu.edu.cn

平台官网：https://rflysim.com

配套教材





教材：戴训华，全权，任雪峰. 智能无人系统控制实践[M]. 北京：电子工业出版社，2026.

本章学习目标

学完第八章能够完成什么

1. 从图像中提取特征，解释识别、跟踪、光流与建图的关系。
2. 读懂传感器数据，并把像素、相机与世界坐标正确联系起来。
3. 用多模态信息估计状态，形成预测、规划与控制决策。
4. 搭建取图与控制流程，用可观察结果检验仿真和迁移效果。

本章目录

背景意义与图像处理基础

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 4 个教学单元

8.1.1 背景意义与基本原理
教材对应：第8.1.1节

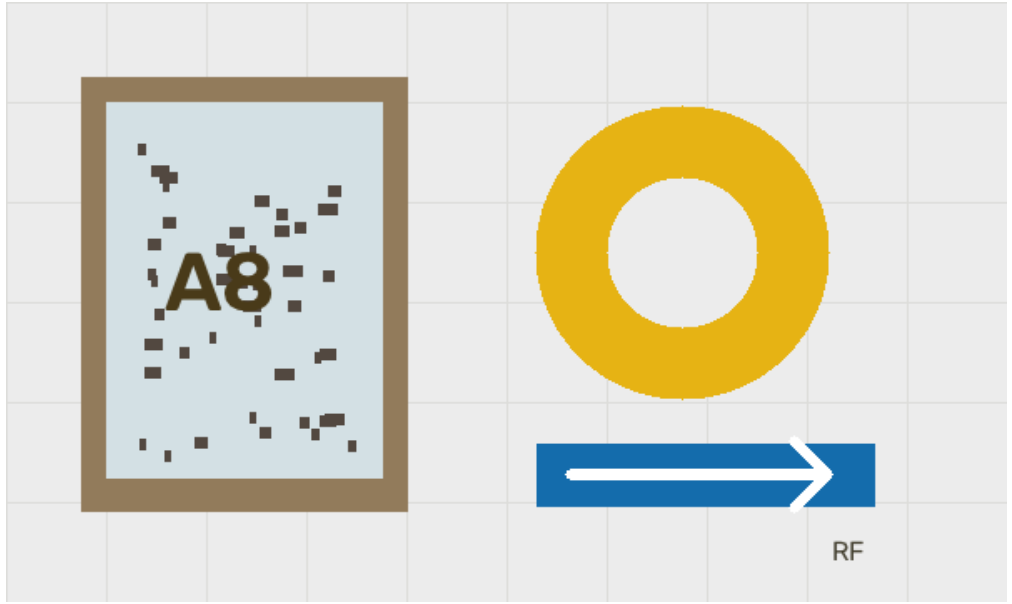
8.1.2 图像预处理与边缘检测
教材对应：第8.1.2.1节（1）–（2）

8.1.3 轮廓提取与图像金字塔
教材对应：第8.1.2.1节（3）–（4）

8.1.4 特征检测、描述与图像配准
教材对应：第8.1.2.1节（5）–（6）

8.1.1 背景意义与基本原理：任务动机

会沿轨迹运动之后，还要知道该往哪里走



场景：寻找并接近黄色标记

已有控制：给定参考，调节运动误差。

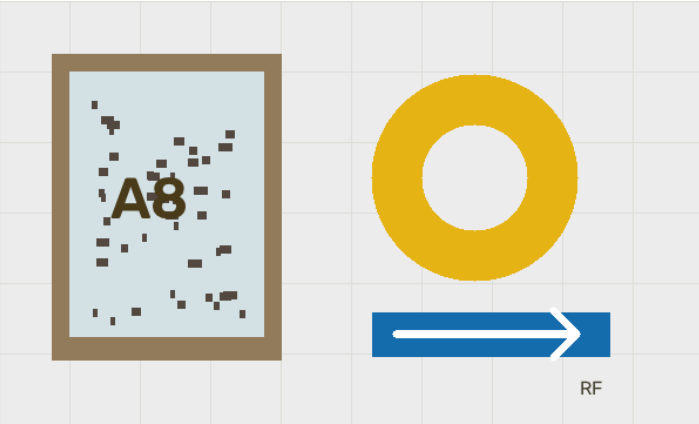
感知新增：目标在哪里？前方有什么？

决策新增：追踪、绕行，还是停下来重观测？

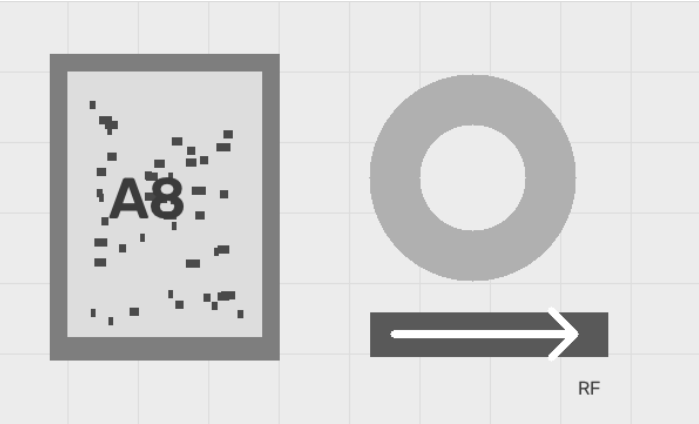
控制让动作可靠，感知与决策让动作有依据。

8.1.1 背景意义与基本原理：原始观测

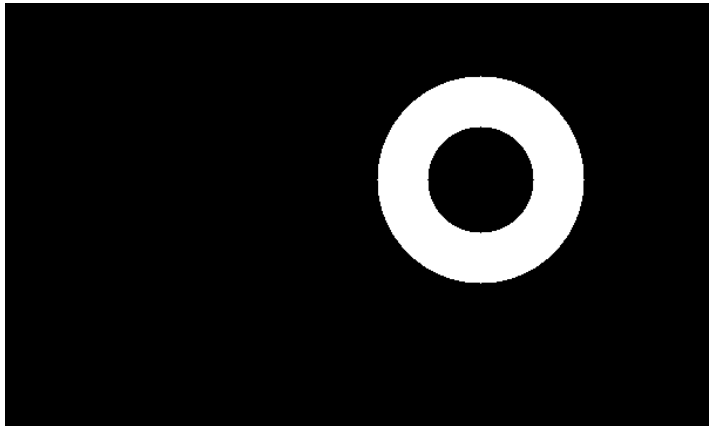
相机先得到像素，距离与含义还需要计算



彩色：保留颜色信息



灰度：只保留亮度信息



掩膜：满足阈值的位置

读图问题：三幅图空间位置相同，但每个像素表达的意义不同。

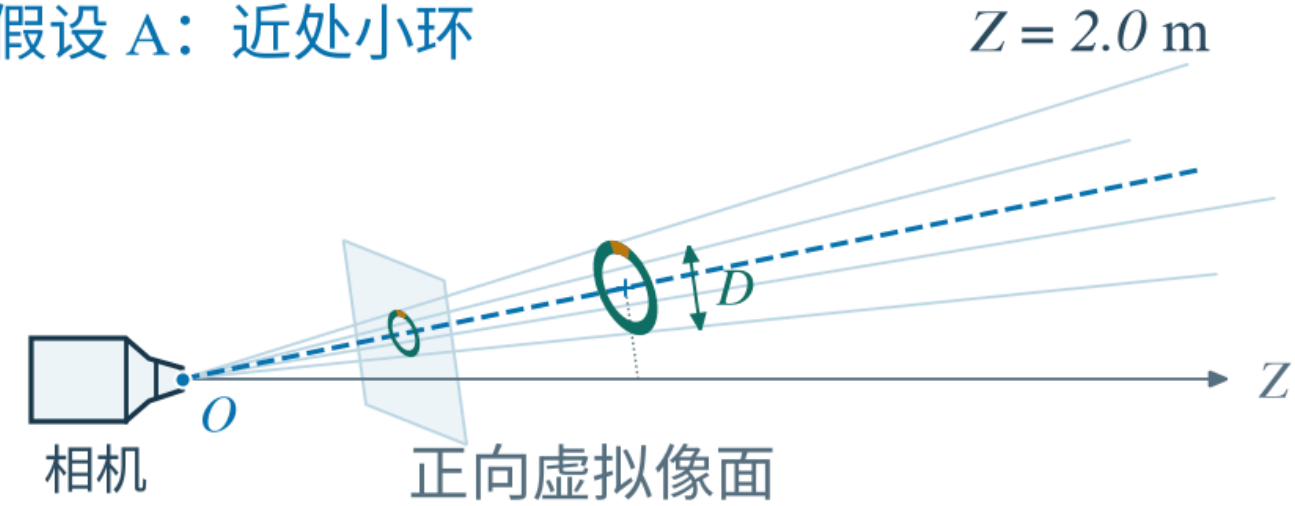
重要区分：黄色区域不是米制位置；颜色相近也不保证属于同一物体。

8.1.1 背景意义与基本原理：投影与深度

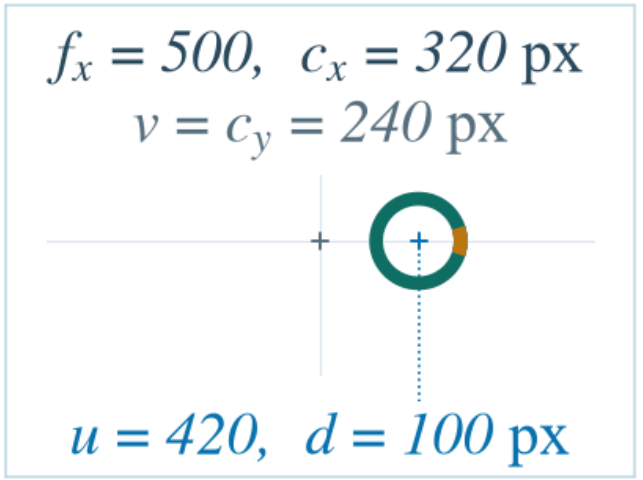
同一像素，可能对应不同距离的物体

一个像素只定一条射线

假设 A：近处小环



投影始终相同



近：$Z = 2,\ X = 0.4,\ D = 0.4$ m

$X/Z = (420 - 320)/500 = 0.2$

远：$Z = 4,\ X = 0.8,\ D = 0.8$ m

$D/Z = 100/500 = 0.2$

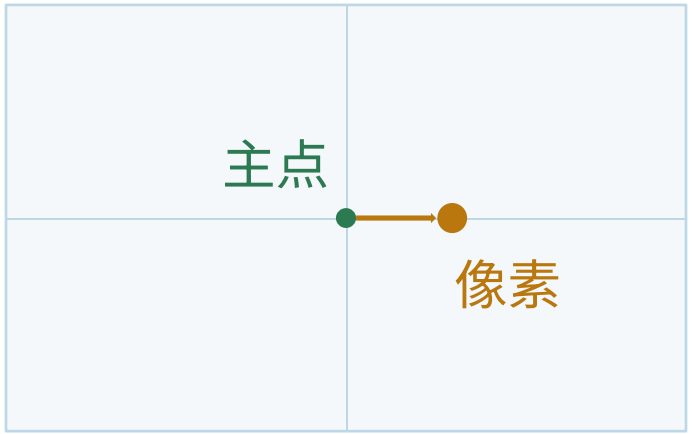
图：相同的投影可以来自不同的距离与尺寸，深度需要额外约束

8.1.1 背景意义与基本原理：空间测量

已知方向，再补深度，才能得到空间位置

理想针孔、已校正像素；本例深度沿相机光轴测量。

图像中的位置



$$u = 420, \quad c_x = 320$$
$$f_x = 500\ \text{px}$$

方向 + 深度 → 相机点

$$\frac{X_c}{Z_c} = \frac{u - c_x}{f_x} = \frac{100}{500} = 0.2$$
$$Z_c = 2\ \text{m} \Rightarrow X_c = 0.4\ \text{m}$$

像素在水平中心线上，本例竖向坐标为0。

再按安装旋转与偏移，转到机体系。

$$^{b}\mathbf{p} = R_c^b \ {}^{c}\mathbf{p} + {}^{b}\mathbf{t}_c$$

8.1.1 背景意义与基本原理：信息层次

从像素到动作，中间要形成可用的状态描述

观测
图像、深度
→ 特征
目标、关键点
→ 估计
位置、速度
→ 决策与控制
参考、动作

例子： 黄色像素 → 目标中心 → 相对方向 → 转向或接近。

接口要带： 数据含义、坐标系、单位、时间戳和有效标志。

8.1.1 背景意义与基本原理：任务边界

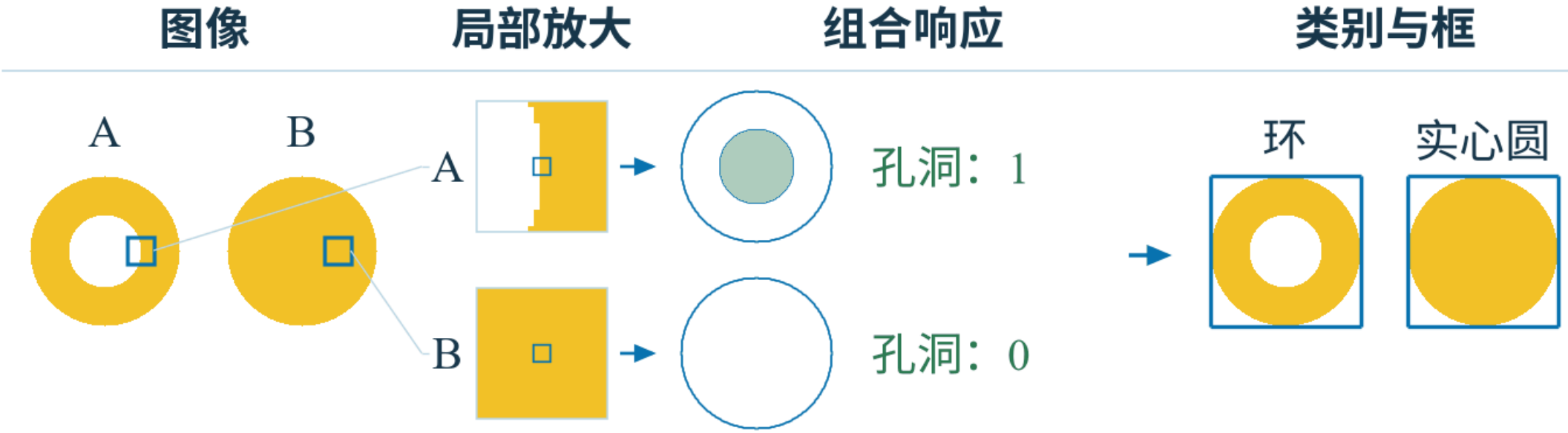
从识别跟踪到预测规划，三类任务各有侧重

任务	回答的问题	输出示例
识别与跟踪	是什么？图中哪里？下一帧在哪？	类别、框、掩膜、轨迹
定位与建图	自己在哪里？周围结构怎样？	位姿、路标、占据地图
预测与规划	接下来会怎样？该怎样走？	未来轨迹、可行路径

这些不是必须串联的固定套餐；任务需求决定保留哪些环节。

8.1.1 背景意义与基本原理：图像特征

颜色相近时，还要结合形状与局部结构



同色不等于同类：边缘与孔洞提供结构线索

固定滤波与区域特征：特征响应示意

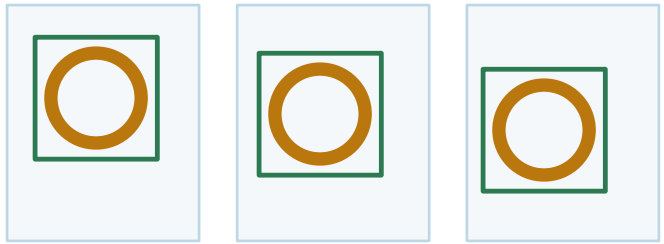
图：同色目标仍有结构差异，联合特征为类别与位置判断提供依据

8.1.1 背景意义与基本原理：训练与推理

模型权重从数据中学习，能力要到新场景检验

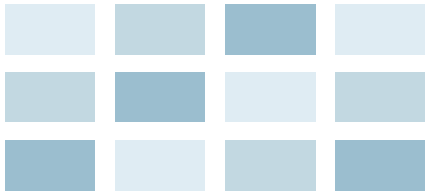
以标记识别为例：离线学习参数，在线只输入新图像。

图像与标注



标注给出类别与范围；
采集覆盖不同背景、
距离和光照。

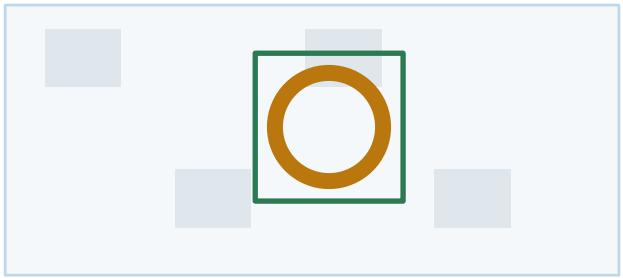
学习与权重



保存的模型参数

比较预测与标注，
按误差调整参数；
权重用于后续推理。

新场景推理



用权重处理新图像，
输出类别与目标框；
不读取该帧正确答案。

按独立场景或采集序列划分训练、验证与测试，避免相邻帧跨组。

8.1.1 背景意义与基本原理：技术全景

沿教材知识图，把“看见”逐步接到“行动”

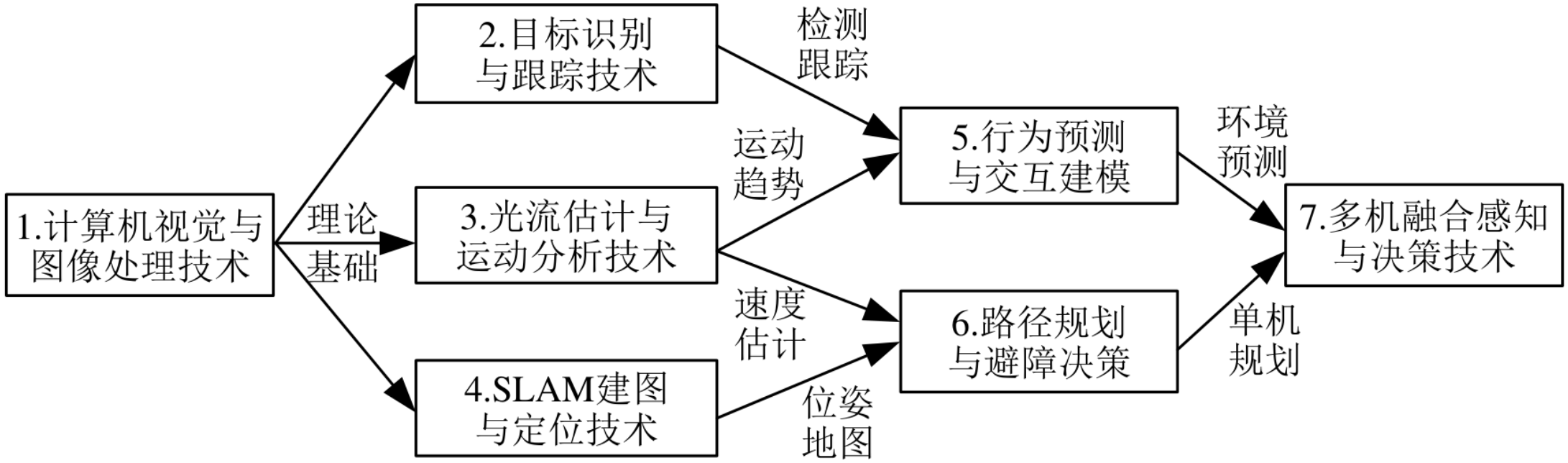
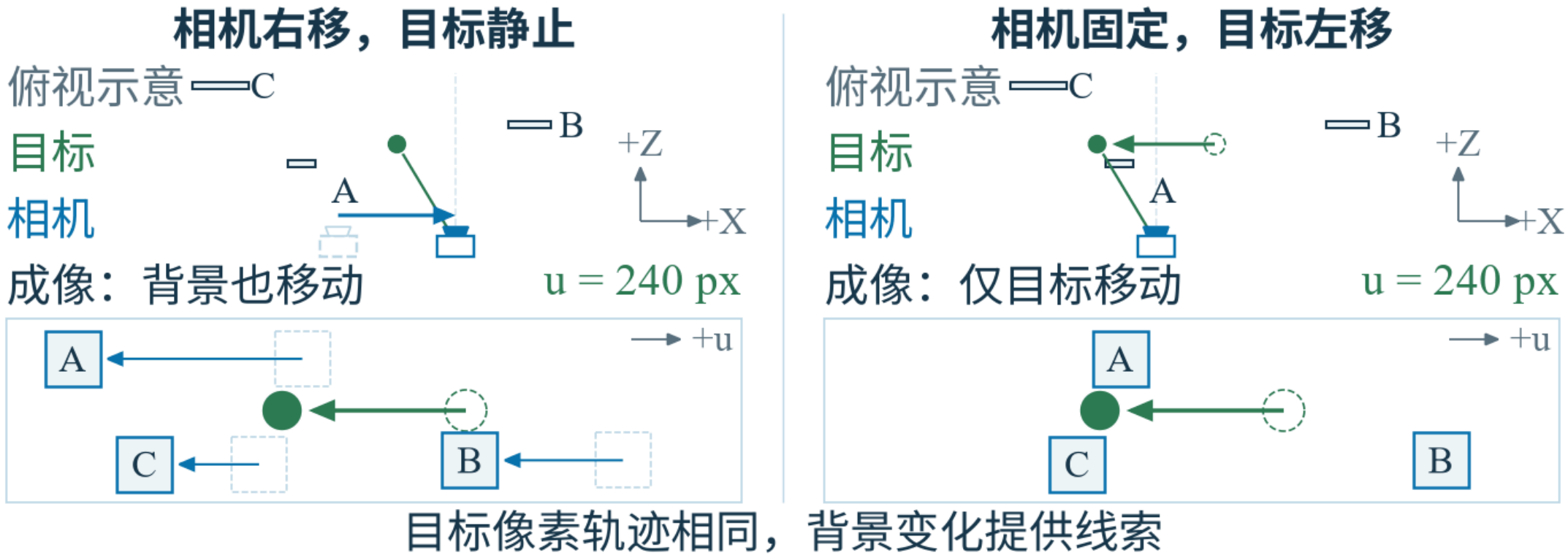


图8.1 多模态感知与智能决策关键技术

SLAM：同步定位与建图，为规划提供位姿与地图。

8.1.1 背景意义与基本原理：跨帧观测

图中的目标移动了，是目标动还是相机动



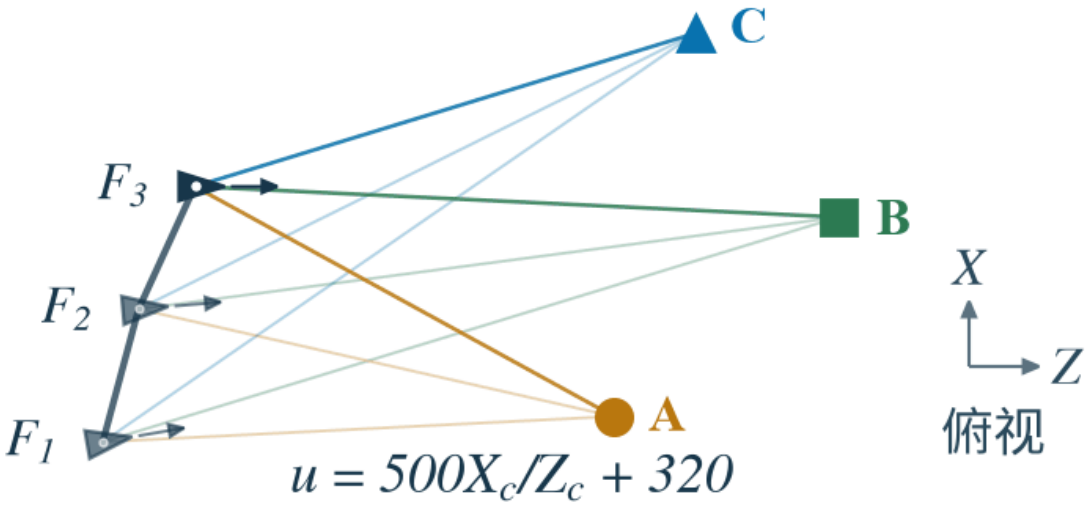
图：相同的目标像素移动，可能来自相机运动，也可能来自目标运动

8.1.1 背景意义与基本原理：定位建图

多个视角共同约束相机位置与环境结构

多帧共同约束

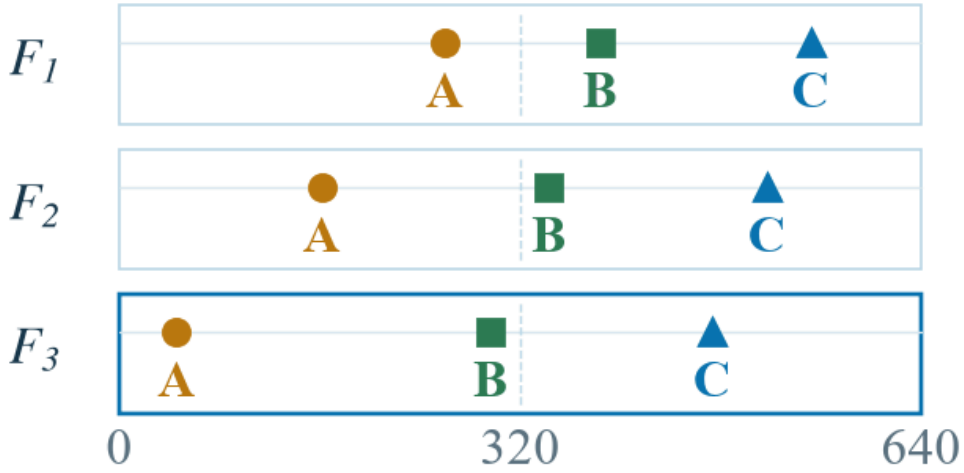
待估：相机位姿 + 路标位置



$$u = 500X_c/Z_c + 320$$

图像中线 $v = 240$

同名点，跨帧对应



多帧约束同一组未知量

相机与地图一起估计；纯单目仍缺绝对尺度

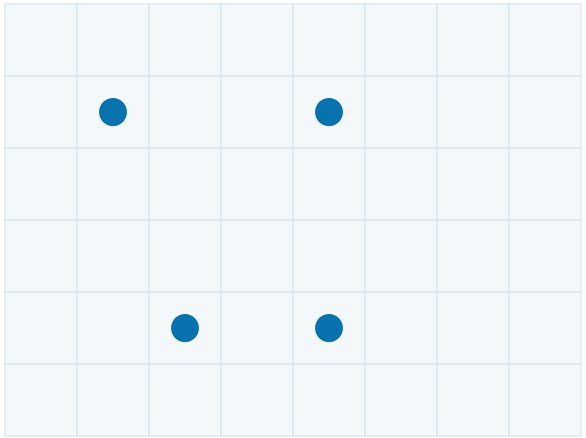
图：跨帧对应约束相对结构，恢复米制尺度还需额外信息

8.1.1 背景意义与基本原理：地图用途

地图能帮助定位，也要能回答哪里可以通过

同一场景、每格1 m；地图保存的信息决定它能回答的问题。

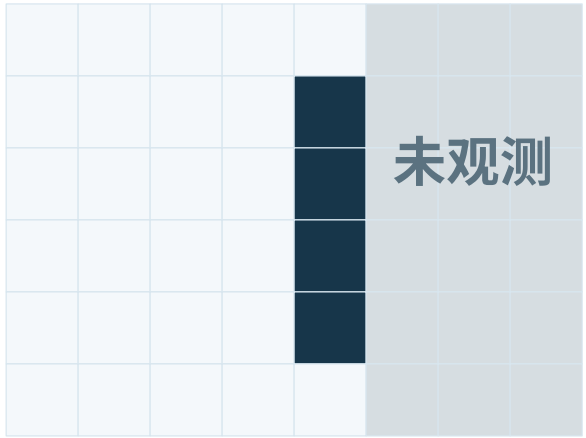
稀疏路标



对应点支持重定位

点之间没有表面定义，
不能直接查询通道净空。

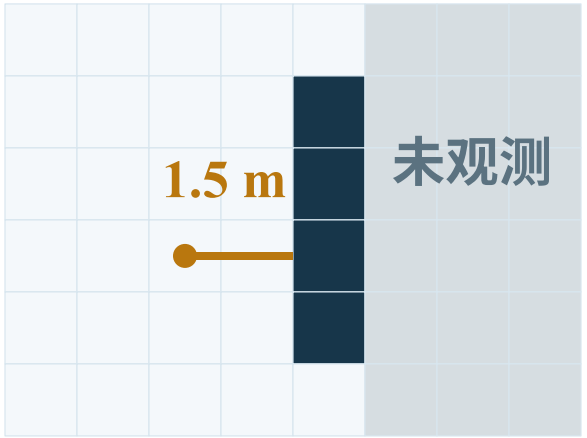
占据与未知



区分占据、空闲与未知

灰格没有观测证据，
不能直接当作可通行。

到障碍的距离

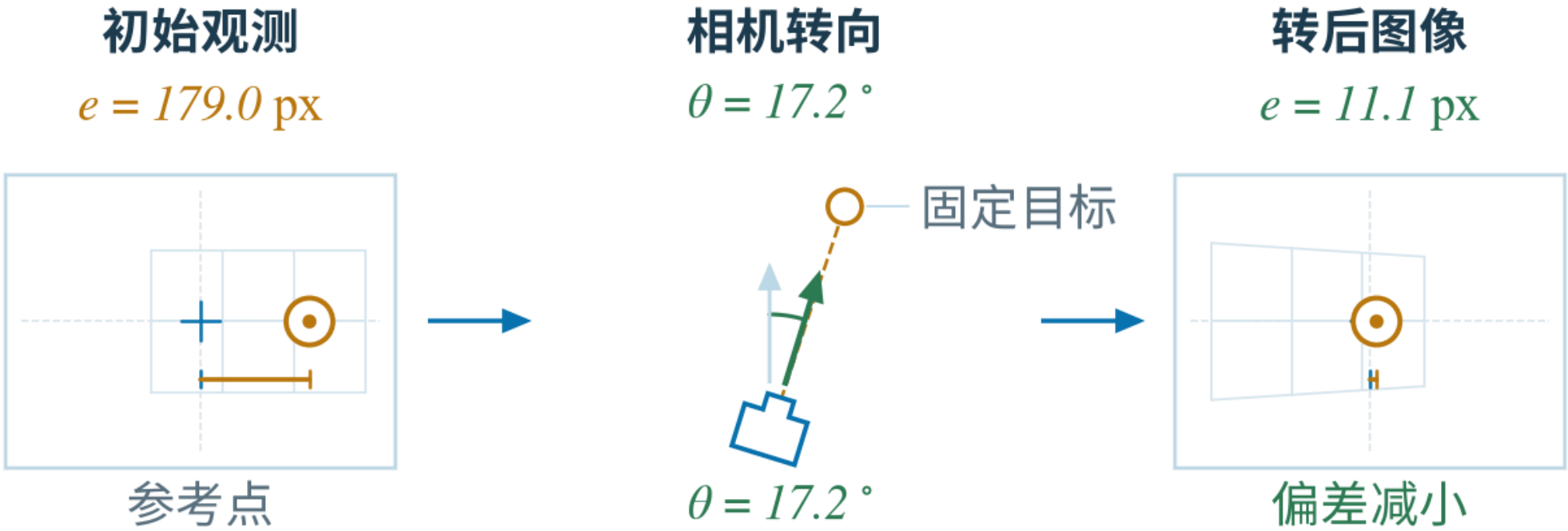


查询已建模障碍距离

还需扣除机体尺寸与余量；
未知区仍需继续观测。

8.1.1 背景意义与基本原理：视觉反馈

观察改变动作，动作也改变下一次观察



三轮修正：179.0 → 69.5 → 27.7 → 11.1 px

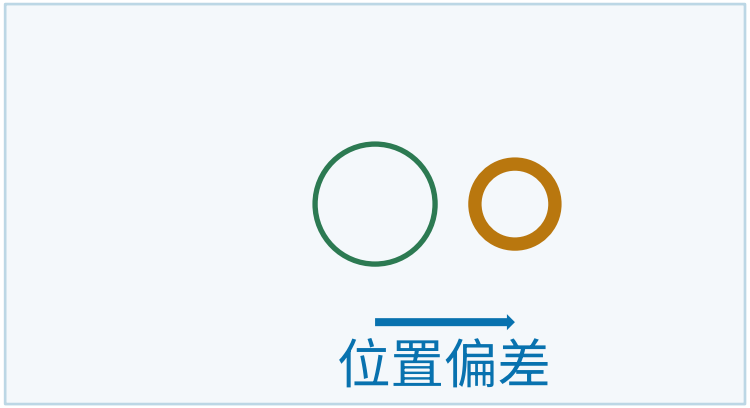
图：每次修正后重新观察，新的像素偏差决定下一次动作。

8.1.1 背景意义与基本原理：视觉伺服

同一个接近任务，可以从两种误差形成动作

目标：接近前方标记，并保持它位于视野中。

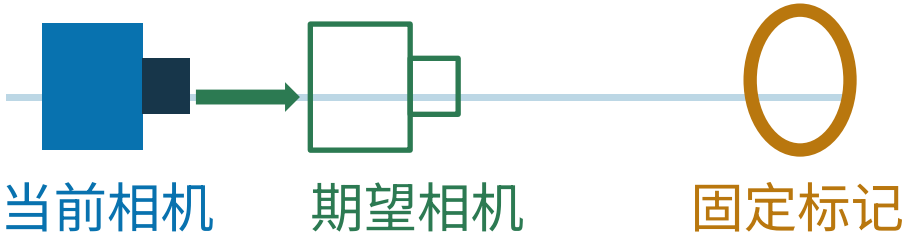
基于图像的视觉伺服（IBVS）



绿：期望图像；黄：当前图像

比较中心、面积或角点等图像特征，
经运动映射生成速度等参考。

基于位置的视觉伺服（PBVS）



先估计目标与相机的相对位姿，
再由位置、姿态误差生成参考。
基本方案不以全局建图为前提。

8.1.1 背景意义与基本原理：全章连接

单一观测有盲区，多模态让信息互相补充

观测来源	主要补充	需要注意
图像	颜色、纹理、目标与场景语义	光照、遮挡、尺度歧义
惯性器件 卫星定位	惯性器件：短时运动 卫星定位：全局位置参考	惯性积分会漂移 卫星信号易受遮挡
深度与激光	距离、表面与障碍几何	量程、无效值与外参

算法解决“怎样提取”，坐标与融合解决“怎样合在一起用”。

8.1.1 背景意义与基本原理：局部与全局

从单机看见，走向多机共享与协同行动

机器人A
发现前方障碍

机器人B
观察另一条通路

交换带坐标和
时间的环境状态

协调路径
避免相互干扰

先在仿真中检查感知、通信与避障效果，
再逐步迁移到真实传感器和载具。

本章目录

背景意义与图像处理基础

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 4 个教学单元

8.1.1 背景意义与基本原理
教材对应：第8.1.1节

8.1.2 图像预处理与边缘检测
教材对应：第8.1.2.1节（1）–（2）

8.1.3 轮廓提取与图像金字塔
教材对应：第8.1.2.1节（3）–（4）

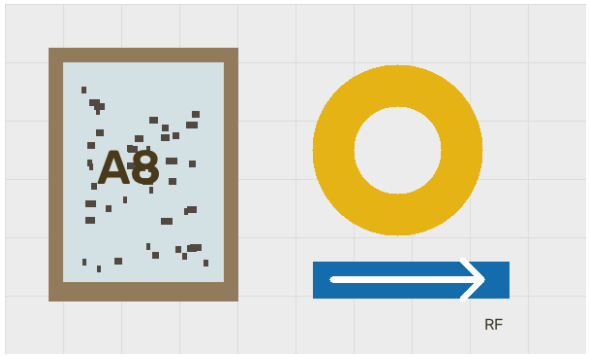
8.1.4 特征检测、描述与图像配准
教材对应：第8.1.2.1节（5）–（6）

8.1.2 图像预处理与边缘检测：处理链

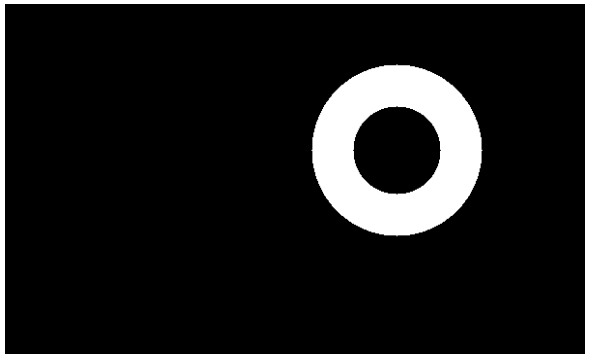
预处理不是美化图片，而是保留任务所需的信息

BGR：蓝、绿、红通道顺序。

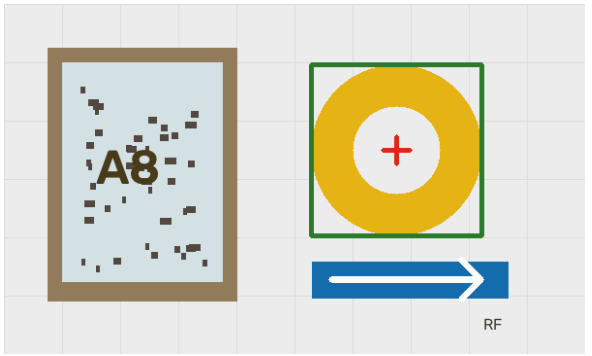
HSV：色相、饱和度、明度，便于按颜色阈值分割。



输入图像



颜色阈值掩膜



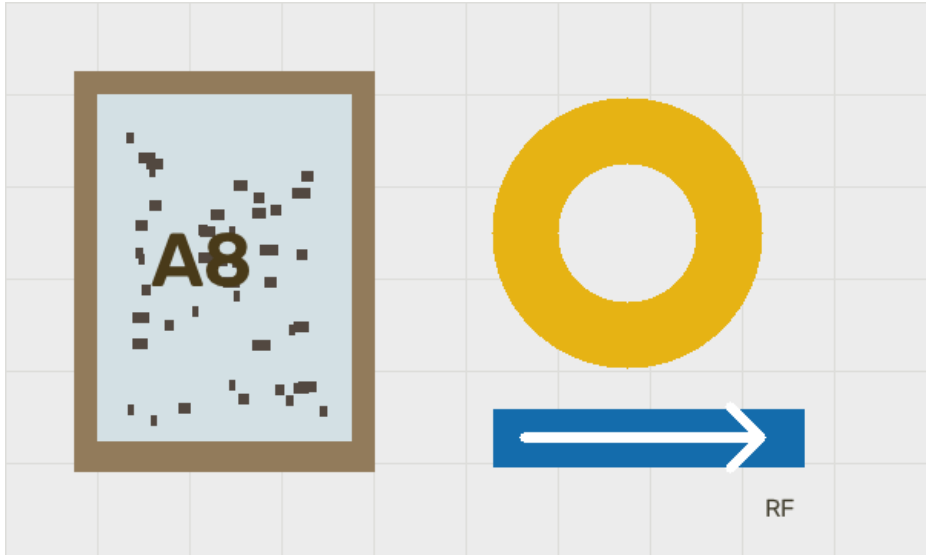
区域与中心

BGR → HSV → 阈值 → 连通区域 → 质心与面积。

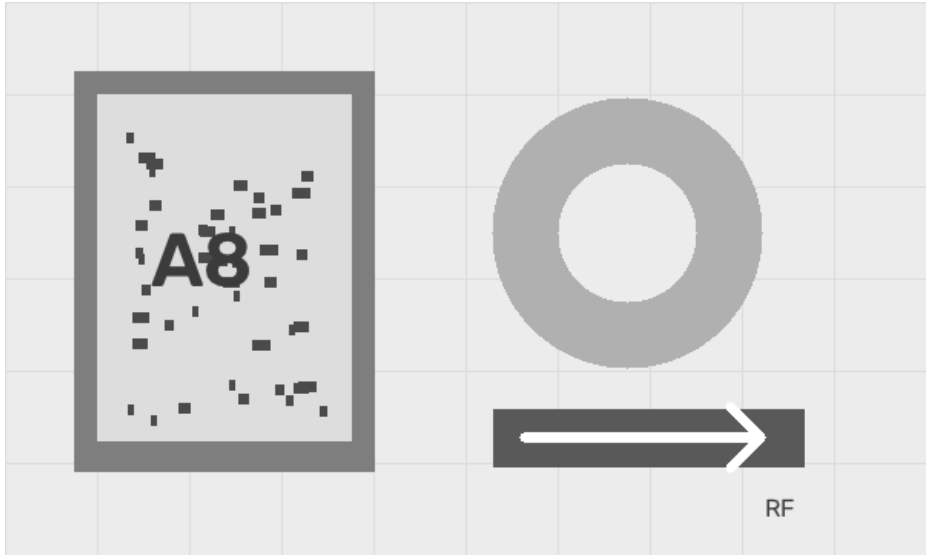
质心：区域像素坐标的均值；面积：区域前景像素数。

8.1.2 图像预处理与边缘检测：图像表达

颜色与灰度：先决定保留哪一种差异



颜色适合区分黄色标记



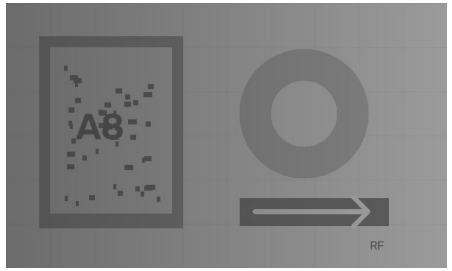
灰度适合提取亮暗结构

OpenCV按BGR读入；灰度将三个通道合成亮度近似。

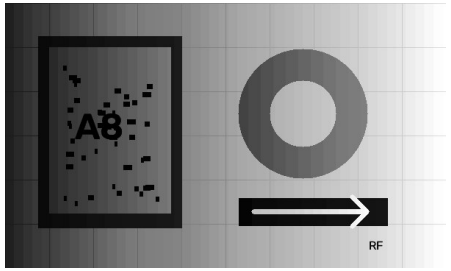
8.1.2 图像预处理与边缘检测：亮度增强

直方图增强重新分配亮度，不会凭空恢复细节

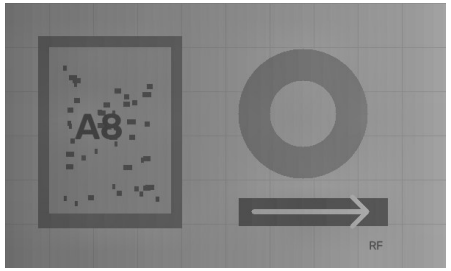
CLAHE：限制对比度的自适应直方图均衡化。



低对比度输入

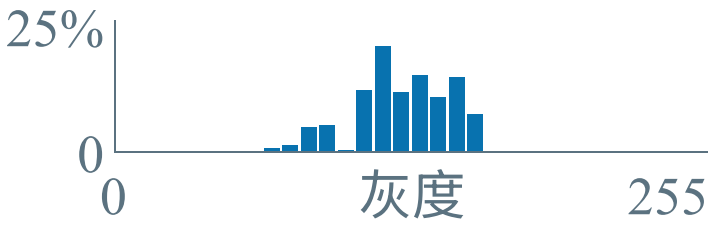


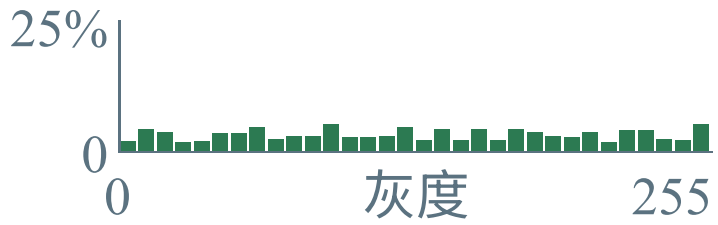
全局均衡化

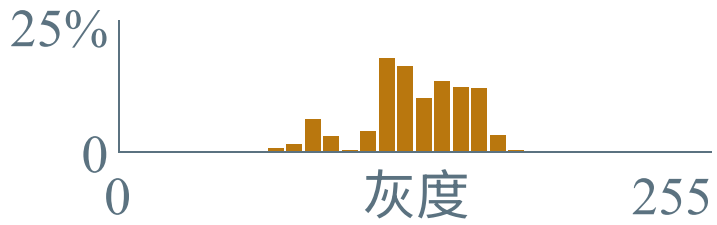


局部CLAHE

纵轴：各灰度档像素数占比（%）；三图同尺度，每档8个灰度值。





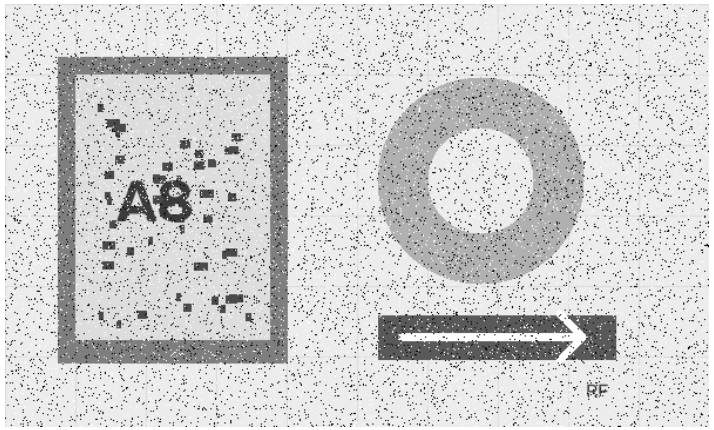


全局：按累计分布映射灰度。

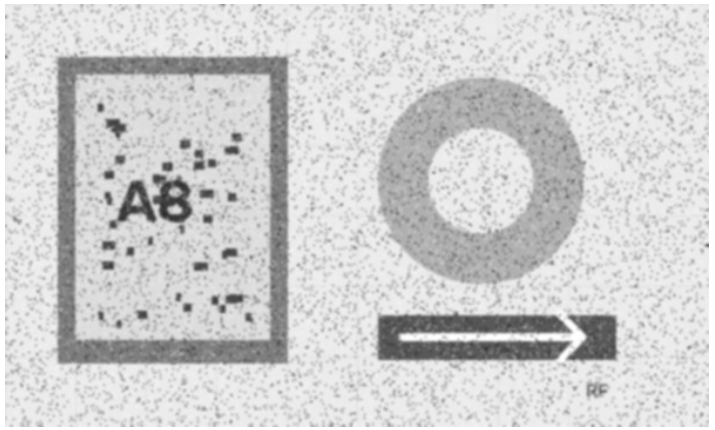
局部：分块均衡，用限幅控制增强。

8.1.2 图像预处理与边缘检测：平滑滤波

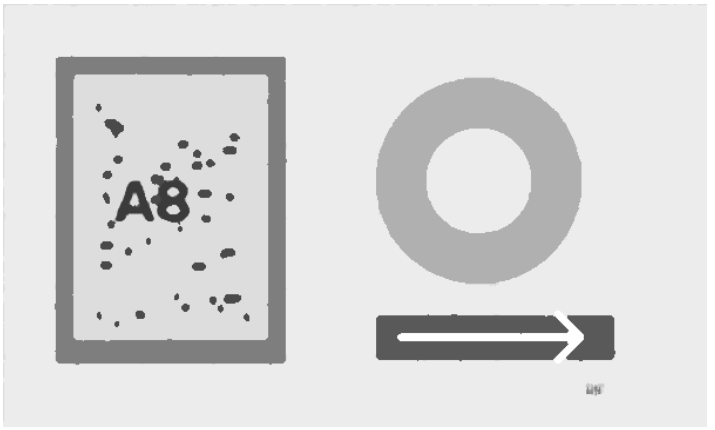
噪声类型不同，合适的滤波器也不同



椒盐噪声输入



高斯滤波 5×5



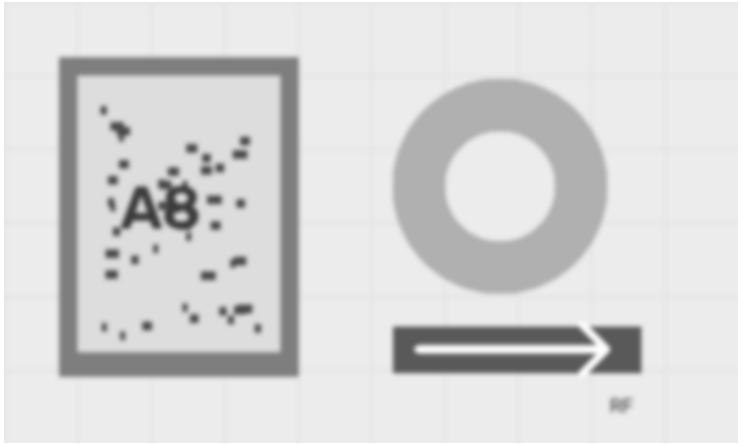
中值滤波 5×5

5×5表示邻域宽、高各5像素。高斯滤波对邻域加权平均。
中值滤波将邻域排序取中间值，适合抑制孤立噪点。

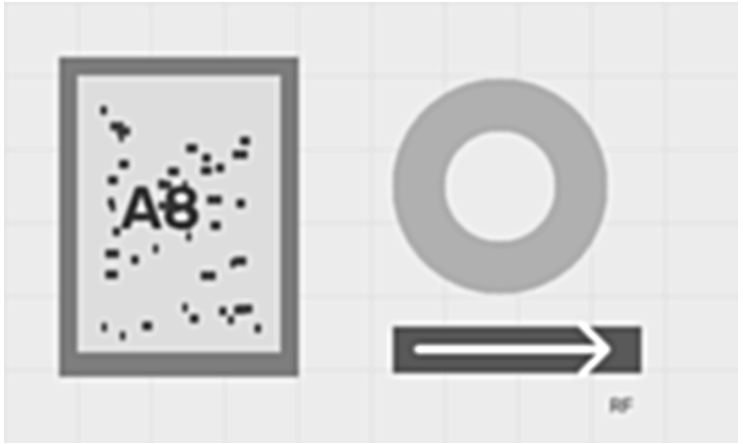
双边滤波同时参考位置和灰度差异，兼顾平滑与保边。

8.1.2 图像预处理与边缘检测：图像锐化

锐化增强已有变化，不能代替清晰成像



模糊输入



高频增强后

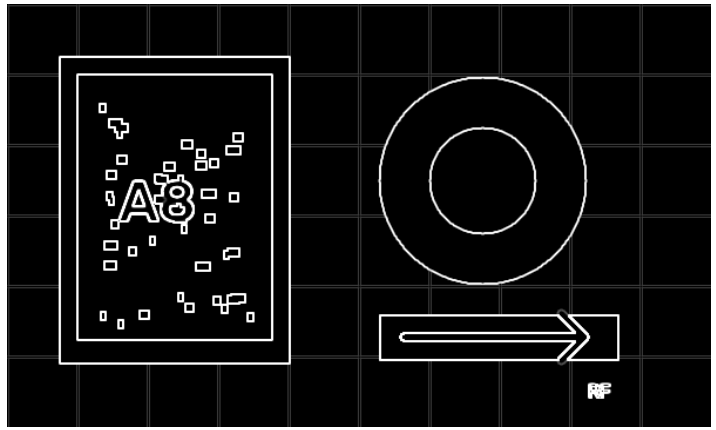
$$I_{\text{sharp}} = I + \alpha(I - G_\sigma * I)$$

I为输入；G_σ为高斯核；α为增强系数；$*$表示卷积。

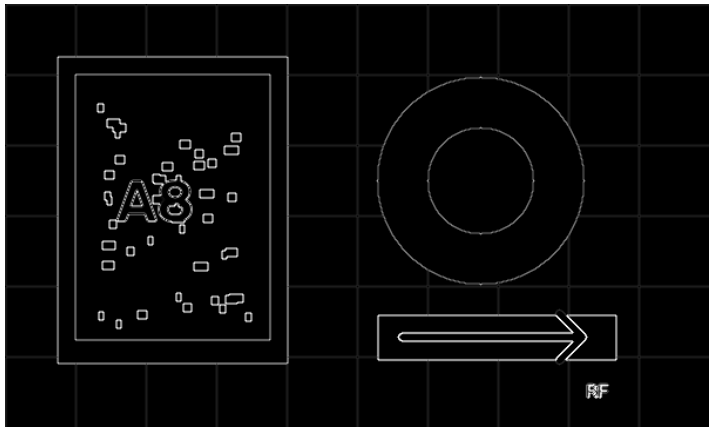
增强系数过大，会出现光晕并放大噪声；先检查输入质量。

8.1.2 图像预处理与边缘检测：梯度算子

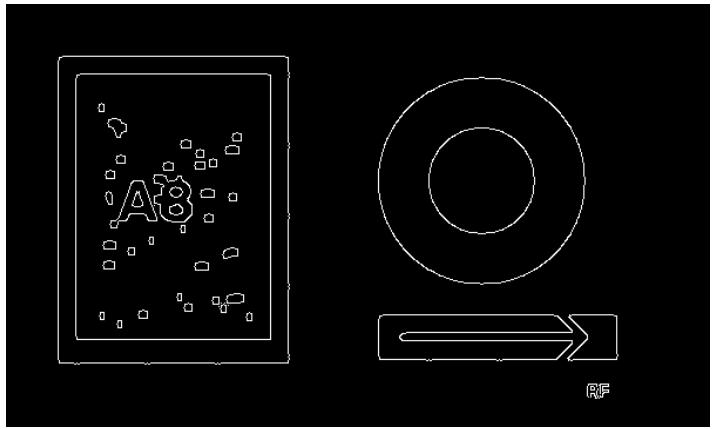
边缘是灰度变化的位置，不等于完整物体



Sobel：梯度幅值



Laplacian：绝对值显示



Canny：细化与连接

Laplacian图将二阶响应取绝对值并截断，仅显示变化强度。
纹理也会形成边缘；遮挡和弱对比度又可能使外轮廓断裂。

8.1.2 图像预处理与边缘检测：多阶段检测

Canny用强响应确立边缘，用连接保留弱响应

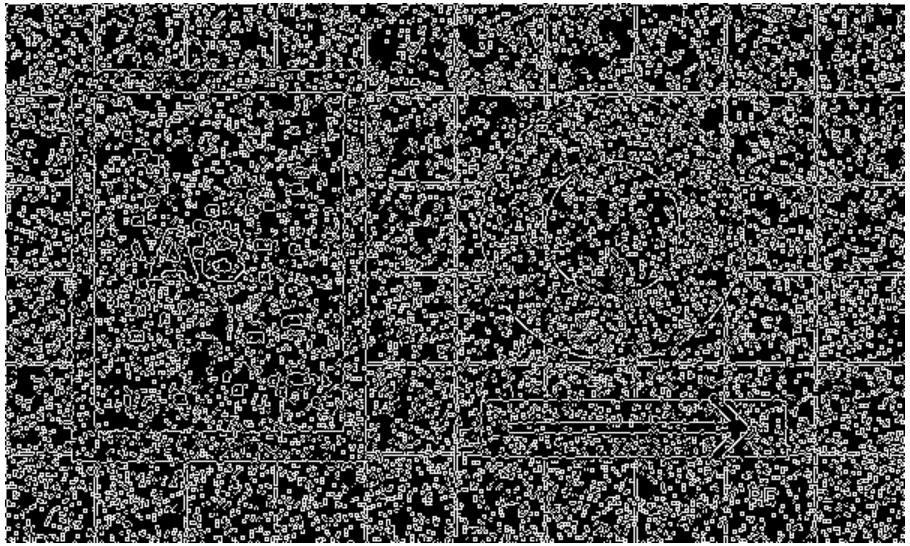
平滑
减小噪声
→
梯度
方向与强度
→
细化
沿梯度取极大
→
双阈值连接
保留连续边缘

高阈值用于确认强响应，低阈值保留与强响应相连的候选点。

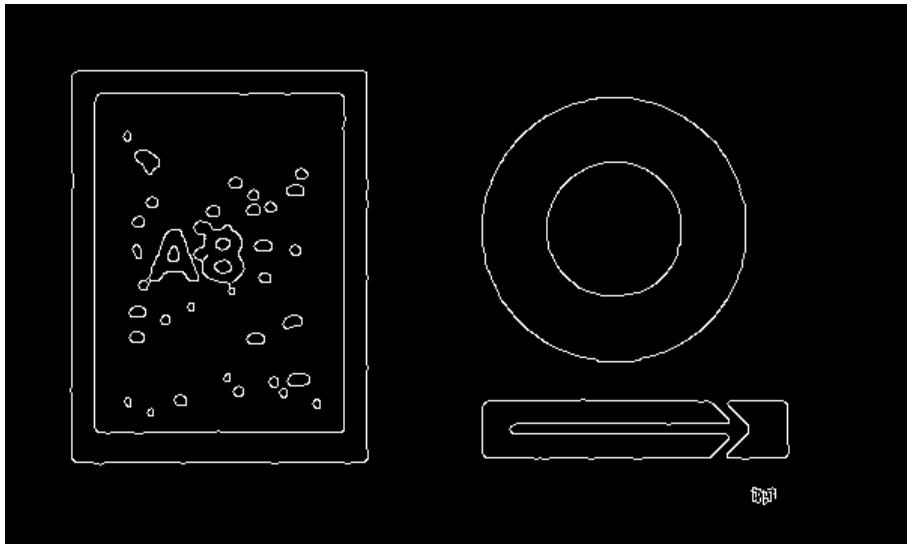
例：低/高阈值60/140；梯度强度160保留，40舍弃。
强度100：与强边缘连通才保留；这里不是原始灰度值。

8.1.2 图像预处理与边缘检测：参数对比

先看噪声，再调阈值，避免把问题交给下一步



带噪输入：阈值30 / 70



中值滤波后：阈值60 / 140

图中同时改变滤波与阈值，展示两者的组合效果。
若要区分各自贡献，应固定其中一项，只改变另一项。

本章目录

背景意义与图像处理基础

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 4 个教学单元

8.1.1 背景意义与基本原理
教材对应：第8.1.1节

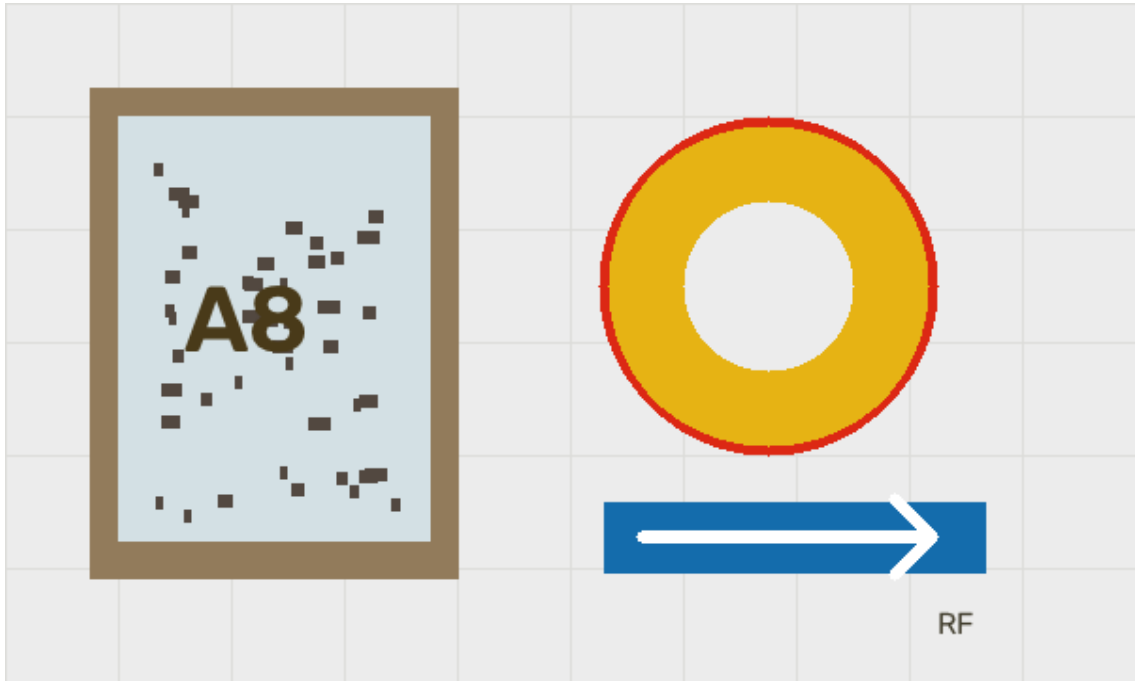
8.1.2 图像预处理与边缘检测
教材对应：第8.1.2.1节（1）–（2）

8.1.3 轮廓提取与图像金字塔
教材对应：第8.1.2.1节（3）–（4）

8.1.4 特征检测、描述与图像配准
教材对应：第8.1.2.1节（5）–（6）

8.1.3 轮廓提取与图像金字塔：轮廓提取

轮廓、区域与几何模型，是三种不同组织方式



颜色阈值分割得到的外轮廓

活动轮廓：从初始曲线出发。
最小化图像项与平滑项总能量。

$$E = E_{\text{image}} + E_{\text{smooth}}$$

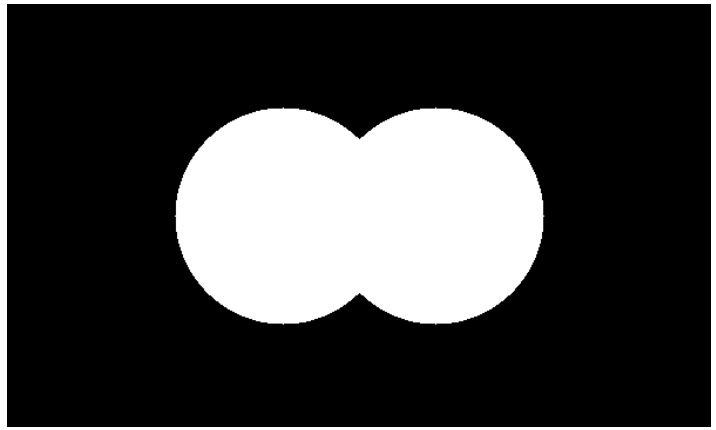
图像项：把曲线吸引向边缘。

平滑项：抑制曲线不规则弯曲。

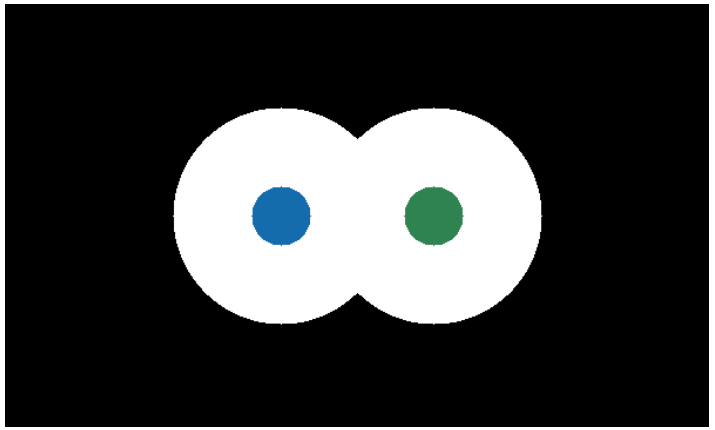
分水岭：标记引导区域分开。
霍夫：参数空间投票找规则形状。

8.1.3 轮廓提取与图像金字塔：区域分割

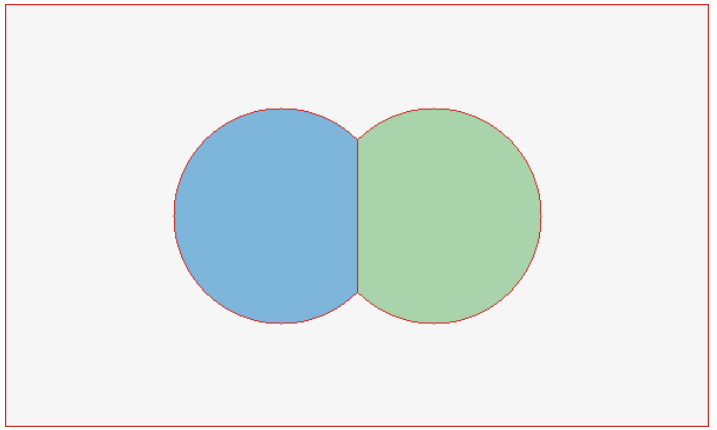
分水岭用标记告诉算法：哪些区域应当分开



相接目标



前景与背景标记



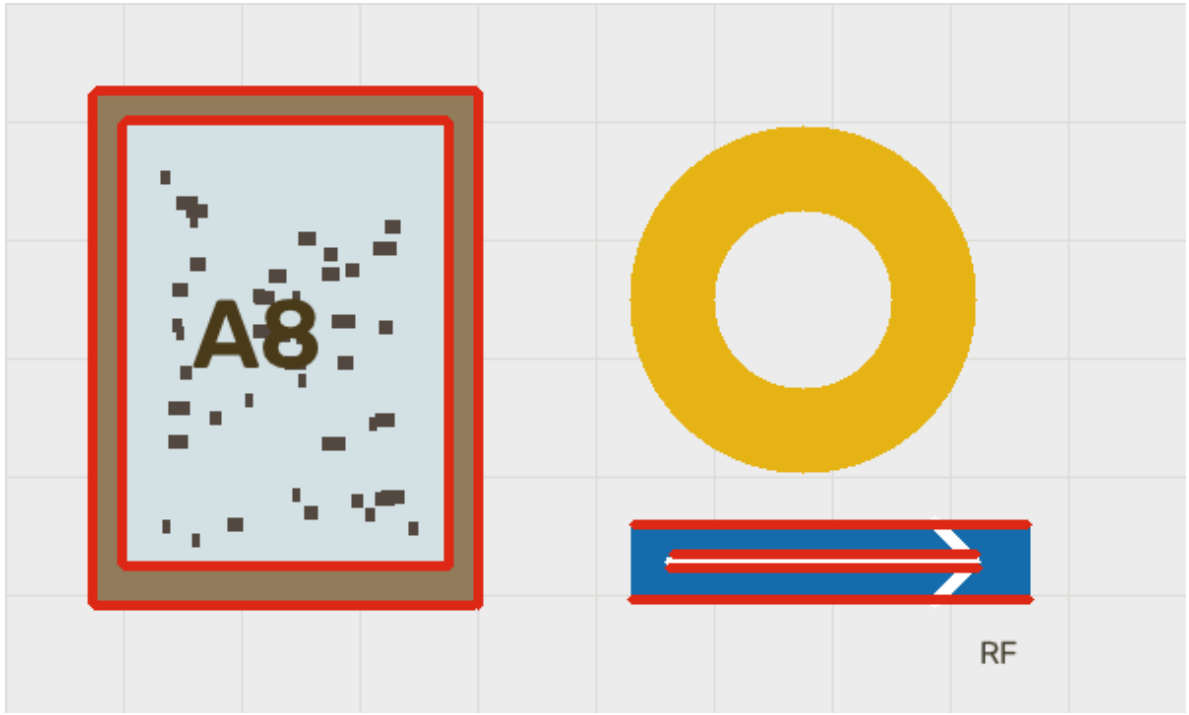
分水岭区域结果

为什么需要标记：仅做连通域分析会把相接目标合并；标记提供分离线索。

彩色区域表示不同标签，红线表示区域边界。

8.1.3 轮廓提取与图像金字塔：规则几何

霍夫投票把散点证据汇集成直线或圆



红线：被多个边缘点支持的直线

$$\rho = x\cos\theta + y\sin\theta$$

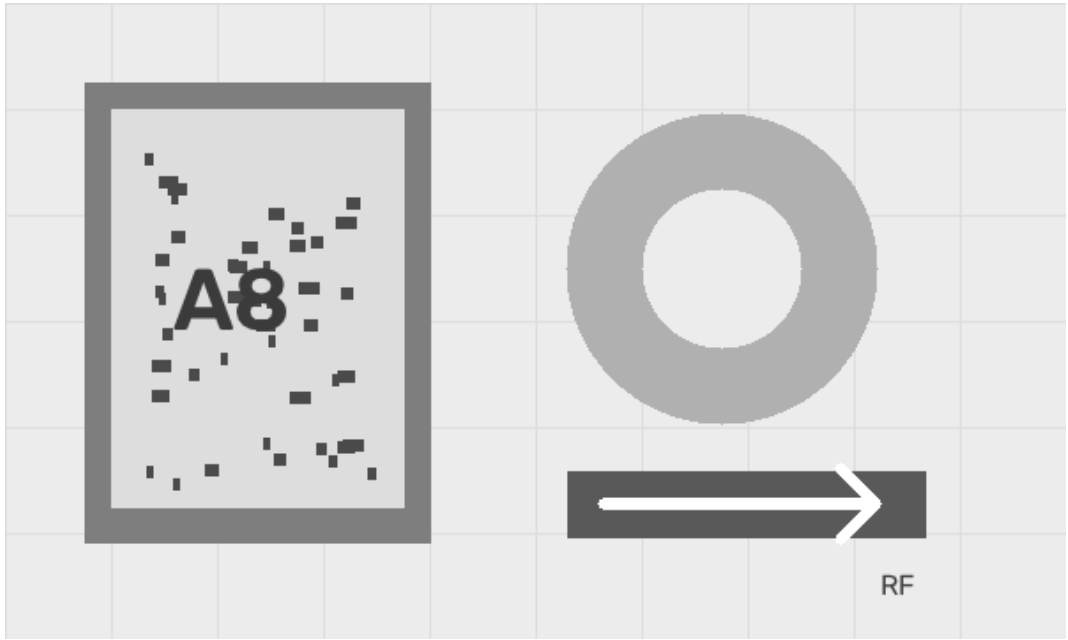
x, y：图像边缘点坐标（像素）。
原点：左上；x向右，y向下。
ρ：有符号距离（像素）。
θ：法线相对x轴的角（弧度）。

每个点支持一组候选直线。
共线点在同一参数附近累积投票。

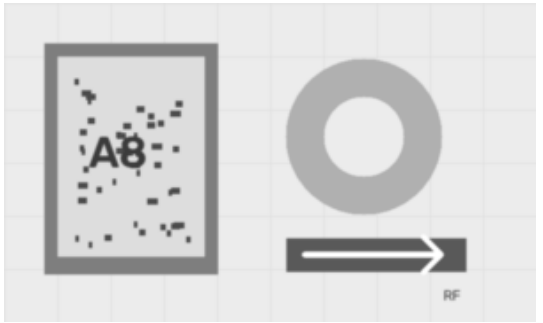
8.1.3 轮廓提取与图像金字塔：多分辨率

图像金字塔把同一场景放到不同尺度观察

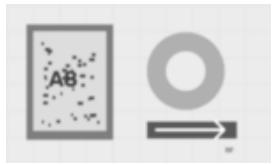
高斯金字塔：先平滑，再将宽和高各下采样一半。



600×360



300×180



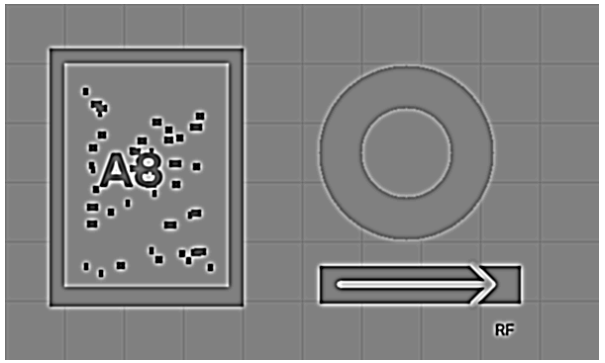
150×90

由粗到细：先粗定位，再细化。
代价：小目标可能在粗层消失。

8.1.3 轮廓提取与图像金字塔：差分与残差

高斯差分找尺度变化，拉普拉斯层保留重建残差

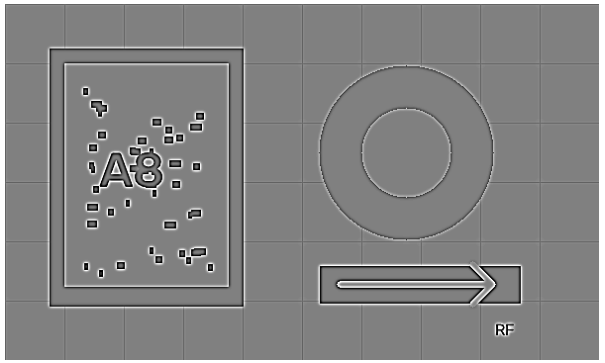
G为高斯图，σ为模糊尺度；up将粗层上采样到同尺寸。



DoG：同尺寸，不同模糊尺度

$$D = G_{\sigma_1} - G_{\sigma_2}$$

D为高斯差分；L_l为第l层残差。



Laplacian层：细层减上采样粗层

$$L_l = G_l - \mathrm{up}(G_{l+1})$$

$$G_l = L_l + \mathrm{up}(G_{l+1})$$

显示：中灰为零，明暗代表正负；仅为读图放大对比。

本章目录

背景意义与图像处理基础

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 4 个教学单元

8.1.1 背景意义与基本原理
教材对应：第8.1.1节

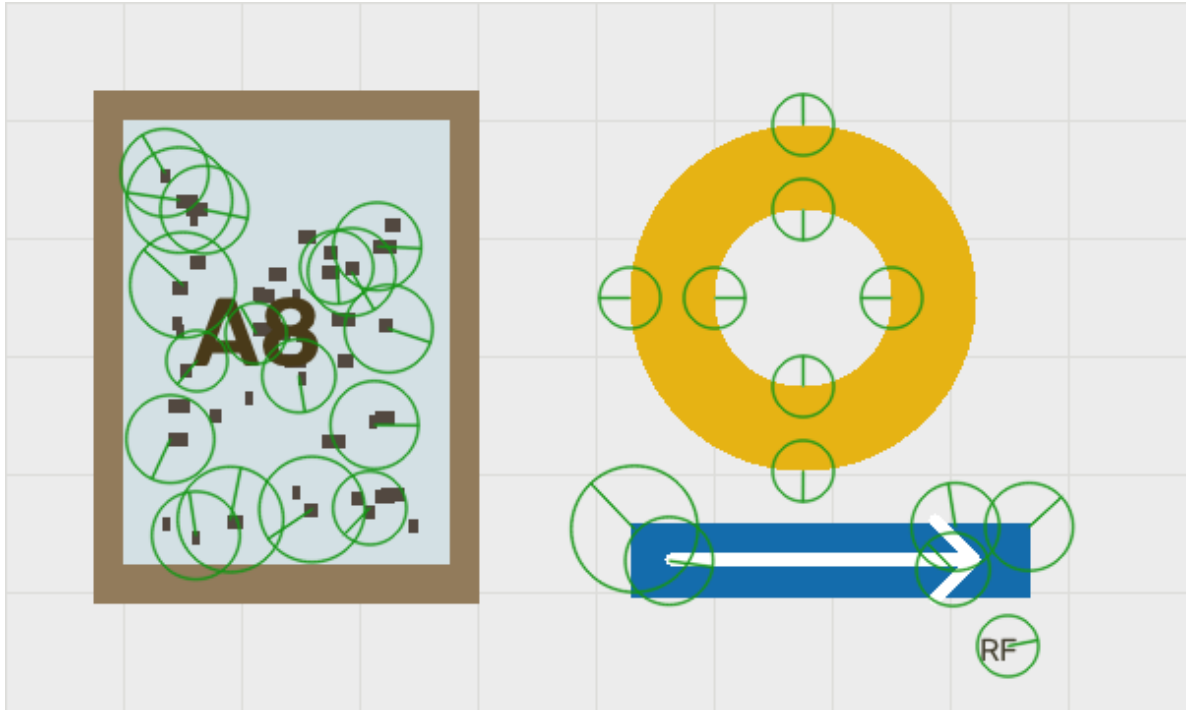
8.1.2 图像预处理与边缘检测
教材对应：第8.1.2.1节（1）–（2）

8.1.3 轮廓提取与图像金字塔
教材对应：第8.1.2.1节（3）–（4）

8.1.4 特征检测、描述与图像配准
教材对应：第8.1.2.1节（5）–（6）

8.1.4 特征检测、描述与图像配准：检测与描述

关键点回答“看哪里”，描述子回答“像不像”



圆和方向线：部分关键点及局部方向

检测器：选出角点、斑点等位置。

描述子：编码邻域外观，便于比较。

匹配与几何：找候选，再检验能否由同一变换解释。

8.1.4 特征检测、描述与图像配准：尺度与旋转

SIFT与SURF利用尺度空间寻找较稳定的特征

方法	定位机制	方向与描述
SIFT	DoG尺度空间极值 再筛除不稳定点	主梯度方向归一 128维局部梯度描述
SURF	Hessian行列式定位 积分图加速近似滤波	邻域主方向归一 累积局部方向响应

Hessian：由二阶导数组成，用行列式定位斑点。

积分图：累积像素和，使矩形区域求和更快。

8.1.4 特征检测、描述与图像配准：轻量特征

FAST与BRIEF分工合作，ORB加入方向与尺度处理

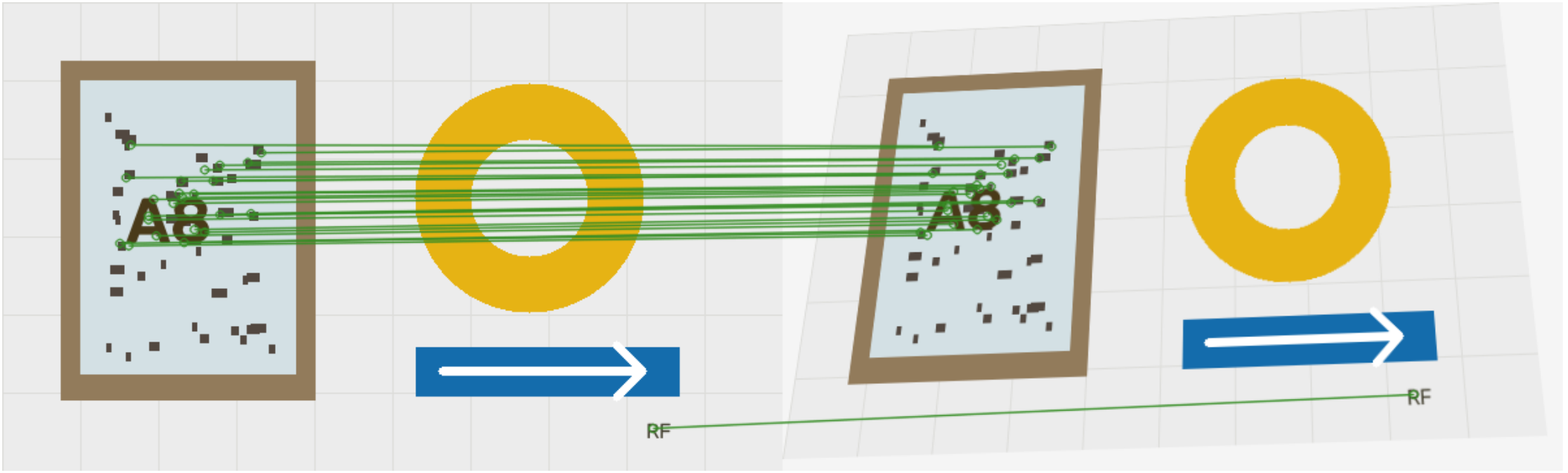
方法	判据与构造
FAST	圆环连续像素段一致亮或暗于中心，且差值超阈值。
BRIEF	比较邻域像素对的灰度，生成二进制串。
ORB	灰度质心给方向；旋转像素对；金字塔处理尺度。

$$d_H(101101,100001) = 2$$

汉明距离：不同位的数量；本例第3、4位不同，共2位。

8.1.4 特征检测、描述与图像配准：特征匹配

外观匹配先给候选，几何检查再判断可信度



连线：距离比检验后的一部分候选匹配

距离比：最近匹配距离除以次近距离，小于0.75才保留候选。

实际计算：两图关键点 443 / 500 个；候选匹配 177 组。

8.1.4 特征检测、描述与图像配准：稳健估计

RANSAC重复随机抽样，寻找较大一致集

随机抽4对拟合候选 → 计算误差区分内外点 → 反复检验选择一致集 → 利用内点重新估计

随机抽样：取4对对应拟合平面单应；每幅图不出现3点共线。

重投影误差：预测点与观测点的像素距离；误差≤阈值为内点。

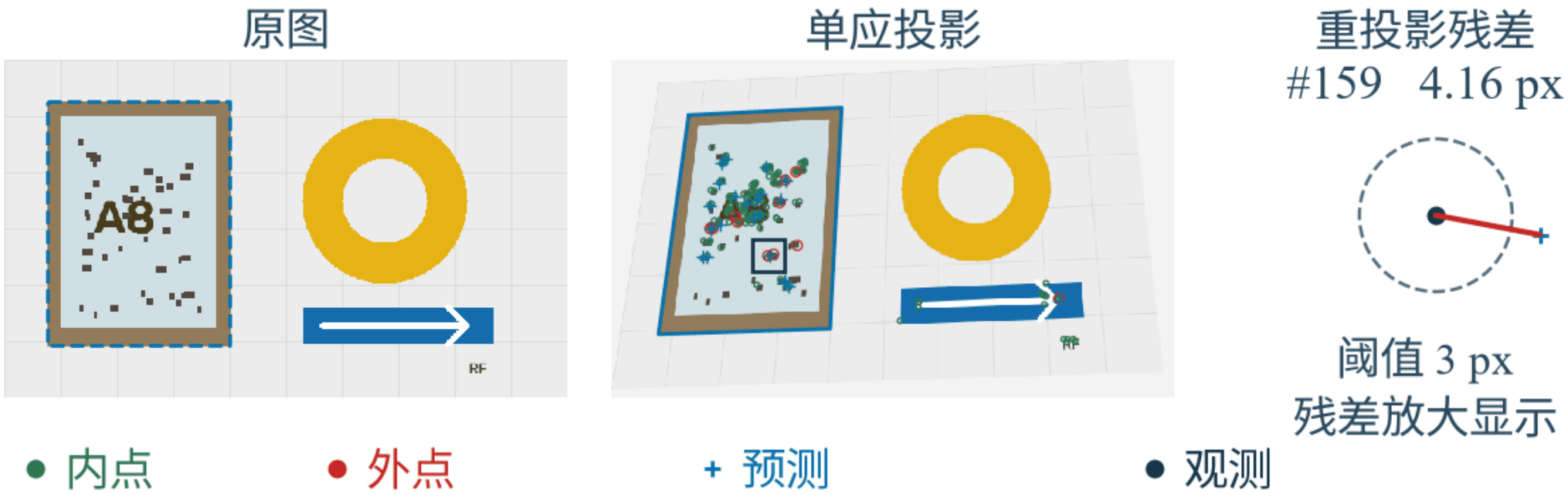
本例结果：177 组候选中 166 组满足3像素阈值；再查内点分布。

数据互相支持，比单个匹配“看起来很像”更有说服力。

8.1.4 特征检测、描述与图像配准：几何一致性

RANSAC用单应投影检查候选匹配

四对样本拟合候选；预测点与观测点相距不超过3 px时计为内点。



图：平面图像匹配的采样、投影与残差（离线计算）。

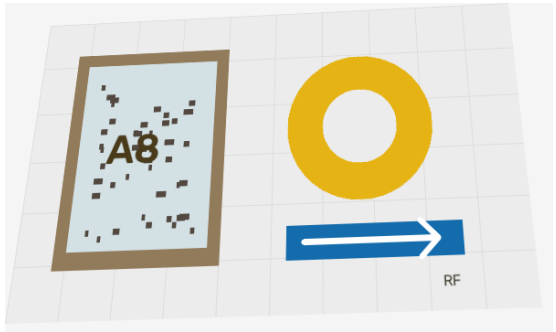
8.1.4 特征检测、描述与图像配准：图像配准

单应矩阵把平面上的对应关系写成一个变换

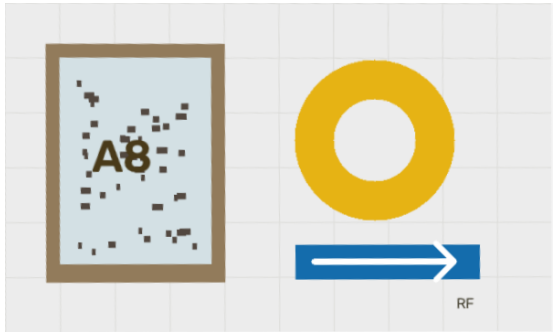
$$\lambda \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = H \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

u, v：原图坐标（像素）。
u', v'：透视图坐标（像素）。
H：3×3单应矩阵。
λ：齐次尺度，随图像点变化。

原图经H映到透视图；反向配准用H^{-1}。



透视变化后的输入



估计单应后反向配准

本例物体共面，因此可用单个单应；一般三维场景不能直接照搬。

8.1.4 特征检测、描述与图像配准：图像拼接

拼接先对齐几何，再处理亮度与接缝

模型	描述变化	适用关系
仿射	平移、旋转、缩放、错切	保持平行线关系
单应	还包含透视变化	共面场景或相机纯旋转

加权平均：重叠区按像素权重混合。

多尺度：分解拉普拉斯层，按高斯权重逐层融合。

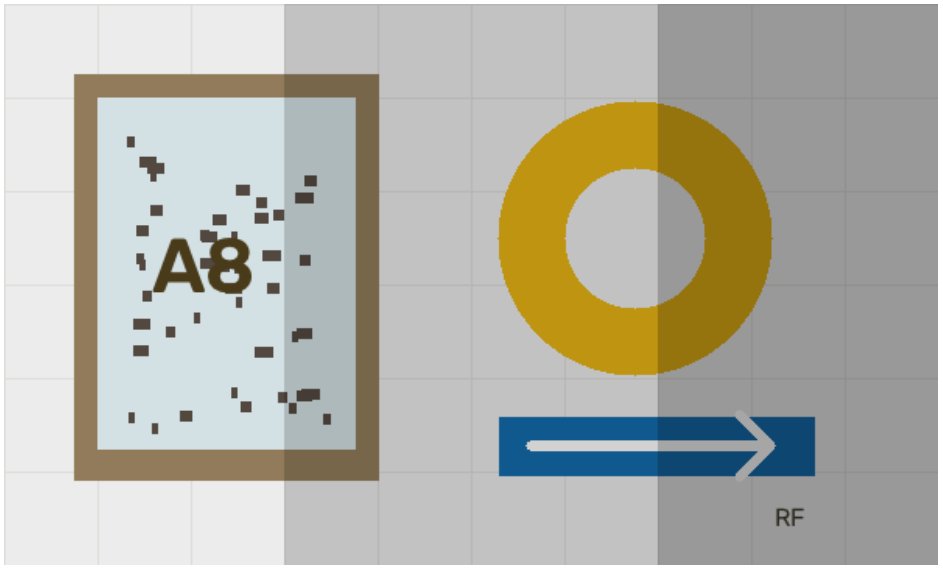
梯度域：在边界约束下重建亮度，匹配源图梯度。

8.1.4 特征检测、描述与图像配准：融合结果

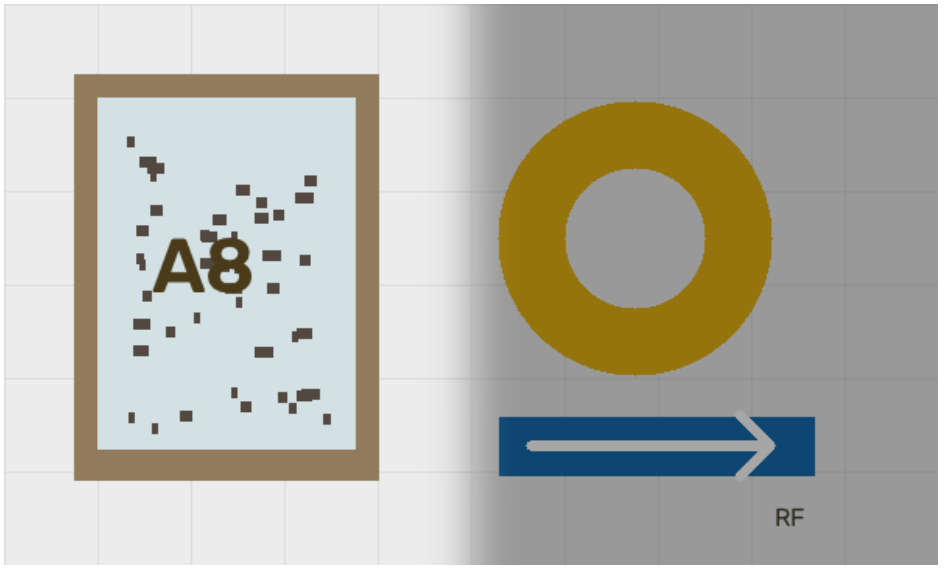
多尺度融合使亮度接缝的过渡更平缓

离线平面裁图：重叠240像素，已按已知平移对齐。

A亮度不变，B乘0.65；下图仅比较融合方式。



等权重平均：亮度阶跃仍明显



多尺度融合：接缝过渡更平缓

8.1.4 特征检测、描述与图像配准：方法选择

三条处理链，说明算法应由任务来选择

任务	合理处理链	验收对象
找到黄色标记	HSV阈值 → 轮廓 → 中心与面积	漏检、误检 中心稳定性
对齐两幅平面图	特征 → 匹配 → RANSAC → 单应	内点分布 重投影误差
寻找规则边界	去噪 → 边缘 → 霍夫或轮廓	连续性、几何偏差

下一节进一步追问：怎样把这些特征变成目标身份、连续轨迹和运动估计？

练习与讨论

练习：给三种现象选择下一步处理

问题1：黄色目标在灰度图里和背景相近，优先改什么？

问题2：图中出现孤立黑白噪点，怎样比较滤波方法？

问题3：两幅图匹配连线很多，但配准后仍严重重影，检查什么？

参考要点

参考要点：先查信息，再查参数，最后查模型

问题1：保留彩色，检查HSV阈值、白平衡与光照；灰度已丢失颜色差异。

问题2：固定带噪输入，比较5×5中值和高斯；同时观察噪点与目标边界。

问题3：先排误配，再看是否有视差或运动目标；确认所选几何模型合适。

本节小结

图像基础最终服务于稳定、可解释的观测

保留信息：颜色、亮度与尺度的选择，要服从目标任务。

组织结构：从局部边缘，到轮廓、区域和可比较的特征。

检验一致：匹配需要几何验证，图像好看不等于测量正确。

图像处理方法与函数参考（OpenCV官方教程）
https://docs.opencv.org/4.x/d2/d96/tutorial_py_table_of_contents_imgproc.html