Chapter Outline

# Multimodal Fusion and Control Decision-Making

**Chapter 8 Multimodal Perception and Intelligent Decision-Making**


8.1 Image Processing Fundamentals

8.2 Recognition, Tracking, Optical Flow

8.3 Localization, Mapping, and Planning

8.4 Common Perception Sensors

8.5 Coordinate Transformations

**8.6 Fusion and Control Decision-Making**

8.7 Tools and Framework Interfaces

8.8 Experiments and Real-World Transfer

8.9 Exercises and Classroom Discussion


## 5 Units in This Section

**8.6.1 Multimodal Perception Modeling**
Textbook: Sec. 8.1.6.1

**8.6.2 Multimodal Information Fusion**
Textbook: Sec. 8.1.6.2

**8.6.3 Trajectory Generation and Obstacle Avoidance Planning**
Textbook: Sec. 8.1.6.3

**8.6.4 Trajectory Tracking and External Control**
Textbook: Sec. 8.1.6.4

**8.6.5 Intelligent Planning and Decision-Making**
Textbook: Sec. 8.1.6.5

# 8.6 Fusion and Control Decision-Making

## 8.6.1 Multimodal Perception Modeling: System Framework

## Measurements, Estimation, Planning, and Execution Close the Loop

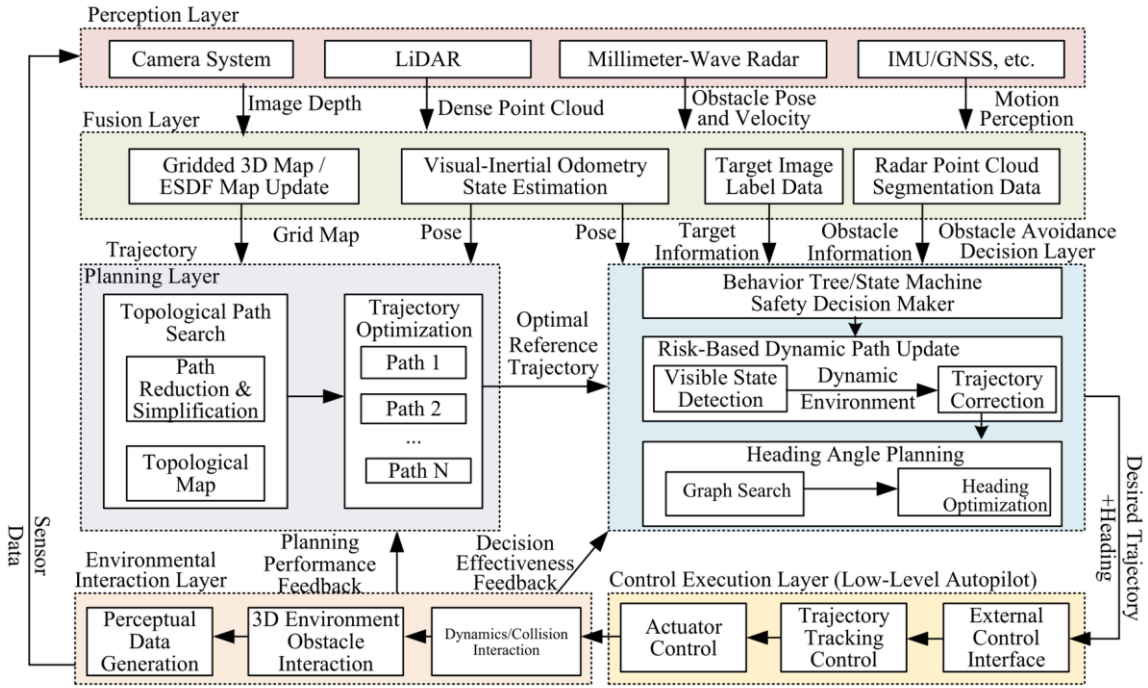



Fig. 8.7 Multimodal perception and decision framework

**Real-time feedback:** actions change state; sensors observe again; estimates and plans update.

**Parallel branches:** maps, poses, and targets feed different planning and decision modules.

**Evaluation:** recorded results support assessment and training on a different timescale from real-time control.

IMU: inertial measurement unit
GNSS: global navigation satellite system
ESDF: Euclidean signed distance field

### 8.6.1 Multimodal Perception Modeling: Framework: Perception and Fusion

# Perception and Fusion Supply Complementary Observations

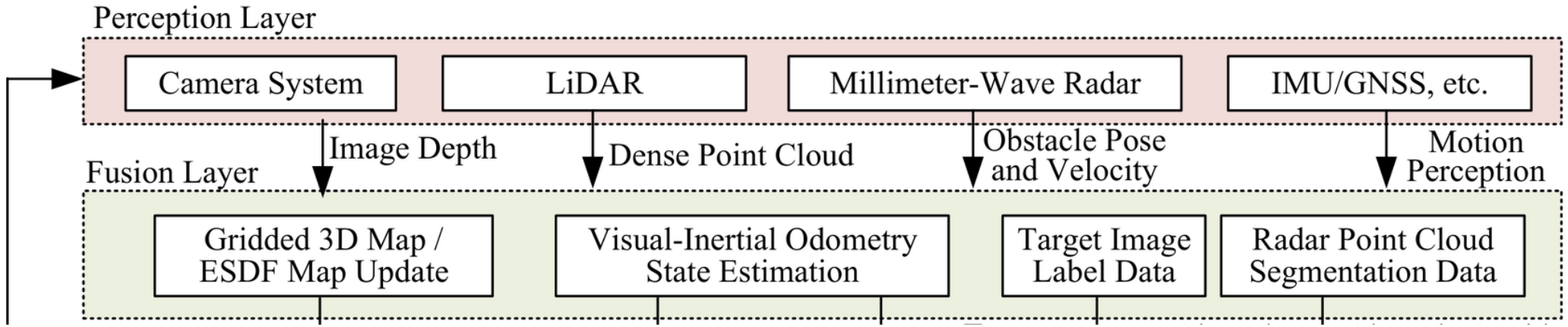



Fig. 8.7 (upper part) Perception and fusion layers

Images, depth, point clouds, radar, and inertial positioning complement one another. Maps, poses, targets, and obstacles are parallel outputs.

LiDAR: light detection and ranging.

### 8.6.1 Multimodal Perception Modeling: Framework: Planning and Decision-Making

# Candidate Trajectories and Risk Decisions Shape the Plan

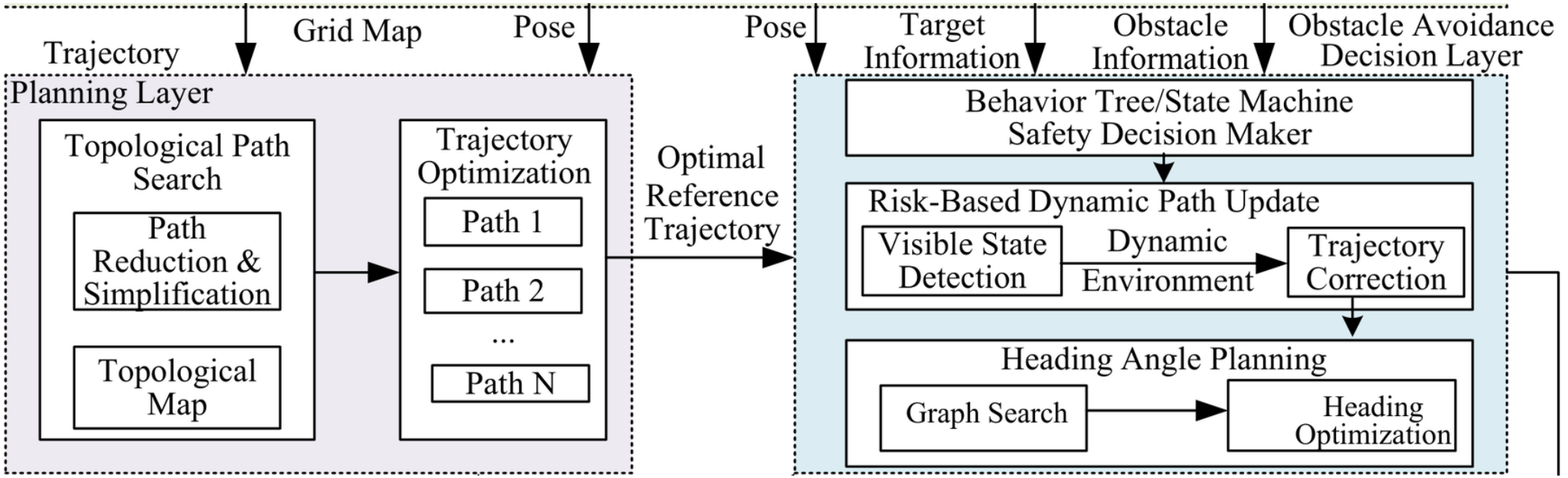



Fig. 8.7 (middle) Planning and obstacle-avoidance decisions

Search proposes paths; optimization refines positions; risk and heading constrain outputs.

### 8.6.1 Multimodal Perception Modeling: Framework: Environment and Control

## Execution Changes the Environment and the Next Observation

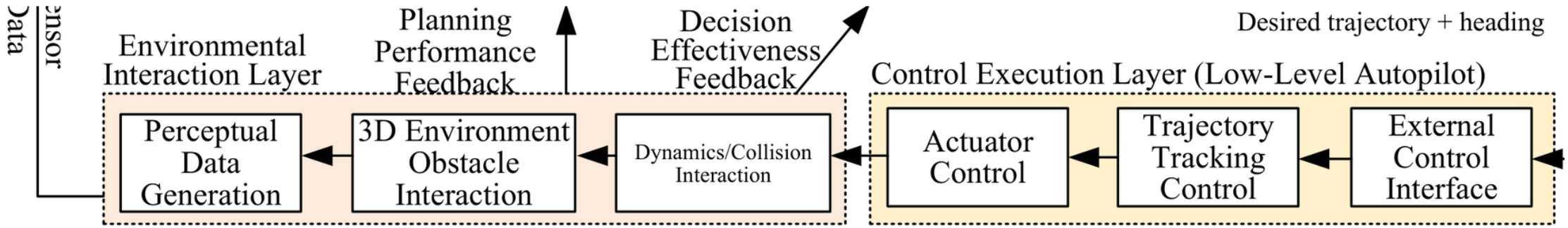



Fig. 8.7 (lower part) Environmental interaction and control execution

Desired trajectory and heading become actions through low-level control. Dynamics, collisions, and environmental interaction change the state; sensors observe again.

### 8.6.1 Multimodal Perception Modeling: Data Inputs

# Multimodal Observations Retain Their Own Data Structures

$$\mathcal{Z}_t = \{\mathcal{I}_t^{\text{cam}}, \mathcal{P}_t^{\text{lidar}}, \mathcal{R}_t^{\text{radar}}, \boldsymbol{\zeta}_t^{\text{imu}}, \mathbf{p}_t^{\text{gnss}}, \ \ldots\} \tag{8.13}$$

| Observation | Mathematical Structure | Processed Uses |
|---|---|---|
| **RGB camera (red, green, blue)** | $\mathcal{I}_t^{\text{cam}} \in \mathbb{R}^{H\times W\times 3}$ | Features, objects, depth |
| **LiDAR point cloud** | $\{\ {}^{c}\mathbf{p}_i\} \subset \mathbb{R}^3$; point matrix $\mathbf{P} \in \mathbb{R}^{3\times N}$ | Maps and object geometry |
| **Millimeter-wave radar** | $\mathcal{R}_t^{\text{radar}} \in \mathbb{R}^k$ | Range, angle, radial velocity |

Model each image, point cloud, and radar measurement first; then map to a common state.

### 8.6.1 Multimodal Perception Modeling: Inertial and Satellite Inputs

# Fused Outputs Differ from Raw Sensor Readings

FRD: forward-right-down; NED: north-east-down; GNSS: global navigation satellite system.

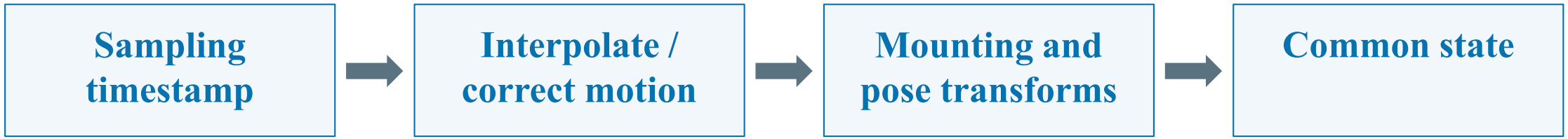
$$\zeta_t^{\text{imu}} = [({}^{b}\mathbf{a}_t)^{\mathsf{T}}, ({}^{b}\boldsymbol{\omega}_t)^{\mathsf{T}}, \mathbf{q}_t^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{R}^{10}$$

IMU (inertial measurement unit): raw nine-axis 3+3+3; fused attitude record 3+3+4.

| Quantity | Dimensions and Units | Physical Meaning |
|---|---|---|
| **Motion acceleration** | 3; m/s² | Body components after bias removal and gravity compensation |
| **Angular rate / quaternion** | 3; rad/s<br>4; dimensionless | Bias-corrected angular rate; **q** is FRD→NED attitude |
| **GNSS latitude, longitude, height** | 3; deg / deg / m | Geographic position converted to a common local NED frame |

### 8.6.1 Multimodal Perception Modeling: Time and Spatial Alignment

# Fusion Needs a Common Time and Coordinate Reference



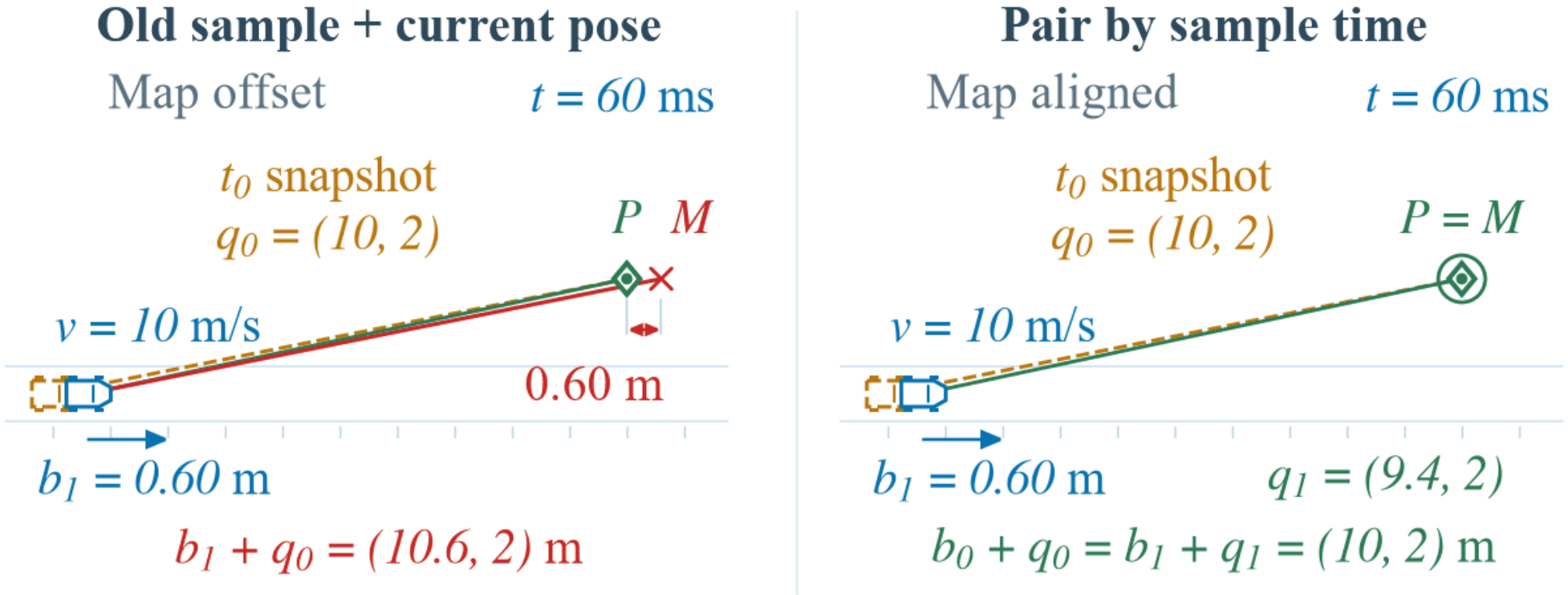

$$\Delta p \approx v\Delta t = 10\ \mathrm{m/s} \times 0.060\ \mathrm{s} = 0.60\ \mathrm{m}$$

**Time:** pair a common time base; interpolate or predict to compensate when needed.
**Space:** sensor extrinsics and vehicle pose at sampling time give common coordinates.

One loop does not synchronize sampling; a map label does not transform coordinates.

## 8.6.1 Multimodal Perception Modeling: Registering Observations and Poses

# The Wrong Pose Makes a Stationary Landmark Appear to Move




One stationary landmark: pairing an old observation with the current pose causes a 0.60 m error.

### 8.6.1 Multimodal Perception Modeling: Camera Systems

# Target Pixels and Valid Depth Give Target Position

| Scale Source | Measurement / Estimation | Key Conditions |
|---|---|---|
| **Stereo / structured light / ranging** | Disparity, coded light, or line-of-sight range | Calibration and valid depth regions |
| **Monocular known size** | Similar triangles using size and pixel extent | Known true size and usable pose |
| **Monocular learned depth** | Estimate depth from image features | Scale / error and training distribution |

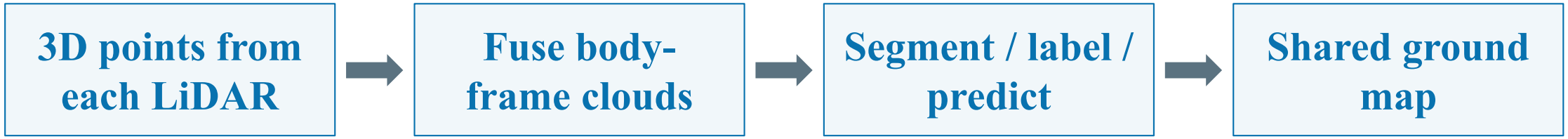
Here the focal lengths are $f_x = f_y = 320$ px; the principal point is $(c_x, c_y) = (320,240)$ px.

$$(u, v) = (480,320), \quad z_c = 4\ \text{m} \implies {}^{c}\mathbf{p} = [2,1,4]^{\mathsf{T}}\ \text{m}$$

Pixel center + depth gives position; orientation needs more geometry or pose data.

### 8.6.1 Multimodal Perception Modeling: LiDAR

# Extrinsics Combine Sensor Point Clouds into One Map




$$^{b}\mathbf{p} = [2,1,0]^{\mathrm{T}} + [0.1,0,0]^{\mathrm{T}} = [2.1,1,0]^{\mathrm{T}}\ \mathrm{m}$$

$$^{e}\mathbf{p} = [-1,2.1,0]^{\mathrm{T}} + [10,20,-30]^{\mathrm{T}} = [9,22.1,-30]^{\mathrm{T}}\ \mathrm{m}$$

Here LiDAR and body axes are parallel; the vehicle is level with yaw 90°. Interpret actual cloud units and axes using the current SDK.

### 8.6.1 Multimodal Perception Modeling: Millimeter-Wave Radar

## Doppler Velocity Gives the Relative Line-of-Sight Component

$$ {}^{c}\mathbf{p}^{\text{radar}} = [r\cos\theta\,, r\sin\theta]^{\mathsf{T}} \tag{8.14}$$

$$ {}^{c}\mathbf{v}^{\text{radar}} = v_r \frac{{}^{c}\mathbf{p}^{\text{radar}}}{\| \quad {}^{c}\mathbf{p}^{\text{radar}} \|} $$

$$ r = 10\ \text{m},\ \theta = 30^\circ,\ v_r = 2\ \text{m/s} \implies \mathbf{p} = [8.6603,5]^{\mathsf{T}}\ \text{m} $$

The radial component is $[1.7321,1]^{\mathsf{T}}$ m/s; positive velocity here is away from the sensor. Tangential velocity is unobserved. This is a 2D model; 3D position also needs elevation or other constraints.

### 8.6.1 Multimodal Perception Modeling: Inertial Propagation

## State Equations Update Position, Velocity, and Attitude

$$ {}^{e}\dot{\mathbf{p}}_t = {}^{e}\mathbf{v}_t \tag{8.15} $$

$$ {}^{e}\dot{\mathbf{v}}_t = \mathbf{R}_b^e(\mathbf{q}_t) \; {}^{b}\mathbf{a}_t $$

$$ \dot{\mathbf{q}}_t = \frac{1}{2}\mathbf{q}_t \otimes [0, ( \; {}^{b}\boldsymbol{\omega}_t)^{\mathsf{T}}]^{\mathsf{T}} $$

${}^{b}\mathbf{a}_t$ is bias-corrected, gravity-compensated motion acceleration (m/s²).
$\mathbf{q} = [q_w, q_x, q_y, q_z]^{\mathsf{T}}$ is a unit quaternion; $\otimes$ is the Hamilton product; $\boldsymbol{\omega}$ is bias-corrected (rad/s).

### 8.6.1 Multimodal Perception Modeling: Specific Force and Gravity Compensation

# A Stationary IMU Clarifies What Acceleration Means

$$ {}^{e}\dot{\mathbf{v}} = \mathbf{R}_b^e(\mathbf{q})( \quad {}^{b}\mathbf{f} - \mathbf{b}_a) + \quad {}^{e}\mathbf{g} $$

${}^{b}\mathbf{f}$: accelerometer specific force; $\mathbf{b}_a$: accelerometer bias; ${}^{e}\mathbf{g}$: gravity acceleration.

$$ {}^{e}\mathbf{g} = [0,0,9.81]^{\mathrm{T}}\ \mathrm{m/s^2}, \qquad {}^{b}\mathbf{f} = [0,0,-9.81]^{\mathrm{T}}\ \mathrm{m/s^2} $$

$$ \mathbf{R}_b^e = \mathbf{I},\ \mathbf{b}_a = \mathbf{0} \implies \quad {}^{e}\dot{\mathbf{v}} = \mathbf{0} $$

Wrongly integrating raw specific force gives −9.81 m/s vertical velocity and −4.905 m displacement after 1 s at rest.

**Rotate specific force and compensate gravity before integration.**

### 8.6.1 Multimodal Perception Modeling: An Inertial Integration Example

# State Results for One 0.1 s Step and Three Steps

$$p_{k+1} = p_k + v_k \Delta t + \frac{1}{2} a \Delta t^2, \quad v_{k+1} = v_k + a \Delta t$$

Northward initial values: $p_0 = 0$ m, $v_0 = 2$ m/s, $a = 1$ m/s$^2$, $\Delta t = 0.1$ s.

| Time / s | Position / m | Velocity / (m/s) |
|---|---|---|
| **0.0** | 0.000 | 2.0 |
| **0.1** | 0.205 | 2.1 |
| **0.2** | 0.420 | 2.2 |
| **0.3** | 0.645 | 2.3 |

Bias 0.1 m/s² over 10 s can add about 5 m of position error; fusion helps limit drift.

# 8.6 Fusion and Control Decision-Making

## 8.6.1 Multimodal Perception Modeling: Satellite Positioning

## Low-Rate Absolute Position Corrects High-Rate Inertial Propagation

**GNSS geographic position** → **NED with a common origin** → **Correct at sample time** → **Propagate to now**

Example: IMU 100 Hz, GNSS 10 Hz; ten propagation steps per 0.1 s.

| Data Availability | Estimator Action | Meaning of the Result |
|---|---|---|
| **Valid GNSS** | Propagate; correct by position | Continuous output; constrain drift |
| **Short GNSS outage** | Keep motion-model prediction | Uncertainty grows over time |
| **GNSS occlusion / multipath** | Assess observations using quality and residuals | Downweight or reject low-confidence observations |

### 8.6.1 Multimodal Perception Modeling: Outputs

## Maps, Vehicle State, and Dynamic Obstacles Serve Different Roles

| Output | Representation and Key Fields | Consumer |
|---|---|---|
| **3D scene map** | OctoMap occupancy; ESDF distances; TSDF surface-distance fusion | Mapping, queries, collision checks |
| **Vehicle state** | Position, attitude, velocity, covariance; VIO / LIO or GNSS–IMU | Navigation and trajectory tracking |
| **Dynamic obstacle set** | ID, position, velocity, semantics, uncertainty, future trajectory | Avoidance and multi-vehicle coordination |

ESDF: Euclidean signed distance field; TSDF: truncated signed distance field.

VIO: visual-inertial odometry; LIO: LiDAR-inertial odometry.

Maps, poses, and obstacles are the parallel fusion outputs in Fig. 8.7.

Chapter Outline

# Multimodal Fusion and Control Decision-Making

**Chapter 8 Multimodal Perception and Intelligent Decision-Making**


8.1 Image Processing Fundamentals

8.2 Recognition, Tracking, Optical Flow

8.3 Localization, Mapping, and Planning

8.4 Common Perception Sensors

8.5 Coordinate Transformations

**8.6 Fusion and Control Decision-Making**

8.7 Tools and Framework Interfaces

8.8 Experiments and Real-World Transfer

8.9 Exercises and Classroom Discussion


## 5 Units in This Section

**8.6.1 Multimodal Perception Modeling**
Textbook: Sec. 8.1.6.1

**8.6.2 Multimodal Information Fusion**
Textbook: Sec. 8.1.6.2

**8.6.3 Trajectory Generation and Obstacle Avoidance Planning**
Textbook: Sec. 8.1.6.3

**8.6.4 Trajectory Tracking and External Control**
Textbook: Sec. 8.1.6.4

**8.6.5 Intelligent Planning and Decision-Making**
Textbook: Sec. 8.1.6.5

### 8.6.2 Multimodal Information Fusion: Fusion Tasks

# Fusion Must Estimate the Same State

| Compatible Quantity | Shared Convention | Confidence Evidence |
|---|---|---|
| **Target position / velocity** | Same object, time, frame, and units | Variance and prediction residuals |
| **Class-probability vector** | Same class order and normalization | Entropy and probability calibration |
| **Scene feature encoding** | Same dimension or a learned common representation | Feature quality and training conditions |

Geometry determines comparability; the confidence model determines the weighting.

### 8.6.2 Multimodal Information Fusion: Confidence-Based Source Selection

# Confidence and Hysteresis Govern Source Switching

| Time | Camera / Confidence | LiDAR / Confidence | Selected Output |
|---|---|---|---|
| **t0** | 10.4 m / 0.90 | 9.8 m / 0.70 | Camera: 10.4 m |
| **t1: occluded** | 10.8 m / 0.20 | 10.1 m / 0.85 | LiDAR: 10.1 m |
| **t2: both invalid** | Invalid | Invalid | Missing; use state prediction |

Evaluation ranks reliability. Hysteresis reduces switching chatter; retain prediction and uncertainty during missing data.

### 8.6.2 Multimodal Information Fusion: Weighted Averaging

# Normalized Weights Produce a Common Position Estimate

$$\sum_s \lambda^{(s)} = 1, \quad 0 \le \lambda^{(s)} \le 1 \tag{8.16}$$

$$\hat{\mathbf{x}}_t = \sum_s \lambda^{(s)} \mathbf{z}_t^{(s)}, \quad \mathbf{z}_t^{(s)} \in \mathbb{R}^n \tag{8.17}$$

Northward positions at the same time: camera $z_{\text{cam}} = 10.4$ m, $\sigma^2_{\text{cam}} = 0.16\ \text{m}^2$;
LiDAR $z_{\text{lidar}} = 9.8$ m, $\sigma^2_{\text{lidar}} = 0.04\ \text{m}^2$.

$$\lambda_s = \frac{1/\sigma_s^2}{\sum_j 1/\sigma_j^2} \implies \lambda_{\text{cam}} = 0.2,\ \lambda_{\text{lidar}} = 0.8$$

$$\hat{x} = 0.2(10.4) + 0.8(9.8) = 9.92\ \text{m}, \quad \sigma^2 = 0.032\ \text{m}^2$$

Weights are dimensionless. Same-dimension position observations must be independent and unbiased; shared errors or correlation change the variance result.

## 8.6.2 Multimodal Information Fusion: Fusion Filtering

# A Dynamic Model Combines History with Current Observations

| Method | Nonlinearity / Distribution | Cost / Conditions |
|---|---|---|
| **Extended Kalman filter (EKF)** | First-order linearization; approximately Gaussian | Low cost; model and linearization matter |
| **Unscented Kalman filter (UKF)** | Deterministic samples propagate mean and covariance | No explicit Jacobians; Gaussian approximation |
| **Particle filter (PF)** | Weighted samples approximate a distribution | Multiple modes; enough particles, sound sampling |

Predict, then correct; missing data allow prediction while uncertainty usually grows.

### 8.6.2 Multimodal Information Fusion: One Filter Update

## Residual and Gain Correct State and Covariance

Initial values: $\hat{x}_0 = 10\ \text{m},\ P_0 = 0.25\ \text{m}^2$; move 1 m every 1 s; $Q = 0.04,\ R = 0.09\ \text{m}^2$.

$P$: variance; $-/+$: prior / posterior; $K$: gain; $\nu$: observation residual.

$$\hat{x}_1^- = 10 + 1 = 11\ \text{m}, \quad P_1^- = 0.25 + 0.04 = 0.29\ \text{m}^2$$

$$K_1 = \frac{P_1^-}{P_1^- + R} = \frac{0.29}{0.38} = 0.763158, \quad \nu_1 = 10.8 - 11 = -0.2\ \text{m}$$

$$\hat{x}_1^+ = \hat{x}_1^- + K_1\nu_1 = 10.847368\ \text{m}$$

$$P_1^+ = (1 - K_1)P_1^- = 0.068684\ \text{m}^2$$

### 8.6.2 Multimodal Information Fusion: Complete Recursive Results

# How Error Changes over Three Predict–Correct Steps

| Step | Obs. / m | Prior / m | Gain | Posterior / m | Variance / m² |
|---|---|---|---|---|---|
| 1 | 10.8 | 11.00000 | 0.76316 | 10.84737 | 0.06868 |
| 2 | 12.2 | 11.84737 | 0.54702 | 12.04026 | 0.04923 |
| 3 | 12.9 | 13.04026 | 0.49786 | 12.97043 | 0.04481 |

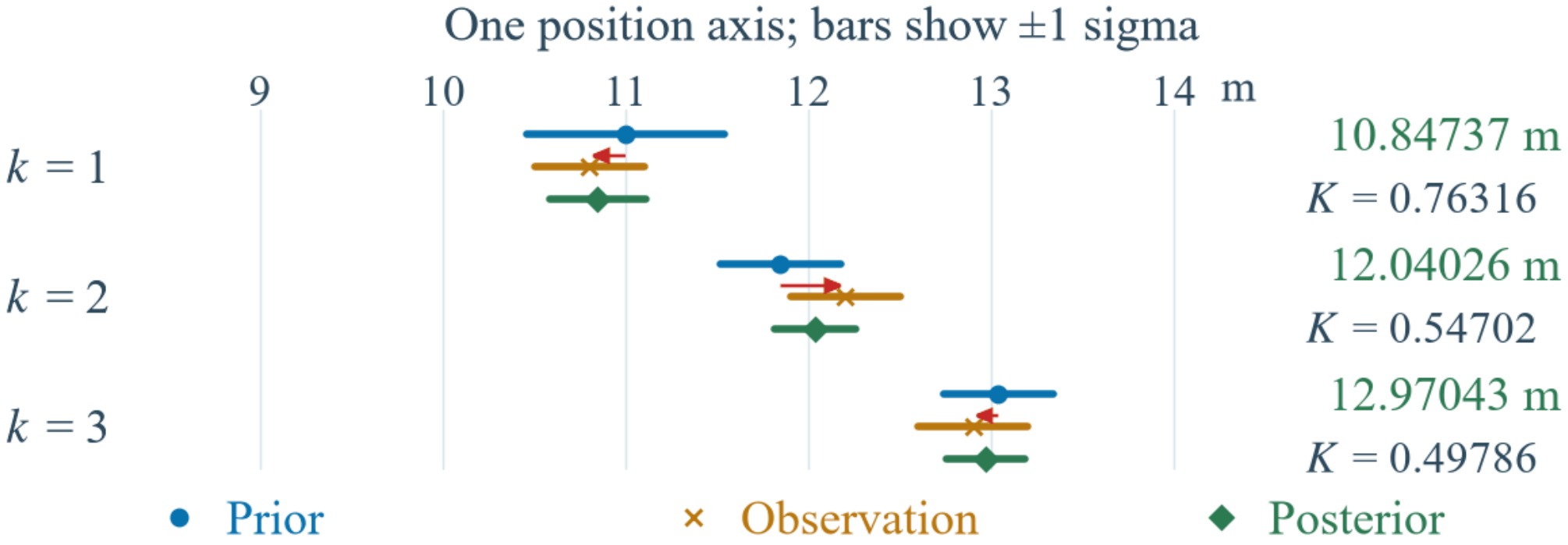
Preset truth: 11, 12, 13 m. Posterior errors: −0.15263, +0.04026, −0.02957 m.

Offline recursive example with specified model and observations.

### 8.6.2 Multimodal Information Fusion: Position Estimation over Time

# Three Predictions and Updates Share One Position Axis

Predict 1 m from each posterior; red arrows show residuals from prediction to observation.




Three recursive position estimates and uncertainty changes (offline calculation).

### 8.6.2 Multimodal Information Fusion: Confidence and Correlation

# Shared Errors Limit the Benefits of Sensor Fusion

**Shared GNSS-origin bias** → **Camera / LiDAR shift together** → **Averaging retains the bias**

$$z_1 = x + b + \epsilon_1, \quad z_2 = x + b + \epsilon_2$$

$x$: true value; $b$: shared bias; $\epsilon_1, \epsilon_2$: each sensor's random error.

$$\lambda_1 + \lambda_2 = 1 \implies \hat{x} = x + b + \lambda_1\epsilon_1 + \lambda_2\epsilon_2$$

Calibrate confidence from residuals and quality; estimate error scales and cross-covariance. Deduplicate sources; predict through gaps; do not treat correlated observations as independent.

**More sensors cannot replace independent information.**

Chapter Outline

# Multimodal Fusion and Control Decision-Making

**Chapter 8 Multimodal Perception and Intelligent Decision-Making**


8.1 Image Processing Fundamentals

8.2 Recognition, Tracking, Optical Flow

8.3 Localization, Mapping, and Planning

8.4 Common Perception Sensors

8.5 Coordinate Transformations

**8.6 Fusion and Control Decision-Making**

8.7 Tools and Framework Interfaces

8.8 Experiments and Real-World Transfer

8.9 Exercises and Classroom Discussion


## 5 Units in This Section

**8.6.1 Multimodal Perception Modeling**
Textbook: Sec. 8.1.6.1

**8.6.2 Multimodal Information Fusion**
Textbook: Sec. 8.1.6.2

**8.6.3 Trajectory Generation and Obstacle Avoidance Planning**
Textbook: Sec. 8.1.6.3

**8.6.4 Trajectory Tracking and External Control**
Textbook: Sec. 8.1.6.4

**8.6.5 Intelligent Planning and Decision-Making**
Textbook: Sec. 8.1.6.5

### 8.6.3 Trajectory Generation and Obstacle Avoidance Planning: Dynamic Obstacle Prediction

# A Prediction Horizon Contains Several Future Positions

$$\hat{\mathbf{p}}_{t+1:t+H}^{(i)} = \text{Predict}(\{\mathbf{p}_{t-k}^{(i)}\}_{k=0}^{K}) \tag{8.18}$$

Input: object $i$ has $K+1$ historical positions and timestamps. Output: positions and times for the next $H$ steps.

| Prediction Approach | Model / Data Conditions | Watch For |
|---|---|---|
| **Constant velocity / Kalman filter** | Short-term motion model and state noise | Maneuver error and covariance growth |
| **Recurrent networks / long short-term memory / Transformer** | Temporal patterns in training data | Distribution shift and prediction uncertainty |

### 8.6.3 Trajectory Generation and Obstacle Avoidance Planning: A Prediction Example

# One-Second Steps Give a Three-Second Position Sequence

$$\mathbf{p}_{t+j} = \mathbf{p}_t + j\Delta t\mathbf{v}_t, \quad \mathbf{p}_t = [3,-1]^{\mathrm{T}}\ \mathrm{m},\ \mathbf{v}_t = [0,1]^{\mathrm{T}}\ \mathrm{m/s}$$

| Future Time / s | Obstacle (N, E) / m | Vehicle Plan (N, E) / m | Same-Time Check |
|---|---|---|---|
| 1 | (3,0) | (2,0) | Centers 1 m apart |
| 2 | (3,1) | (3,1) | Coincide; reject |
| 3 | (3,2) | (4,2) | Centers 1 m apart |

Compare positions at the same future time; add safety margins for object size and uncertainty.

### 8.6.3 Trajectory Generation and Obstacle Avoidance Planning: The Obstacle-Avoidance Objective

## Waypoint Differences and Obstacle Costs Play Different Roles

$$\min_{\mathbf{x}_{0:T}} \mathcal{J} = \sum_{t=0}^{T-1} \|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2 + \lambda \sum_{t=0}^{T} C(\mathbf{x}_t) \tag{8.19}$$

Boundary conditions: $\mathbf{x}_0 = \mathbf{x}_{\text{start}}$, $\mathbf{x}_T = \mathbf{x}_{\text{goal}}$. Here $\mathbf{x}_t \in \mathbb{R}^3$ is position in m.

| Term | Role | Units Here |
|---|---|---|
| **Squared waypoint difference** | Penalizes displacement; does not ensure continuous acceleration | $m^2$ |
| **Obstacle penalty C** | Nonnegative penalty that grows near obstacles | $m^2$; $\lambda$ is dimensionless |

### 8.6.3 Trajectory Generation and Obstacle Avoidance Planning: Obstacle Costs

# A Penalty Converts Safety Clearance into Obstacle Cost

Start / goal: (0, 0) / (4, 0) m. Obstacle: center (2, 0) m, radius 0.5 m. Required clearance: 0.5 m.

Signed surface clearance: $d(\mathbf{x}_1) = y - 0.5$ m; $\lambda = 10$. Cost units: m².

$$C(\mathbf{x}) = \max(d_{\text{safe}} - d(\mathbf{x}), 0)^2, \quad d_{\text{safe}} = 0.5\ \text{m}$$

$$\mathbf{x}_1 = [2\ \text{m}, y]^{\mathsf{T}},\ y \geq 0: \quad C(\mathbf{x}_1) = \max(1\ \text{m} - y, 0)^2$$

| Midpoint y / m | Motion Cost / m² | Obstacle Cost / m² | Total J / m² |
|---|---|---|---|
| **0** | 8 | 1 | 8+10×1=18 |
| **1.5** | 12.5 | 0 | 12.5 |

## 8.6.3 Trajectory Generation and Obstacle Avoidance Planning: Complete Optimization Results

# The Soft-Cost Minimum Still Violates Safety Clearance

$$J(y) = 8\ \mathrm{m}^2 + 2y^2 + 10\max(1\ \mathrm{m} - y, 0)^2$$

For $0 \le y \le 1$ m, $J'(y) = 24y - 20$ m. Setting the derivative to zero gives:

$$y^* = 5/6\ \mathrm{m}, \quad J(y^*) = (29/3)\ \mathrm{m}^2 \approx 9.6667\ \mathrm{m}^2$$

$$d_{\mathrm{segment}} = \frac{(2\ \mathrm{m})y}{\sqrt{4\ \mathrm{m}^2 + y^2}} \implies d_{\mathrm{segment}}(5/6\ \mathrm{m}) = 10/13\ \mathrm{m}$$

| Candidate | Segment Clearance / m | Required: 0.5 m |
|---|---|---|
| **Soft minimum: y = 5/6** | 0.26923 | Fails |
| **Detour: y = 1.5** | 0.70000 | Pass; check dynamics |

### 8.6.3 Trajectory Generation and Obstacle Avoidance Planning: Solving and Replanning

## Static Local Optimization and Dynamic Replanning Have Roles

| Solver | Core Mechanism | Shared Limits |
|---|---|---|
| **Gradient descent** | Iterate along the negative gradient | Step size, initialization, local minima |
| **TrajOpt / CHOMP** | Sequential convexification / covariant gradients | Nonconvex constraints and collision checks |
| **GPMP2** | Gaussian processes and sparse factor graphs | Models, initialization, constraint formulation |

Dynamic scenes need prediction and replanning. Check continuous clearance, position/velocity/acceleration (PVA), and actuator limits.

**Lower soft cost does not make a trajectory executable.**

Chapter Outline

# Multimodal Fusion and Control Decision-Making

**Chapter 8 Multimodal Perception and Intelligent Decision-Making**


8.1 Image Processing Fundamentals

8.2 Recognition, Tracking, Optical Flow

8.3 Localization, Mapping, and Planning

8.4 Common Perception Sensors

8.5 Coordinate Transformations

**8.6 Fusion and Control Decision-Making**

8.7 Tools and Framework Interfaces

8.8 Experiments and Real-World Transfer

8.9 Exercises and Classroom Discussion


## 5 Units in This Section

**8.6.1 Multimodal Perception Modeling**
Textbook: Sec. 8.1.6.1

**8.6.2 Multimodal Information Fusion**
Textbook: Sec. 8.1.6.2

**8.6.3 Trajectory Generation and Obstacle Avoidance Planning**
Textbook: Sec. 8.1.6.3

**8.6.4 Trajectory Tracking and External Control**
Textbook: Sec. 8.1.6.4

**8.6.5 Intelligent Planning and Decision-Making**
Textbook: Sec. 8.1.6.5

### 8.6.4 Trajectory Tracking and External Control: Position, Velocity, and Acceleration

## A Quintic Time Function Generates Continuous PVA

Single-axis displacement: 4 m over 4 s. Normalized time: $\tau = t/(4\ \mathrm{s}), \quad 0 \le t \le 4\ \mathrm{s}$.

$$p(t) = 4(10\tau^3 - 15\tau^4 + 6\tau^5)\ \mathrm{m}, \quad \tau = t/(4\ \mathrm{s})$$

$$v(t) = (30\tau^2 - 60\tau^3 + 30\tau^4)\ \mathrm{m/s}$$

$$a(t) = (60\tau - 180\tau^2 + 120\tau^3)/4\ \mathrm{m/s^2}$$

At t = 1 s: P = 0.4140625 m; V = 1.0546875 m/s; A = 1.40625 m/s².

### 8.6.4 Trajectory Tracking and External Control: Trajectory Samples

## Whole-Second References from the Same Time Function

| Time / s | P / m | V / (m/s) | A / (m/s²) |
|---|---|---|---|
| 0 | 0.0000000 | 0.0000000 | 0.00000 |
| 1 | 0.4140625 | 1.0546875 | 1.40625 |
| 2 | 2.0000000 | 1.8750000 | 0.00000 |
| 3 | 3.5859375 | 1.0546875 | -1.40625 |
| 4 | 4.0000000 | 0.0000000 | 0.00000 |

Midpoint speed reaches 1.875 m/s; acceleration changes from positive to negative. Whole-second samples do not check continuous limits.

### 8.6.4 Trajectory Tracking and External Control: Choosing Control Quantities

# The Interface Level Determines Which Inner Loops Remain

| Reference | Typical Task | External Controller Handles |
|---|---|---|
| **Position P** | Navigation, station keeping, waypoints | Timed references and velocity / acceleration feedforward |
| **Velocity V / acceleration A** | Visual servoing, avoidance, feedforward | Feedback error and limits |
| **Attitude / thrust** | Large maneuvers, sensor line of sight | Translational stabilization and thrust limits |
| **Angular rate / thrust** | Highly dynamic low-level control | Attitude stabilization and greater safety responsibility |

Control level, feedback state, and flight-controller inner loops determine feasibility.

### 8.6.4 Trajectory Tracking and External Control: External Control Interfaces

## Offboard References Need Frames, Units, and a Continuous Stream

| Message Semantics | Reference Frame and Units |
|---|---|
| **Position / velocity / acceleration** | Local NED; m / $m \cdot s^{-1}$ / $m \cdot s^{-2}$ |
| **Attitude / angular rate** | FRD→NED quaternion; body angular rate in rad/s |
| **Mode keepalive** | >2 Hz; stream for >1 s before entry; check loss-of-link action |

PX4 v1.16 rotates MAVLink BODY_NED velocity references by heading only; this differs from a full FRD rotation with roll and pitch.

Conditions here are for PX4 v1.16; check the actual firmware and SDK:
https://docs.px4.io/v1.16/en/flight_modes/offboard

### 8.6.4 Trajectory Tracking and External Control: Visual Feedback Control

# Scale Converts Pixel Error into a Velocity Reference

Ideal stationary target: $Z = 4$ m, $f_x = 320$ px, $e_{u,0} = 160$ px, $K_p = 0.5\ \mathrm{s}^{-1}$.

$$X = \frac{Z}{f_x} e_u = \frac{4}{320}(160) = 2\ \mathrm{m}, \quad v_{\mathrm{right}} = K_p X = 0.5(2) = 1\ \mathrm{m/s}$$

$$e_{u,k+1} = (1 - K_p \Delta t) e_{u,k} = 0.95 e_{u,k}, \quad \Delta t = 0.1\ \mathrm{s}$$

| Quantity / Step | k0 | k1 | k2 | k3 |
|---|---|---|---|---|
| **Pixel error / px** | 160.00 | 152.00 | 144.40 | 137.18 |
| **Rightward error / m** | 2.00000 | 1.90000 | 1.80500 | 1.71475 |
| **Right velocity / m/s** | 1.000000 | 0.950000 | 0.902500 | 0.857375 |

## 8.6.4 Trajectory Tracking and External Control: Joint Position and Attitude Control

# Motion Trajectories and Line-of-Sight Goals Jointly Constrain Actions

MPC: model predictive control.

**Position / PVA task** → **Attitude / line of sight** → **Joint predictive optimization** → **Low-level tracking**

| Constraint Type | Objective or Limit Here | Check Before Execution |
|---|---|---|
| **Position and heading** | Move north; camera faces east, yaw = 90° | Define translation and orientation separately |
| **Attitude and input** | Limit tilt, thrust, and angular rate | Match the model to the control interface |
| **MPC horizon** | Predict states; choose a feasible control sequence | Online computation time and fallback strategy |

Chapter Outline

# Multimodal Fusion and Control Decision-Making

**Chapter 8 Multimodal Perception and Intelligent Decision-Making**


8.1 Image Processing Fundamentals

8.2 Recognition, Tracking, Optical Flow

8.3 Localization, Mapping, and Planning

8.4 Common Perception Sensors

8.5 Coordinate Transformations

**8.6 Fusion and Control Decision-Making**

8.7 Tools and Framework Interfaces

8.8 Experiments and Real-World Transfer

8.9 Exercises and Classroom Discussion


## 5 Units in This Section


**8.6.1 Multimodal Perception Modeling**
Textbook: Sec. 8.1.6.1

**8.6.2 Multimodal Information Fusion**
Textbook: Sec. 8.1.6.2

**8.6.3 Trajectory Generation and Obstacle Avoidance Planning**
Textbook: Sec. 8.1.6.3

**8.6.4 Trajectory Tracking and External Control**
Textbook: Sec. 8.1.6.4

**8.6.5 Intelligent Planning and Decision-Making**
Textbook: Sec. 8.1.6.5

### 8.6.5 Intelligent Planning and Decision-Making: Decision-Making Methods

## Decision Models and Implementation Architectures Can Combine

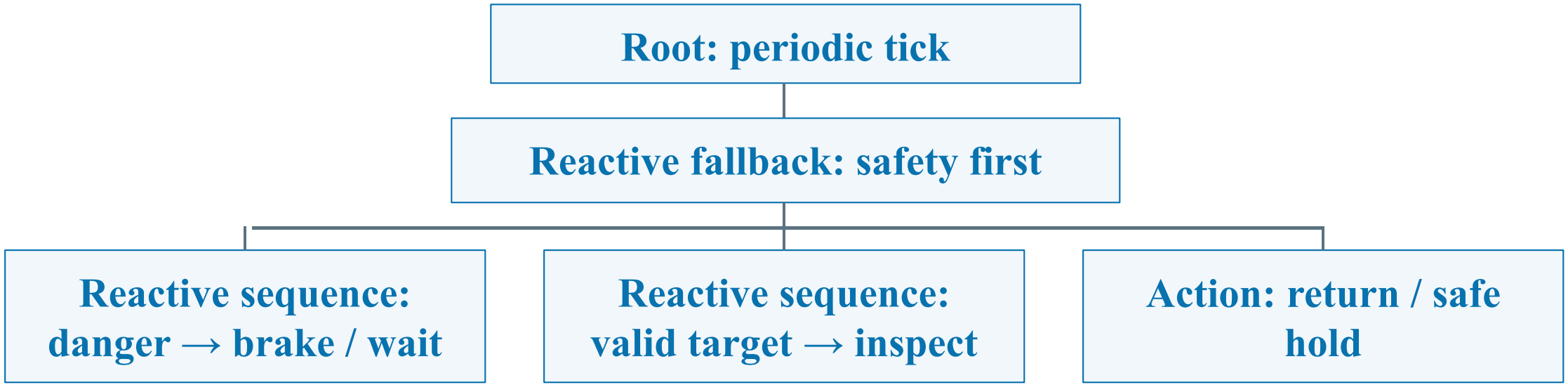
FSM: finite state machine; HFSM: hierarchical FSM; DRL: deep reinforcement learning.

| Approach | Input → Output | Conditions and Limits |
|---|---|---|
| **Behavior tree (BT)** | Conditions / action status → node | Reuse modules; define status |
| **FSM / HFSM** | Events / guards → task state | Clear structure; complex at scale |
| **DRL** | State / reward → action policy | Training, generalization, safety |
| **End-to-end** | Raw images / clouds → actions | Joint optimization; harder diagnosis |
| **Foundation model** | Language / state → symbolic flow | Semantic capability; compute, latency, validation |

Explicit task logic, learned local policies, and stable control can form separate layers.

### 8.6.5 Intelligent Planning and Decision-Making: Behavior Trees

# Behavior Trees Advance by Success, Failure, or Running




**Sequence:** return on Failure / Running; succeed only if all children return Success.
**Fallback:** return on Success / Running; fail only if all children return Failure.
**Parallel:** trigger children concurrently; configure success / failure thresholds for the task.
**Reactive nodes:** recheck conditions each tick and interrupt the current action when needed.

### 8.6.5 Intelligent Planning and Decision-Making: State Machines and Hierarchical State Machines

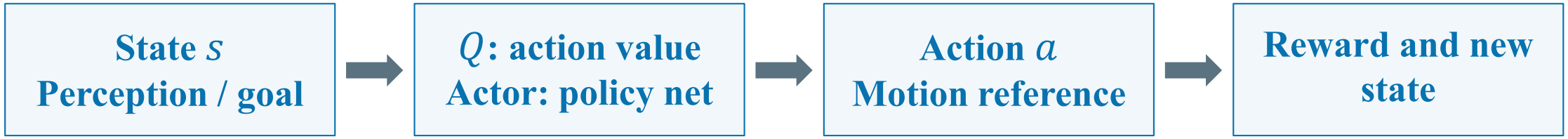
# States and Guards Determine Behavior Transitions

| Event / Guard | State Transition | Action |
|---|---|---|
| **Valid target; clear route** | Search → track | Generate a target-tracking reference |
| **Predicted conflict** | Track → detour | Generate a checked detour trajectory |
| **Observation uncertainty too high** | Detour → wait | Brake; wait for valid information |
| **Task complete / return condition** | Any permitted state → return | Plan return; maintain feedback |

Preset events: search → track → detour → wait → return.
An HFSM can divide tracking into approach, align, and capture.

### 8.6.5 Intelligent Planning and Decision-Making: Learned Decision-Making

## Reinforcement Learning and End-to-End Architectures Can Overlap




$$r = \text{progress} - 0.1 - 100\,\mathbf{1}_{\text{collision}}$$

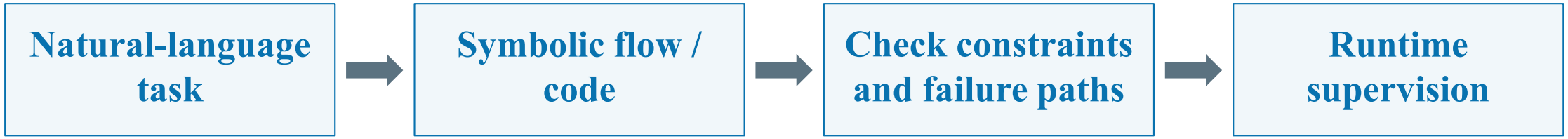
progress: meters advanced.
$\mathbf{1}_{\text{collision}}$: 1 if collision; else 0.

| Action and Result | One-Step Reward | Interpretation |
|---|---|---|
| **Advance 1 m; no collision** | 0.9 | Progress 1 minus action cost 0.1 |
| **Advance 1 m; collision** | -99.1 | Subtract collision penalty 100 too |
| **Wait; no collision** | -0.1 | No progress reward; action cost remains |

End-to-end maps raw perception to actions. Learned reward is not a hard safety constraint.

### 8.6.5 Intelligent Planning and Decision-Making: Foundation-Model Decision Logic

# Language Tasks Become Inspectable Symbolic Flows




**Task:** inspect A → inspect B → return; confirm completion at each stop before advancing.
**Exceptions:** wait, replan, or return if targets are invalid, risk is high, or energy is low.
**Deployment:** check logic, computation latency, and compression effects; keep high-rate stabilization independent.

Language models can assist initial task logic; observation feedback and safety constraints remain active at runtime.

### 8.6.5 Intelligent Planning and Decision-Making: A Complete Decision-Chain Example

## Preset Observations Produce Checkable Actions Across the Loop

| Time | Perception / Fusion / Prediction | Decision | Planning and Control Output |
|---|---|---|---|
| t0 | Same-axis estimate 9.92 m; clear route | Track | Reference checked for clearance and PVA |
| t1 | Conflict predicted at 2 s; valid map | Detour | Reject conflict; solve and check a detour |
| t2 | Occluded; uncertainty above threshold | Wait | Limited braking; retain safe feedback |
| t3 | Observations recover; detour route clear | Track | Replan; continue feedback correction |
| t4 | Inspection complete; return conditions met | Return | Return trajectory; check link loss and energy |

Offline: BT/FSM task logic, local learned actions, and real-time control have distinct roles.

Exercises and Discussion

# Four Checks on Measurement, Fusion, Planning, and Control

1. Stationary inertial case: level FRD body, $\mathbf{f} = [0,0,-9.81]^{\mathrm{T}}\ \mathrm{m/s^2}$, no bias. What is NED motion acceleration?

**2. Radar and fusion:** does radial velocity determine tangential velocity?
Fuse independent unbiased positions 10.4 / 9.8 m with variances 0.16 / 0.04 m².

3. Soft-cost path: for $y = 5/6$ m, segment-to-center distance is 10/13 m and radius is 0.5 m. Is 0.5 m clearance met?

4. Control and decisions: $e_u = 160$ px, $Z = 4$ m, $f_x = 320$ px, $K_p = 0.5\ \mathrm{s^{-1}}$. Find rightward velocity. Who organizes task switching?

## Reference Points

# Physical Meaning and Execution Constraints Determine the Answers

1. Motion acceleration: $\mathbf{R}(\mathbf{f} - \mathbf{b}_a) + \mathbf{g} = \mathbf{0}$ $\mathrm{m/s^2}$. Do not integrate raw specific force directly.

**2. Observability and weighting:** tangential velocity is undetermined. Weights: 0.2 / 0.8; estimate: 9.92 m; variance: 0.032 m².

**3. Continuous safety:** clearance 0.26923 m < 0.5 m. Reject the path; add continuous-clearance and dynamic constraints.

**4. Interface roles:** rightward velocity is 1 m/s. Task logic selects track / wait states; the stable feedback loop executes.

## Section Summary

# Reliable Estimates, Feasible Trajectories, and Decisions Work Together

**Measurement and fusion:** align quantities, time, frames, and objects first. Confidence and correlation determine useful information.

**Prediction and planning:** compare at the same future time; verify continuous clearance, velocity, acceleration, and input limits.

**Control and decisions:** PVA and interface conventions govern execution; explicit logic and learned policies organize task switching.

**Key distinctions:** specific force ≠ motion acceleration; radial ≠ full velocity; soft cost ≠ safety guarantee. Next: tools and frameworks in 8.7.