Glider MDP

The Outliers
Gary Ellingson & Dan Koch

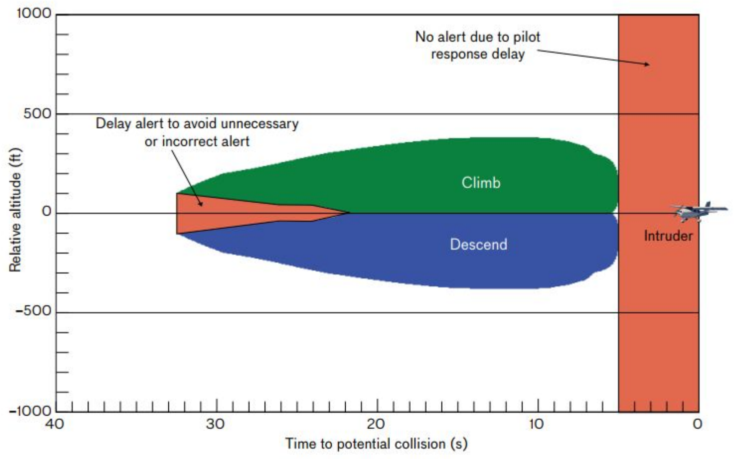
Problem Statement

- Markov decision process
- Longitudinal control of a glider to reach a target
- Target: 1 meter above ground at the end of the room
- States observed with motion capture

Objectives & Motivation

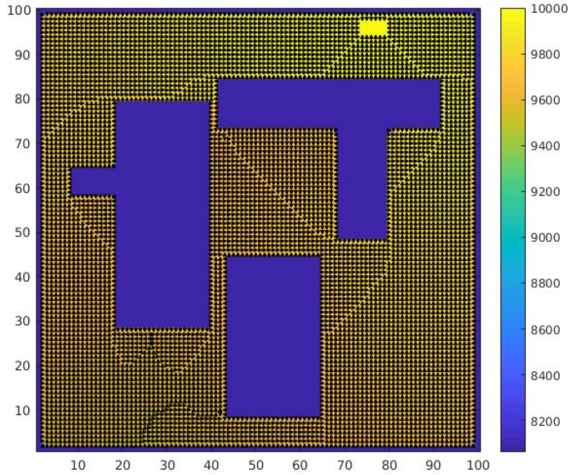
- Learn about Markov Decision Processes
- Apply them to a “real” problem

$$V_\infty(\mathbf{x}) = \gamma \max_{\mathbf{u}} \left[r(\mathbf{x}, \mathbf{u}) + \sum_j V(\mathbf{x}_j)\, p(\mathbf{x}_j | \mathbf{x}, \mathbf{u}) \right]$$



Challenges

- Transition probability model
  - Based on UAV book dynamics
  - Use distributions of points to represent uncertainty
- Computational tractability
  - State dimensionality
  - Pre-compute transition probabilities
  - Position-invariant computation
- Hardware / delay
  - Arduino PPM trainer port setup



State space

$$\mathbf{x} = \begin{bmatrix} x \\ h \\ u \\ w \\ \theta \\ q \end{bmatrix}$$

$$\mathbf{u} = \delta_e$$

- State space: longitudinal states
- Action space: elevator deflection

- Discretize state space into coarse grid
  - x and h: 0.25m step size, u and w: 1.0m/s step size, θ: 10deg step size
  - Resulting grid: 23*10*7*5*5*5 = 201,250 elements
- Discretize action space
  - $\delta_e \in \{-30°, -15°, 0°, 15°, 30°\}$
  - Resulting search space: 201,250*5 = 1,006,250 elements

System dynamics

Assuming $\phi = p = r = v = 0$ and with $\alpha = \arctan(w/u)$, $V_a = \sqrt{u^2 + w^2}$:

$$\dot{u} = -qw - g\sin\theta + \frac{\rho V_a^2 S}{2m}\left[C_{X_0} + C_{X_\alpha}\alpha + C_{X_q}\frac{cq}{2V_a} + C_{X_{\delta_e}}\delta_e\right]$$

$$\dot{v} = qu + g\cos\theta + \frac{\rho V_a^2 S}{2m}\left[C_{Z_0} + C_{Z_\alpha}\alpha + C_{Z_q}\frac{cq}{2V_a} + C_{Z_{\delta_e}}\delta_e\right]$$

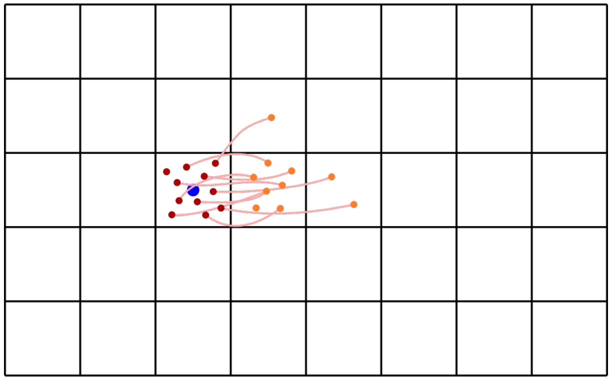
$$\dot{q} = \frac{1}{2J_y}\rho V_a^2 cS\left[C_{m_0} + C_{m_\alpha}\alpha + C_{m_q}\frac{cq}{2V_a} + C_{m_{\delta_e}}\delta_e\right]$$

$$\dot{\theta} = q\cos\theta$$

$$\dot{h} = u\sin\theta - w\cos\theta$$

Transition probability model

- Propagate collection of noisy points through dynamics
    - Uncertainty on aerodynamic coefficients
    - Uncertainty on actual state value within grid cell
- Probability equal to fraction of points that end up in a given grid cell
- Used 500 sample points

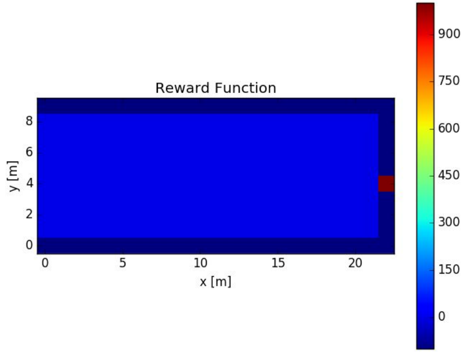


Value iteration

- Standard value iteration

$$V_{\infty}(\mathbf{x}) = \gamma \max_{\mathbf{u}} \left[r(\mathbf{x}, \mathbf{u}) + \sum_j V(\mathbf{x}_j)\, p(\mathbf{x}_j | \mathbf{x}, \mathbf{u}) \right]$$

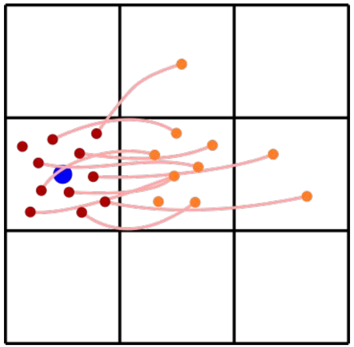
- Reward function:
    - -100 at walls and ceiling
    - +1000 for hitting target



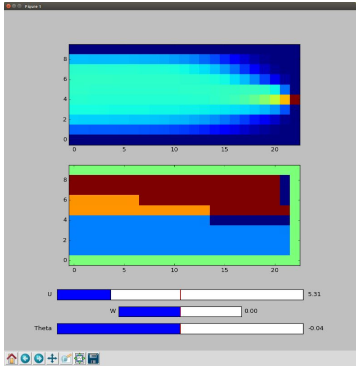
Computational speedups

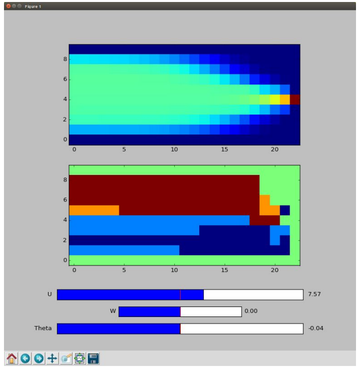
- Pre-compute transition probabilities
  - Over 1 million transition probability function calls per iteration
- Position-invariant transition probability

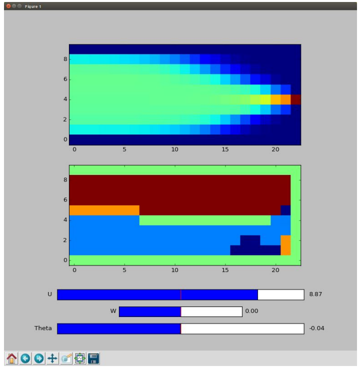
$$p\left(\begin{bmatrix} x' \\ z' \\ u' \\ w' \\ \theta' \\ q' \end{bmatrix} \middle| \begin{bmatrix} x \\ z \\ u \\ w \\ \theta \\ q \end{bmatrix}, \delta_e\right) = p\left(\begin{bmatrix} x' - x \\ z' - z \\ u' \\ w' \\ \theta' \\ q' \end{bmatrix} \middle| \begin{bmatrix} 0 \\ 0 \\ u \\ w \\ \theta \\ q \end{bmatrix}, \delta_e\right)$$

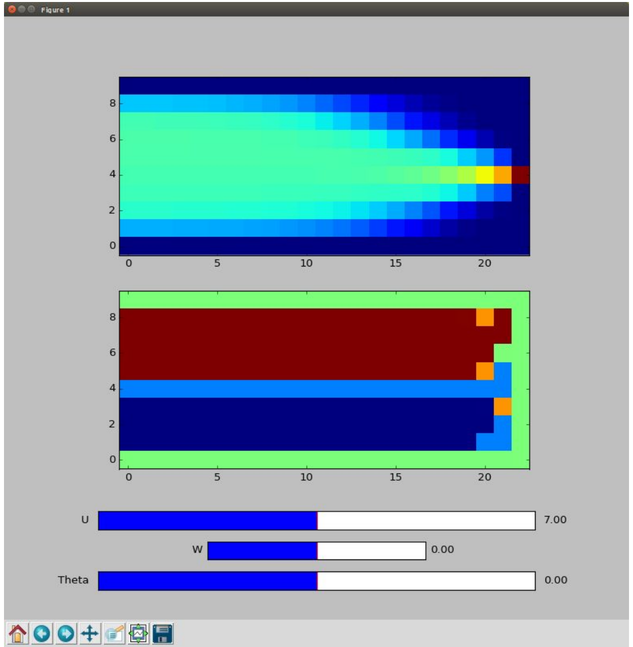


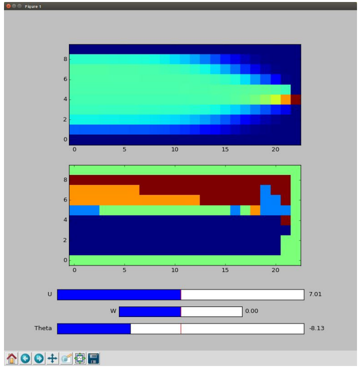
Value iteration results

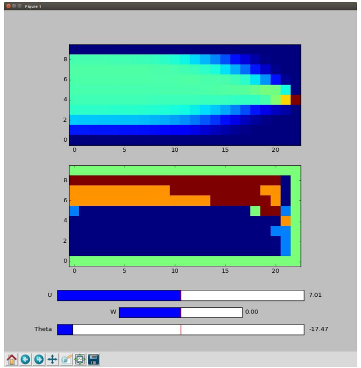


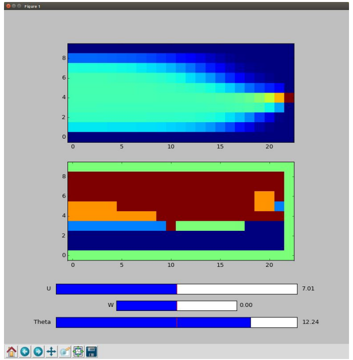




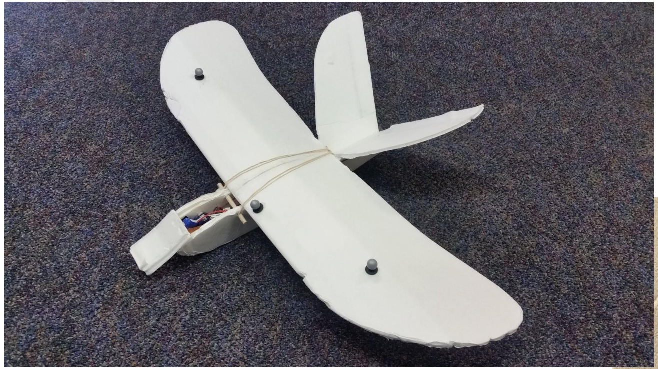


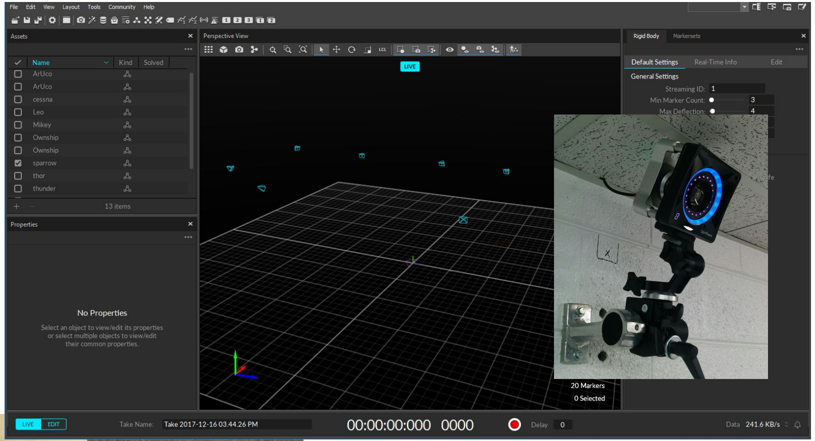


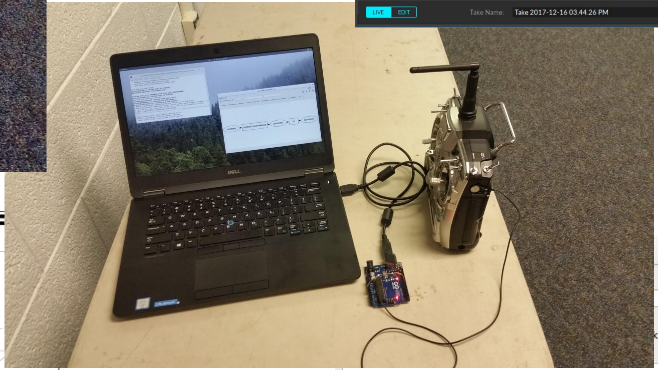


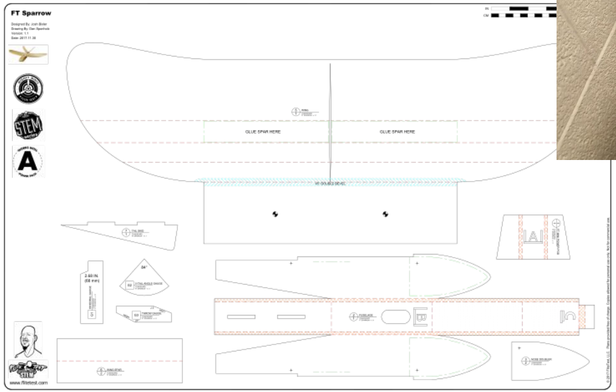


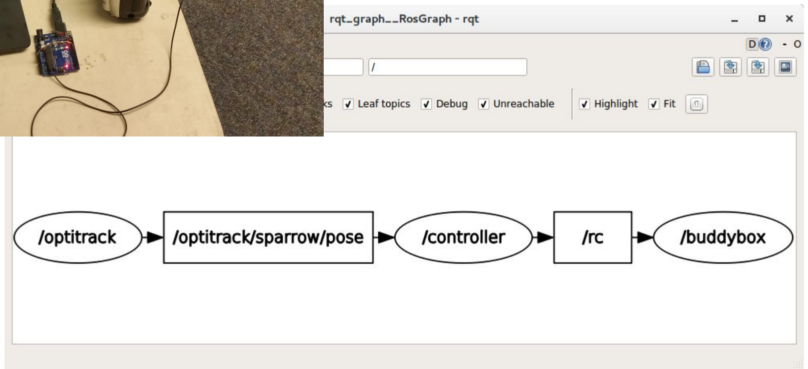
Hardware overview





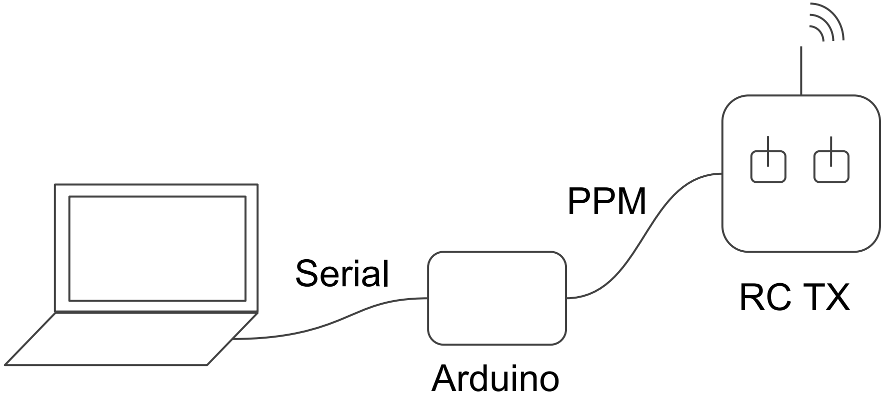






Hardware challenges

- RC buddybox delay
- Servo delay
- Airplane (weight)
- Size of the room



Hardware results

Conclusions

What did we learn?

- Computational tricks are important for non-trivial problems
- Accurate modeling is key

Why was it useful?

- Learning
- Developed buddybox system

What else would we try if we had more time?

- Model the actuator delay
- Policy iteration, other variants
- Continuous-time MDPs
- Continuous state space