Control system for in-vivo micro scale surgical robot

Sheil Sarda

Jan 2023

Outline

Assumed Inputs for localization and navigation

Proposed methodology for

- (a) Localization
- (b) Path Planning
- (c) Trajectory Generation
- (d) Control System

Design Criteria and Key Risks

Appendix

.01

Assumed Inputs for localization and navigation

Assuming access to an intraoperative MRI feed / fluoroscopy

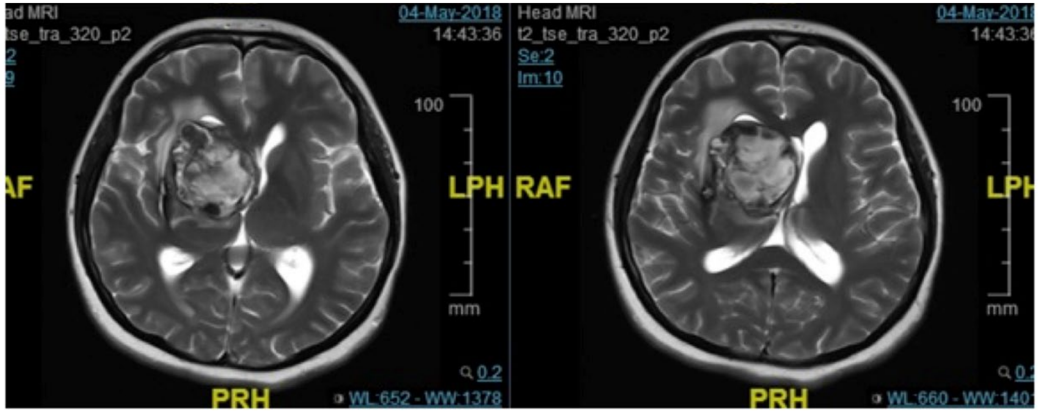


Fig. 1 Preoperative T2flair inspection showing occupancy with hemorrhage

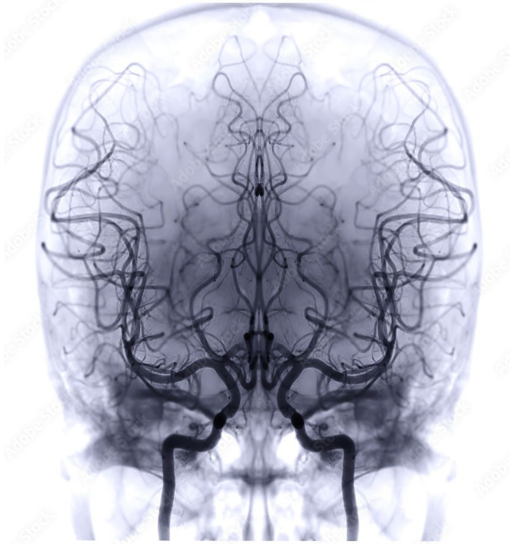


Fig. 3 Scan from cerebral angiography (performed using fluoroscopy

Fig. 2 Intraoperative MRI scan (T1) showing the occupancy and hemorrhage were totally removed

Fig. 4 Patient is positioned on an examination table during fluoroscopy

Source: Intraoperative MRI-assisted neuro-port surgery for the resection of cerebral intraparenchymal cavernous malformation (*Chinese Neurological Journal; 9/11/19)*

.02

Proposed methodology for:

(a) Localization

(b) Path Planning

(c) Trajectory Generation

(d) Control System

Designing an occupancy grid schema to keep track of obstacles in the workspace (in-vivo)

Potential schema for occupancy map:

Field	Description
Grid	2x2 array; cells represent square area of environment
➢ **Occupancy**	Probability cell is occupied by obstacle
➢ **Metadata**	• Type/classification of features (based on human anatomy) • Probability of a feature being a benign / malignant tumor
Time_Last_Updated	Timestamp map was last refreshed
Resolution	Resolution of map (corresponding to real-world scale)

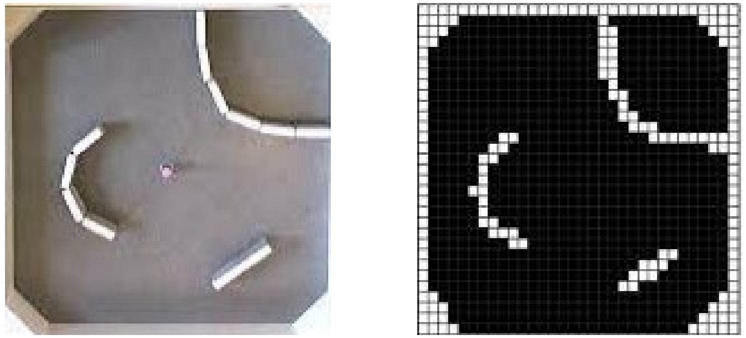


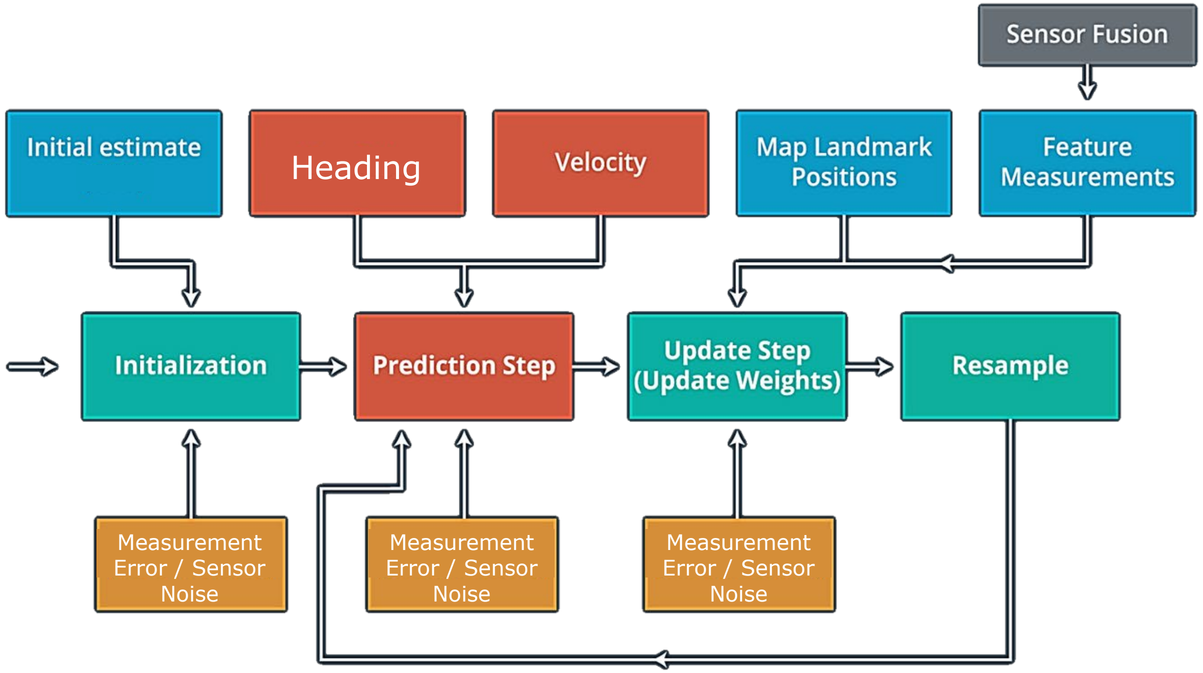
Fig. 5 Illustrative example of how an image of the environment (left) can be mapped into a binary occupancy grid (right)

Potential Algorithm to initialize and refresh occupancy grid:

- Maintain large occupancy grid for localization + smaller local map for locations of objects in the vicinity of the robot; Using a static state Bayes filter to estimate the state of each cell in the occupancy grid
- Have non-interruptible threads constantly parsing the image stream to refresh the occupancy grid
- Latency of occupancy map update loop needs to be lower than frequency of image stream updates
  - Ensures control systems using the occupancy map to plan next steps have latest inputs to work off
  - Needs to be enforced as a minimum performance guarantee, and achieved in practice by writing thread parallel code
- As a failsafe, control algorithm checks if input is out of date based on some key latency threshold, and halts if so

We can populate the occupancy grid using image filtering techniques & localize the robot using Bayesian estimators

Sketch of potential adaptive control localization algorithm

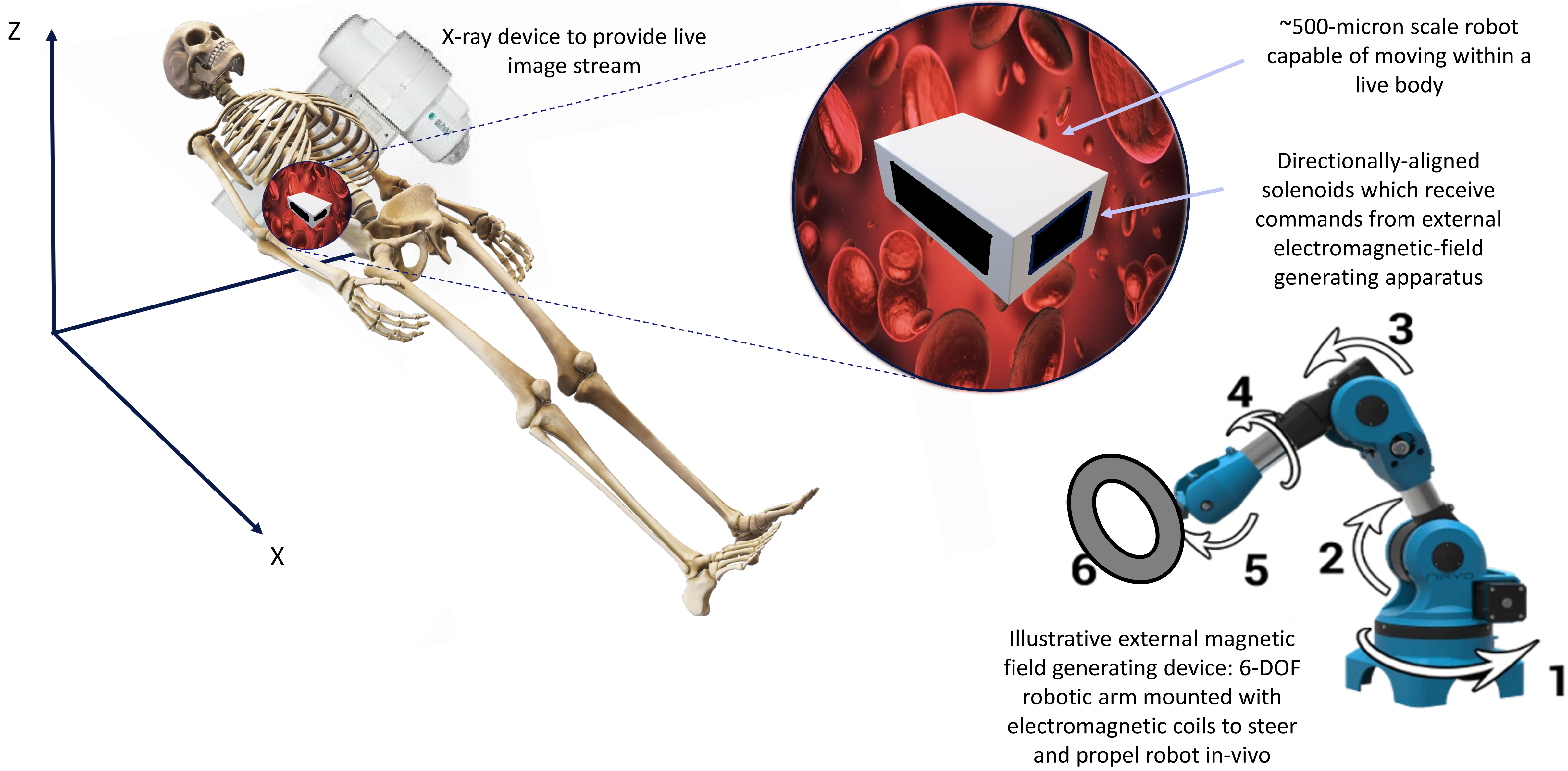


Potential image-filtering techniques to localize robot using X-rays, MRI, etc.

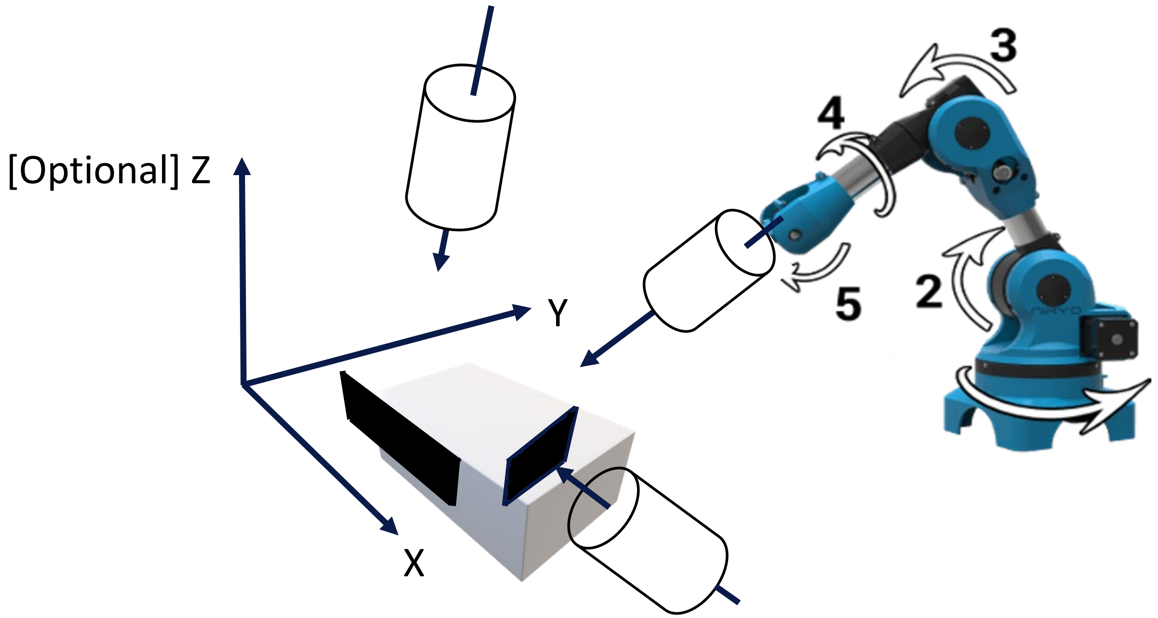
1. Measurement models can be used to specify the likelihood of robot and target location measurements obtained from the imaging system
2. Motion models to estimate the robot's future position given current state of the target and robot (via probability density functions for next states)
3. Bayesian filters to estimate the current robot and target state given output from measurement and motion models

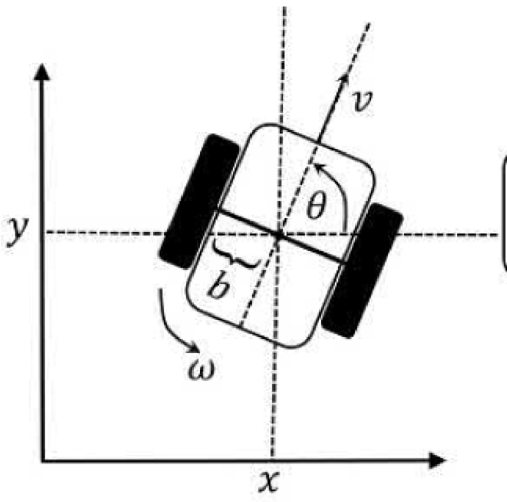
Source: High-performance localization software for autonomous vehicles (GitHub); Surgical Robotics, Systems Applications and Visions (Springer)

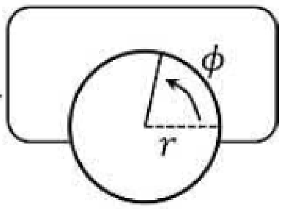
Illustrative Model of robot, magnetic field controller, X-ray



We can model this system in 2D as a differential drive mobile robot with a heading angle and velocity







$$\begin{aligned} \dot{x} &= \cos(\theta)v \\ \dot{y} &= \sin(\theta)v \\ \dot{\theta} &= \omega \end{aligned} \qquad \begin{aligned} v &= \frac{r}{2}(\dot{\phi}_R + \dot{\phi}_L) \\ \omega &= \frac{r}{2b}(\dot{\phi}_R - \dot{\phi}_L) \end{aligned}$$

(Assuming no wheel slipping)

Key considerations and limitations of this kinematic model

- One key difference between the differential drive robot and our design is that the wheels are replaced by external permanent magnets mounted on actuators such as the end-effector of a 6-DOF robotic arm
- Robot’s velocity and torque are based on the rotation of the externally mounted magnetic X, Y coils
- We choose the 2D analogy as a simplifying assumption, and subsequently our mechanical objective is to maneuver along a path towards a goal point, staying within a maximum acceptable deviation at all waypoints

Source: Robot Modeling and Control (Spong, Hutchinson, Vidyasagar)

Potential field and gradient descent approach to path planning

Overview and formulation of potential field approach

- Treat robot as a point particle in configuration space under the influence of a potential field
  - By design, the potential field attracts the robot to the final configuration and repels the robot from boundaries of the free configuration space
  - If possible, ensure potential field does not have local minima
- In practice, we implement this using a distinct attractive and repulsive potential field:
  - Attractive field: (i) monotonically increasing with distance to goal configuration; (ii) well-defined gradient; (iii) zero attractive force at the goal configuration
  - Repulsive field: to prevent robot from colliding with obstacles and when robot is far away from obstacles to exert minimal force. To facilitate this, define repulsive potential fields at each obstacle + every DH frame of the robot

Generating path using gradient descent

- As a refresher, this involves taking small steps in the direction of the negative gradient and repeating until the final configuration is reached; the configurations explored in this method form waypoints in our path
- Terminate when configuration reaches sufficiently near the goal configuration based on a tolerance selected for the task
- The problem that plagues all gradient descent algorithms is the possible existence of local minima in the potential field.
  - To detect these local minima, we can check if successive configurations lie within a small region of the configuration space, there is likely a local minima nearby
  - We can escape the local minima by defining a random walk once if/when the planner gets stuck; the random walk can be generated by adding a small fixed constant to current configuration

Interpolate between the waypoints generated by gradient descent using quintic polynomial trajectories

- For each joint in the robot, we define a trajectory in the configuration space which connects the current state to the next waypoint while satisfying velocity and acceleration constraints
- Smooth and continuous trajectories are essential to avoid impulsive movements which may excite vibrational modes in the manipulator & reduce tracking accuracy
- Given we have 6 constraints (2x for initial and final configuration; 2x for initial and final velocities; 2x for initial and final accelerations), we need a fifth order polynomial

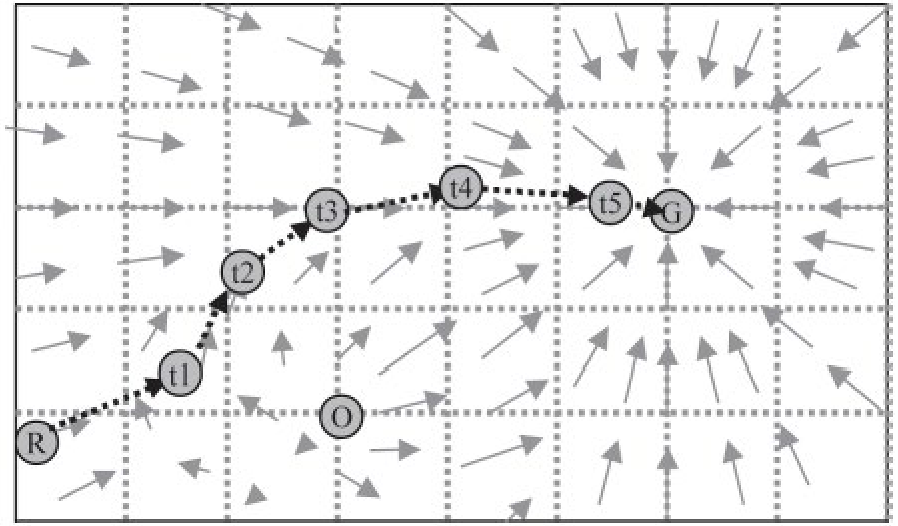


Fig. 6 Waypoints from start to goal configuration generated by gradient descent on potential fields

$$q(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5$$

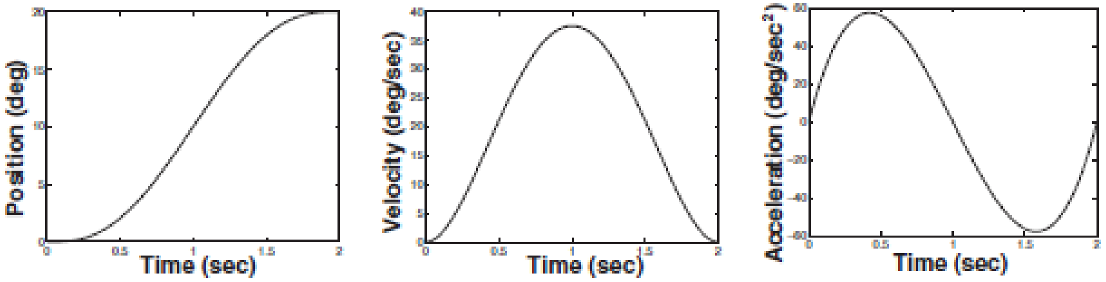
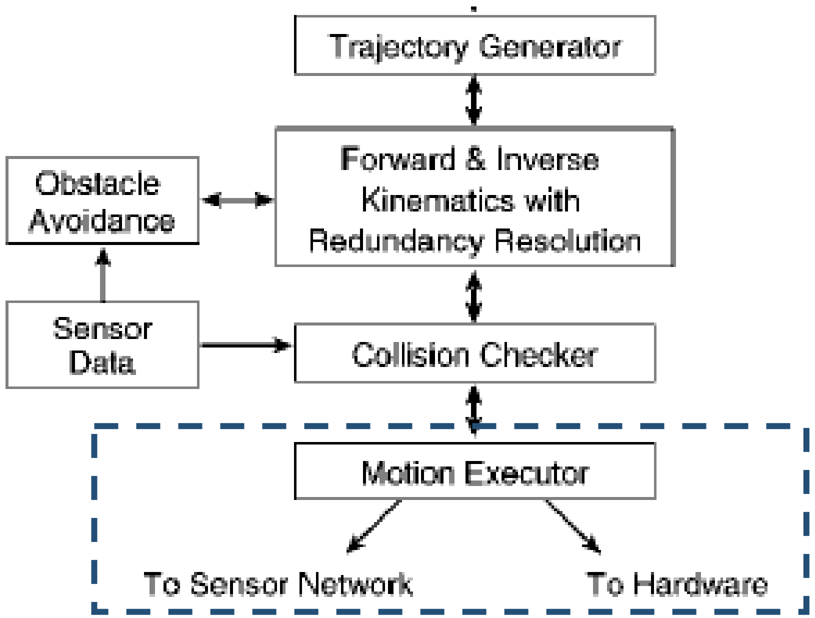


Fig. 7 General form of quintic polynomial (top) and quintic polynomial trajectory, velocity and acceleration profile (bottom)

Using PID loops to ensure robot follows the quintic polynomial-based approach discussed earlier

- Control loop sends waypoint position commands to the external device which generates electromagnetic field; position commands are translated into motor commands such as targeted angular velocity in the X and Y direction
- PID loops for each control variable use the robot’s position as feedback to speed up / slow down the robot, ensuring it reaches the waypoint quickly and accurately



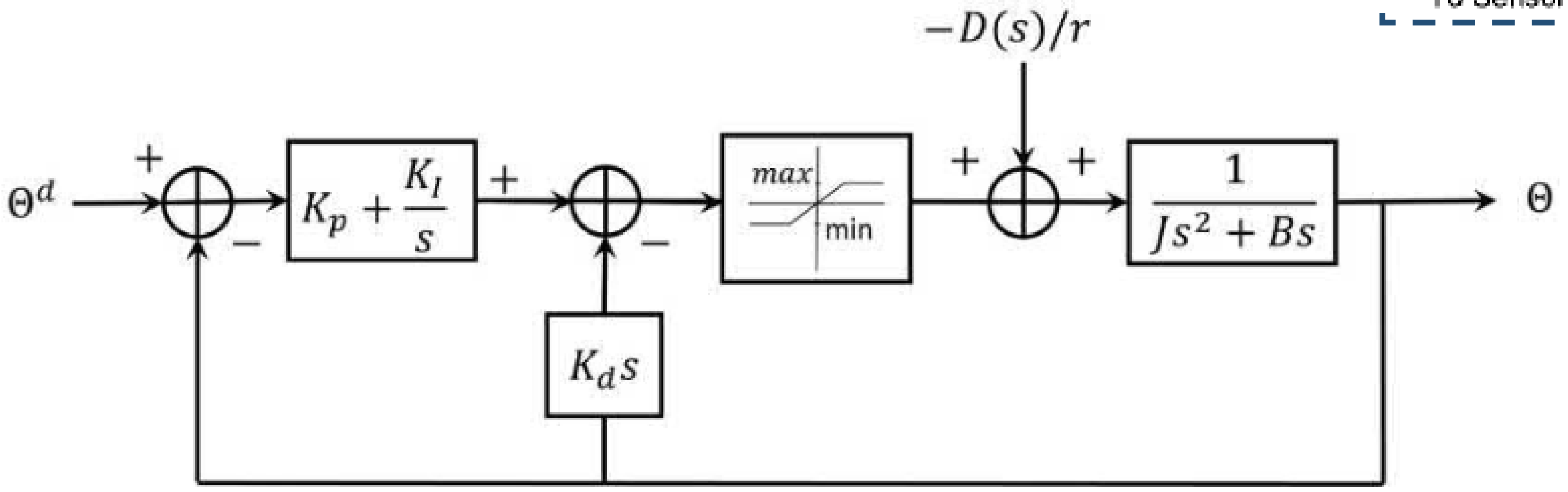


Fig. 8 (Top) PID loops exist within the motion executor module in this process diagram; (Left) Example PID loop for a control variable theta

.04
Design Criteria and Key Risks

Pros / Cons of medical vision modes which can be used for localization

While we are using X-ray imaging streams for this task, several other options also exist:

Methodology	(1) Fluoroscopy	(2) CT Scans	(3) Ultrasound	(4) MRI
Feasibility	• Our microbot, being of high density, should be visible in X-ray images • 3+ images + knowledge of human anatomy required to create 3D reconstruction	• Using continuous beam mode, CT scanners can provide low resolution ~10fps image streams • Can acquire multiple 3D slices in continuous beam mode	Can create 3D volume using tracked sweep of 2D imaging	Key advantage over (1) and (2) is absence of harmful radiation
Autonomous (Y/N)	Images must be acquired under command of human operator			To be further researched
Key risks	• Exposes patient and physician to excessive radiation unless precautionary measures are implemented		• Scanning may cause tissue deformation • Images tend to be noisy due to reflections, air pockets, etc.	• Acquisition of images is typically not real-time • Any metal creates large void in image

- The problem of localizing objects using the above medical imaging techniques images might seem straightforward, yet practical implementations have seldom appeared
- Usually, a cascade of basic image filters (such as thresholding, edge detection, image smoothing and noise removal filters) are combined with more sophisticated feature detections routines, such as a variant of Hough transform

Source: *Surgical Robotics, Systems Applications and Visions* (Springer)

Pros / Cons of potential field approach to path planning

Commonly used control systems to enable trajectory following in robotics include:

Path Planning Algorithm	Description + Pros	Cons + Key Risks
Potential Field w/ Gradient Descent	• This approach is well-suited for large configuration spaces with several obstacles where it’s hard to map free configuration space	• Susceptible to failure due to the presence of local minima in the potential field
A*	• Uses a heuristic function to prioritize exploring nodes in the robot’s free configuration space which have a lower estimated distance to the goal	• Computationally expensive when: (i) search space is large; (ii) dimensions of the search problem are high
RRT / RRT*	• Fully randomized approach, which runs several iterations of randomly adding points to a tree (if no collision) until a path to goal is found or time expires	• By default, RRT can result in quite long and windy paths which require being pruned down to the shortest possible collision-free path

- Ultimately decided to proceed with potential field approach to path planning since the configuration space for our robot may have many obstacles based on human anatomy that we are trying to steer clear of
- This would make generating the free configuration space map (which A* and RRT rely on) a computationally intensive exercisex

Pros / Cons of closed loop PID control vs adaptive control methods

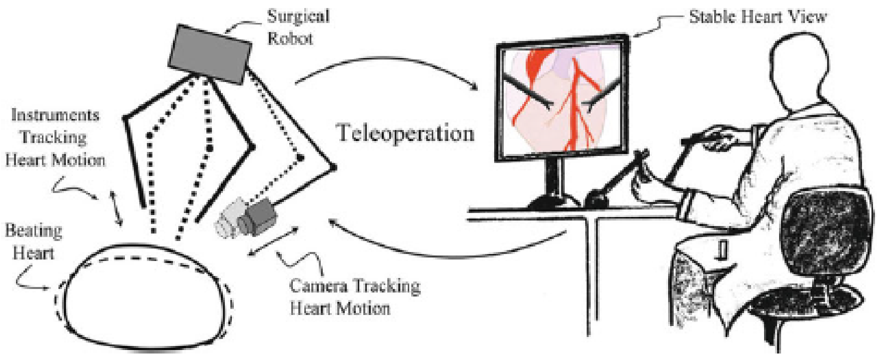
Commonly used control systems to enable trajectory following in robotics include:

Control System	Description + Pros	Cons + Key Risks
PID Control	• Initial setup only involves gain tuning for every control variable (can be semi-automated) • Light compute load vs adaptive controllers • Gain-scheduling can help controller work for different state-space regions	• Gains are good for small region of state-space and system is unstable outside this region • Integrator wind-up occurs when large change in setpoint occurs, resulting in severe overshoot • Optimal control / stability not guaranteed
Model Predictive Control (MPC)	• Computes optimal control trajectory using linear approximation of system + recomputes when approximation is unlikely to be accurate	• Generating system model for our robot might be difficult given anatomical constraints • In general, this approach is sensitive to model inaccuracies and tuning of the cost function
Linear Quadratic Regulator (LQR)	• Finds optimal control input for linear systems • Handles constraints on control inputs and states	• Just like in MPC, generating system model for our use case might be challenging • Linearization of nonlinear systems leads to significant performance degradation

Broader surgical robotics issues which might require special consideration when designing our system

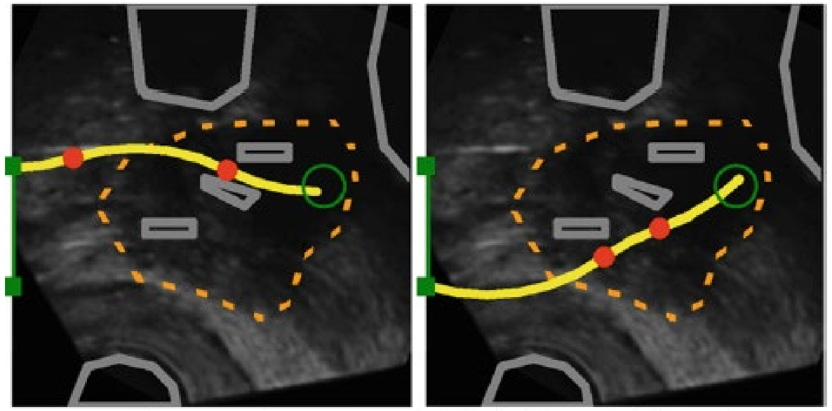
Motion canceling in medical interventions

- Model-Based Active Relative Motion Cancelation is required during coronary artery bypass graft (CABG) surgery. The tools need to track and manipulate a fast-moving target with very high precision. During free beating, individual points on the heart move as much as 7–10mm. Although the dominant mode of heart motion is on the order of 1–2 Hz
- Query whether movement in the human brain may require a similar motion canceling solution to ensure smooth and predictable mobility of our robot



Planning for deformable tissues

- Inserting needles into soft tissues causes surrounding tissues to displace and deform. Ignoring these deformations can result in substantial placement error
- Query if this is relevant for our use case where the robot may interact with soft tissues in the brain
- Perhaps formulating the problem as a Partially Observable Markov Decision Process (POMDP) might be helpful, since these can be solved to yield optimal control policies under uncertainty



Shortest path Maximizing probability of success

Source: *Surgical Robotics, Systems Applications and Visions* (Springer)

.05
Appendix

SLAM using onboard sensor data [if possible to have onboard odometry]

Pseudocode for SLAM:

I. Image acquisition: X-ray images of the environment are captured using an x-ray sensor.

II. Feature detection and extraction: Features, such as edges or corners, are detected and extracted from the x-ray images.

III. Feature matching: Features from consecutive x-ray images are matched to determine the motion of the sensor.

IV. Pose estimation: The sensor's motion is used to estimate its position and orientation within the environment.

V. Map building: The sensor's position and orientation, along with the features extracted from the x-ray images, are used to build a map of the environment.

VI. Loop closure: When the sensor revisits an area, loop closures are detected and used to refine the map and sensor's position.

Some of the challenges of visual SLAM on x-ray imagery include the low contrast and texture of x-ray images, as well as the high computational demands of processing x-ray images in real-time. To overcome these challenges, researchers have proposed various methods such as using a combination of x-ray and visible light images, preprocessing the images to enhance features, and using efficient optimization algorithms.

Alternatives to 5th degree polynomial-based trajectory generation (Cubic polynomial, Minimum-time trajectory)

7.5.1 Trajectories for Point-to-Point Motion

As described above, the problem is to find a trajectory that connects the initial and final configurations while satisfying other specified constraints at the endpoints, such as velocity and/or acceleration constraints. Without loss of generality, we will consider planning the trajectory for a single joint, since the trajectories for the remaining joints will be created independently and in exactly the same way. Thus, we will concern ourselves with the problem of determining $q(t)$, where $q(t)$ is a scalar joint variable.

We suppose that at time t_0 the joint variable satisfies

$$q(t_0) = q_0 \quad (7.9)$$

$$\dot{q}(t_0) = v_0 \quad (7.10)$$

and we wish to attain the values at t_f

$$q(t_f) = q_f \quad (7.11)$$

$$\dot{q}(t_f) = v_f \quad (7.12)$$

Thus, we consider a cubic trajectory of the form

$$q(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$$

Minimum-Time Trajectories

An important variation of the LSPB trajectory is obtained by leaving the final time t_f unspecified and seeking the "fastest" trajectory between q_0 and q_f with a given constant acceleration α, that is, the trajectory with the final time t_f a minimum. This is sometimes called a **bang-bang** trajectory since the optimal solution is achieved with the acceleration at its maximum value $+\alpha$ until an appropriate **switching time** t_s at which time it abruptly switches to its minimum value $-\alpha$ (maximum deceleration) from t_s to t_f.

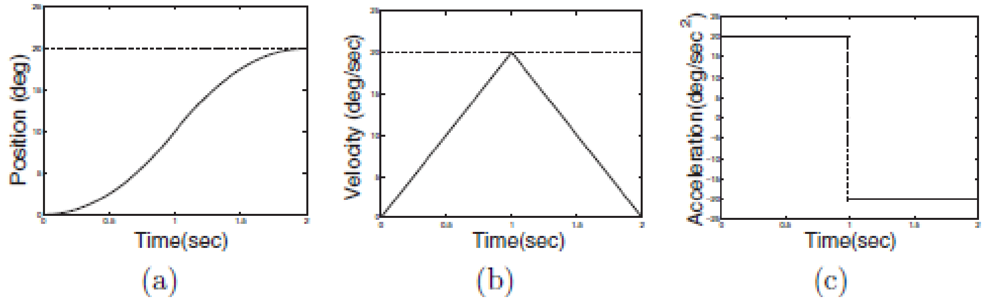
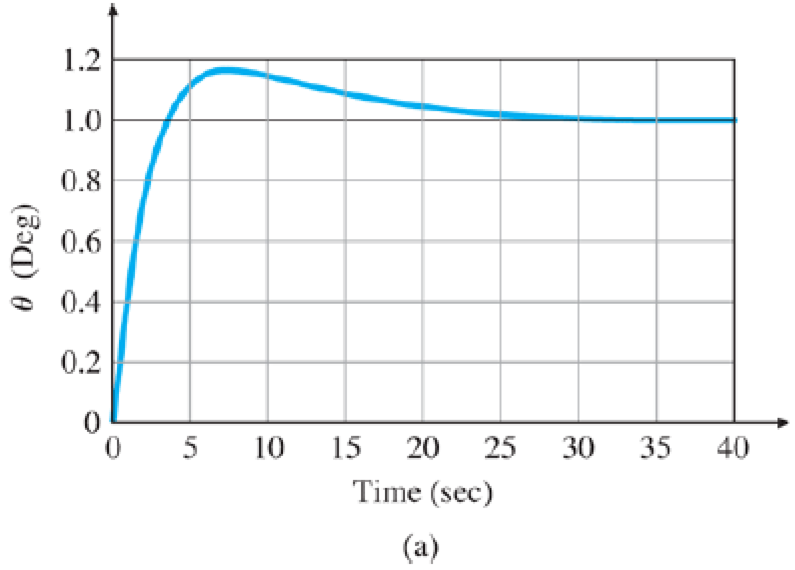
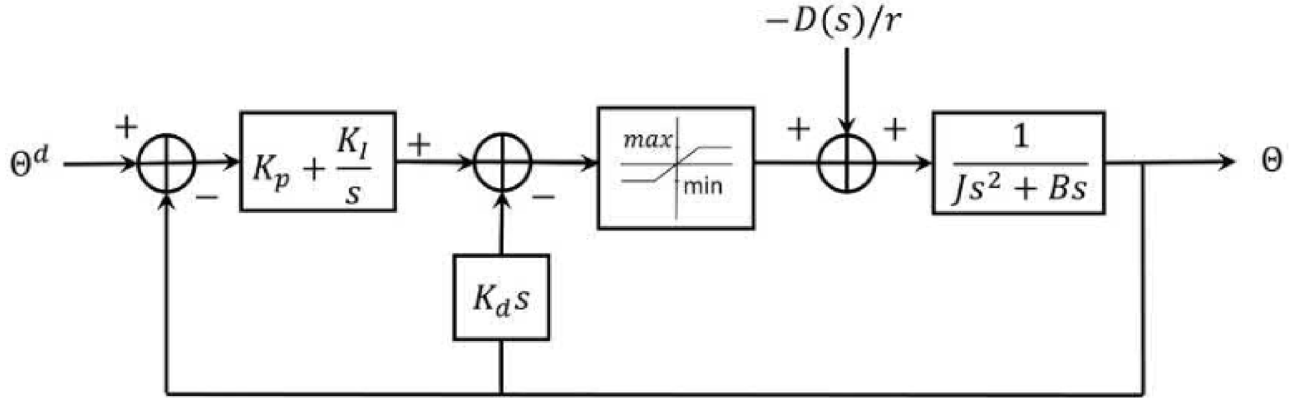


Figure 7.22: (a) Minimum-time trajectory. (b) Velocity profile for minimum-time trajectory. (c) Acceleration profile for minimum-time trajectory.

Deeper Dive into PID control and potential stability considerations

Figure 6.69 Transient response for PID example: (a) unit step command response; (b) step torque disturbance response





A second consideration affecting high-frequency gains is that many systems have high-frequency dynamic phenomena, such as mechanical resonances, that could have an impact on the stability of a system. In very-high-performance designs, these high-frequency dynamics are included in the plant model, and a compensator is designed with a specific knowledge of those dynamics. A standard approach to designing for unknown high-frequency dynamics is to keep the high-frequency gain low, just as we did for sensor-noise reduction. The reason for this can be seen from the gain–

Source: Feedback Control of Dynamic Systems (Franklin, Powell, Emami-Naeni)