

Engineering Physics 353 Final Report

Team 14 - “Team 4”

Nathan Van Rumpt, Eric Souder

April 2023

1 Background

The final project for Engineering Physics 353 is the creation of a ROS package to control a simulated robot as it attempts to drive around a streetscape - composed of an inner and outer loop - and record license plate numbers off of parked cars.



Figure 1: Competition track layout

The robot must be capable of navigating entirely autonomously while avoiding collisions with pedestrians and other vehicles in the environment. Points were awarded for correctly reporting a license plate over a ROS topic (6 points for plates on the outer ring, and 8 points for plates on the inner ring) and for completing a full lap, and deducted for hitting pedestrians, other cars, or driving off the road. Our team attempted to create a robot that would be primarily capable of very reliably navigating the track and collecting the maximum amount of points. As a secondary objective, we aimed to have the robot complete as quickly as possible, as tie scores would be broken by the robot with the fastest time.

2 Overall Architecture

I'll do this section.

3 Navigation

Robot navigation was accomplished with an almost entirely classical control system, using minimal neural network based approaches and instead using manually tuned computer vision and classical control techniques.

The navigation controller was composed of two ROS nodes: The 'State Machine' node was composed of a finite state machine and the 'Navigate' node contained a set of navigation algorithms for different scenarios, selecting an appropriate algorithm based on the state reported by the state machine. States were communicated internally between the two nodes via a ROS topic.

3.1 State Machine

The finite state machine responsible for controlling the robot's navigation state was composed primarily of a number of state classes, representing a unique navigational state and a state machine class responsible for orchestrating the transitions between them. Each state class, such as the 'Pave Navigate' or 'Junction Wait' states, is a subclass of an abstract parent class which defines an 'evaluate transition' interface responsible for determining the next state to enter and either returns itself to remain in the current state or generates a new state object to transition into. The state machine class calls this method each frame to determine which state the robot should be in, and then broadcasts that state to the navigation node via a ROS topic. The overall state machine diagram is shown in Figure 2.

Three main types of navigational states existed: active navigation states, stopped states, and preprogrammed states. In preprogrammed states, where the robot executes a constant movement, the state machine transitions based off of a simple timer to move to the next state once the movement had been in progress for the preprogrammed length of time. In

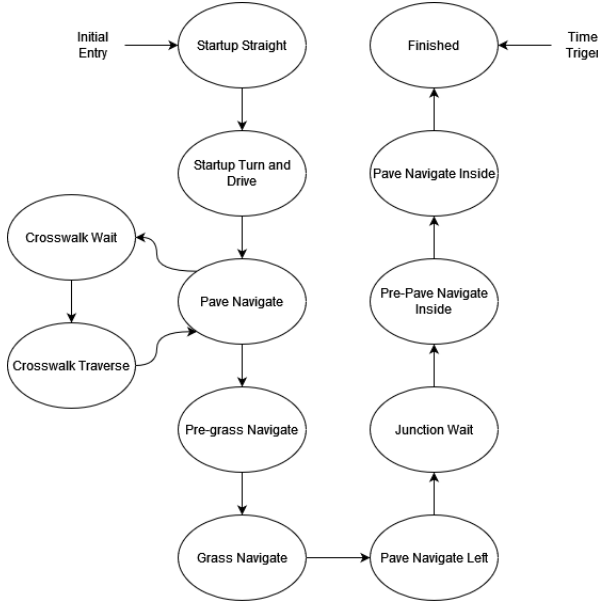


Figure 2: Diagram of the robot state machine.

active navigation states, where the navigation algorithm is interpreting each frame on the fly to determine a which movement to command, or in stopped states, where the robot is waiting for an event to occur before proceeding, the state machine interprets each frame and makes a transition decision based on it.

3.1.1 Pavement and Grass Detection

In order to transition from the pavement navigation states to the grass navigation state and vice versa, the state machine must be able to determine if the robot is driving on grass or on pavement. This detection is accomplished with a small neural network, implemented within a ‘SurfaceDetector’ class. Training data was 400 images captured from the robot camera, evenly split between grass and pavement surfaces. Data was acquired by manually driving the vehicle around with a simple ros node activated to save each camera frame. Because of the goal of simplicity for the neural network, all images were preprocessed by significant compression, images were resized to 0.025

Layer (type)	Output Shape	Num. Params.
Flatten	(None, 1632)	0
Dense	(None, 1632)	2665056
Dense	(None, 1)	1633

Figure 3: Summary of surface detection neural network

times their original width and 0.05 times their original height, this resulted in a 36 by 32 image, which was further cropped to remove the top 18 rows of pixels, focusing only on the ground and removing the area above the horizon.

The simple model was composed of only three layers: An input flatten layer, a hidden dense layer, and a output dense layer composed of a single neuron, with a total of 2,666,689 trainable parameters, as shown in Figure 3. The model was trained with a learning rate of 1×10^{-4} and a validation split of 0.2 over 10 epochs. Loss and accuracy curves by epoch are shown in Figure 4.

Because of the small size and binary output of the network, little work was deemed necessary to validate it’s performance and there was no need for extensive error analysis. A robot was manually driven around the track and the output of the network’s predictions was monitored to ensure it was sensible before it was integrated with the state machine to trigger transitions between pavement and grass navigation states.

3.1.2 Crosswalk and Pedestrian Detection

When navigating over the first section of paved track, two crosswalks with pedestrians must be safely crossed without hitting any pedestrians. The robot must detect the presence of a crosswalk and transition to a ‘Crosswalk Wait’ state in order to avoid hitting a pedestrian. Detection of both crosswalks and pedestrians is performed by methods within the ‘PedestrianDetector’ class. First, crosswalks were detected using simple HSV colour thresholding. [figure here]. A crosswalk was considered ‘detected’ when 1500 total pixels matching the range of colours associated with the red crosswalk stop line were detected

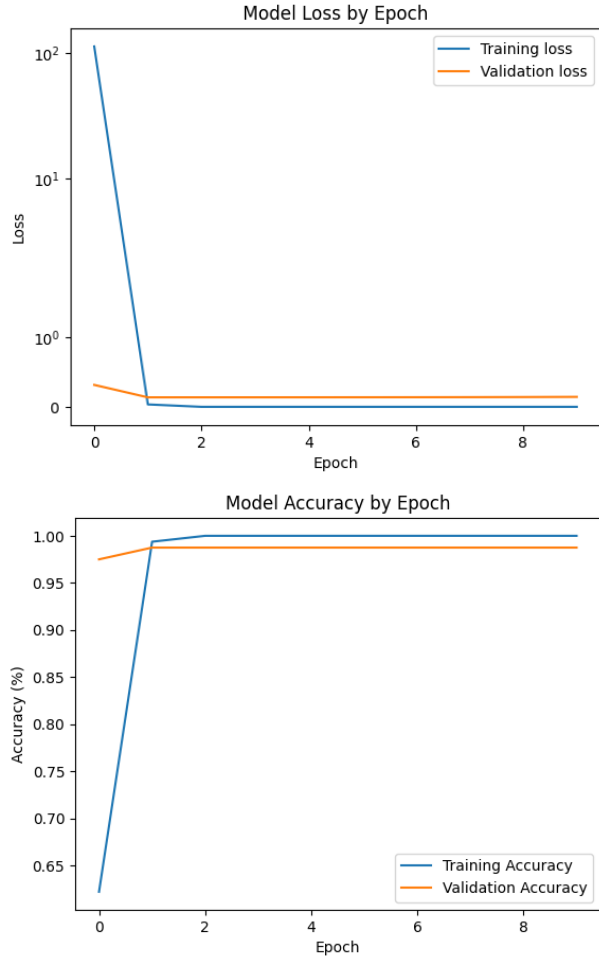


Figure 4: Loss and accuracy of grass and pavement detection network by training epoch.

in the bottom 100 rows of the image. Although 1500 pixels is less than 2 full horizontal rows - much less than the size of the detection area - the need for a large area was the result of the relative high velocity which meant that the robot may only receive one or two frames to detect the crosswalk, so the larger area was chosen as a compromise between guaranteeing detection of the stop line and constantly stopping at around the same location. It also provides detection capability if the line is approached at an angle. This hedges against a possible 10 point loss from hitting a pedestrian or other parts of the navigation system fail.

Once the crosswalk was detected and the robot was in the ‘Crosswalk Wait’ state, we must detect pedestrians to determine when it is safe to cross. Pedestrians were detected using HSV thresholding to detect their blue jeans, similarly to the crosswalk detection. [figure here]. However, since the location (not just boolean presence) of the pedestrians was needed, erosion and dilation were used to merge the thresholded areas together into fewer, larger ‘blobs’. The algorithm then finds the contours of each ‘blob’ and selects the largest as the most likely choice to be the pedestrian. Each frame, A bounding box is drawn around it, and its position is found.

As a result of the repeatability of our pavement navigation algorithms, the edges of the crosswalk were able to be hard-coded into the pedestrian detector. The state machine tracks the position of the detector, and waits for it to be found in the crosswalk and then to exit the crosswalk. At this point, we can guarantee the maximum amount of time that the pedestrian will not be in the crosswalk, and the state machine is then able to transition to a ‘Crosswalk Traverse’ sprint across the crosswalk before returning to the normal pavement navigation state.

3.1.3 Junction and Truck Detection

In order to transition to the inner loop of track, the robot must turn through the junction to the inner loop, stop within the junction, wait for the truck to pass safely, and then proceed into the junction. This is accomplished through state transitions using methods with the ‘JunctionDetector’ class.

When the vehicle exits the grass segment, the state machine switches to the ‘Pave Navigate Left’ mode, tracking the left side of the track rather than the right side as it does for other pavement navigation states. After three seconds of this, the state machine begins attempting to go through a procedure to see if has reached the junction with the inner track by monitoring the width of the pavement in front of it. The pavement width is determined by HSV thresholding to the colour of the pavement and counting the distance between the right- and leftmost valid pixels at a horizontal line 150 pixels up from the bottom of the frame. The algorithm used to evaluate if we are in the junction is shown in figure 5, which works by exploiting the characteristic pattern of the pavement appearing to widen, narrow, and then widen again as we turn into the junction [figure here!]. The state machine then transitions to a stopped ‘Junction Wait’ state.

In the ‘Junction wait’ state, the state machine waits to detect the truck in the appropriate position to safely start the inner loop navigation. Truck detection is based on a very similar methodology as pedestrian pant detection, using HSV colour thresholding followed by erosion, dilation, and contour finding. The state machine will proceed to the next state (moving into the inner loop) once the truck reaches a position in the frame of $x < 400$ pixels and $y < 700$ pixels, a zone that was experimentally determined safe. These values were able to be hard-coded because the position of the robot when waiting at the junction remained extremely consistent throughout runs.

3.2 Navigation Algorithms

During the active navigation states, the navigation node processes each frame from the camera and uses it to determine a steering angle to publish based on an algorithm tailored for the state it is in. The two main types of active navigation nodes, for paved surfaces and grass surfaces employed relatively similar processes to determine steering angles by attempting to find the edge of the track and hold it at a constant bearing using a proportional controller to determine angular control and holding a constant speed. The

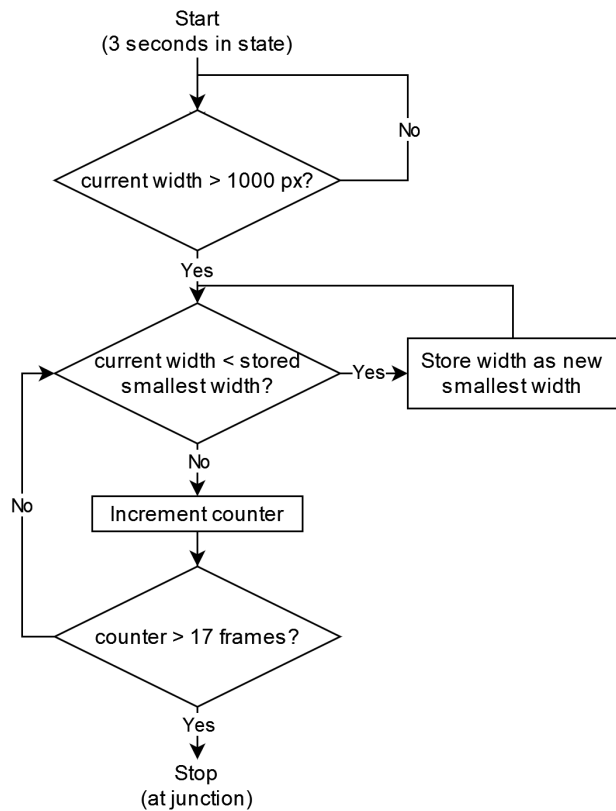


Figure 5: Flowchart showing junction detection algorithm based on pavement width.

main difference between the two was the method used to find the edge of the track.

3.2.1 Pavement Navigation

During pavement navigation, the edge of the track is found through an image processing pipeline to output a binary image with the edge of the track as white pixels and the rest of the image as black pixels. This binary image is then processed to find the center of the track, which is used to determine the steering angle. The image processing pipeline is shown in Figure 6. The HSV threshold, greyscale, and binary threshold steps convert our frame into a single binary image that we are able to run edge detection and line finding on, while the dilation and erosion steps remove ‘holes’ in the pavement caused by the environment’s 1 m grid lines. Sobel edge detection is used to find the edges of the pavement, and then we crop the image to only include the pavement area of interest, ignoring the sky and trees.

A Hough transform is used to find the lines that best fit the edges, and these lines are then drawn onto a new, blank image which we can then convert to binary and use a simple weighted average to find the centroid of the edge, restricting ourselves to a zone between a certain number of pixels from the bottom of the image - to avoid looking too far into the future or focusing too much on lines immediately close to robot - and to either the left or right side of the image depending on if we are navigating based on the left or right side of the track.

We navigate based on the right side in almost all pavement areas, as it does not have any junctions interrupting the lines, but on the left side of the track in the ‘Pave Navigate Left’ state, as we want to follow the left line into the junction in order to transition to the inner loop.

3.2.2 Grass Navigation

Grass navigation was accomplished with a similar, but more simple pipeline than was used in the pavement navigation, shown in Figure 7. However, instead of attempting to create lines that fit the edges of the track, we instead use thresholding and other

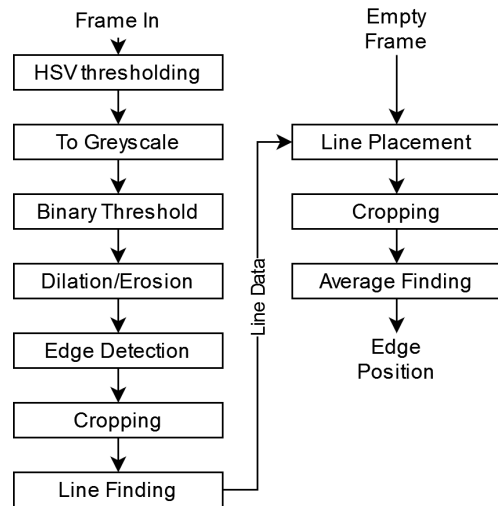
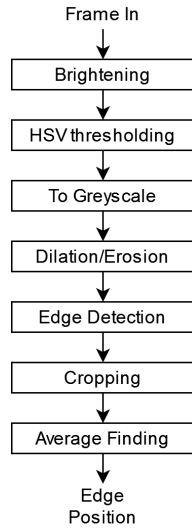


Figure 6: Pavement edge detection image processing pipeline.

classical techniques to simply isolate the pixels that match the colour of the lines, and then find the average position of the right line. A key step in being able to do this was to pass all frames through a brightening step that brings all images up to a standard contrast, as frames from the camera vary in contrast over different areas of the grassy section. Sobel edge detection is also used as a final step to reduce the impact of any large blocks of pixels in the final image other than the line, as they will only contribute an outline’s worth of pixels, rather than the entire block.

3.3 Performance

Through testing and iteration, we found our navigation system to be reliable once tuned, but not necessarily robust to modifications. For instance, we were able to run hundreds of laps on the track without any failures, but would not be able to change the speed without significant trial-and-error based retuning of multiple components and values. Additionally, we found that the navigation system was sensitive to the battery status of the computer the simulation was running on, and would routinely fail if the com-



specific plates used in the competition. The pipeline is shown in Figure 8.

Figure 7: Grass edge detection image processing pipeline

puter was not charged. We also encountered reliability challenges as we increased the speed of the robot. On the pavement sections, we were not able to reliably exceed 0.5 m/s, and on the grass sections, we were not able to reliably exceed 0.3 m/s without the robot losing the line and failing. We believe this to be a result of amount of track that the robot would cover between frames - the design of this navigation system, since it relies only on the current frame, is unable to use data about its previous navigation inputs or outputs to determine current outputs, and thus is unable to compensate for the robot's momentum. While these speed limits bounded the performance and lap time of the robot, we still found it extremely reliable within the limits of its maximum speed.

4 License Plate Detection

Plate detection was accomplished using a series of handcrafted filters to extract the individual plate character bounding boxes. This was chosen since it gave a high degree of control over the output, and allowed us to easily tune the system to work on the

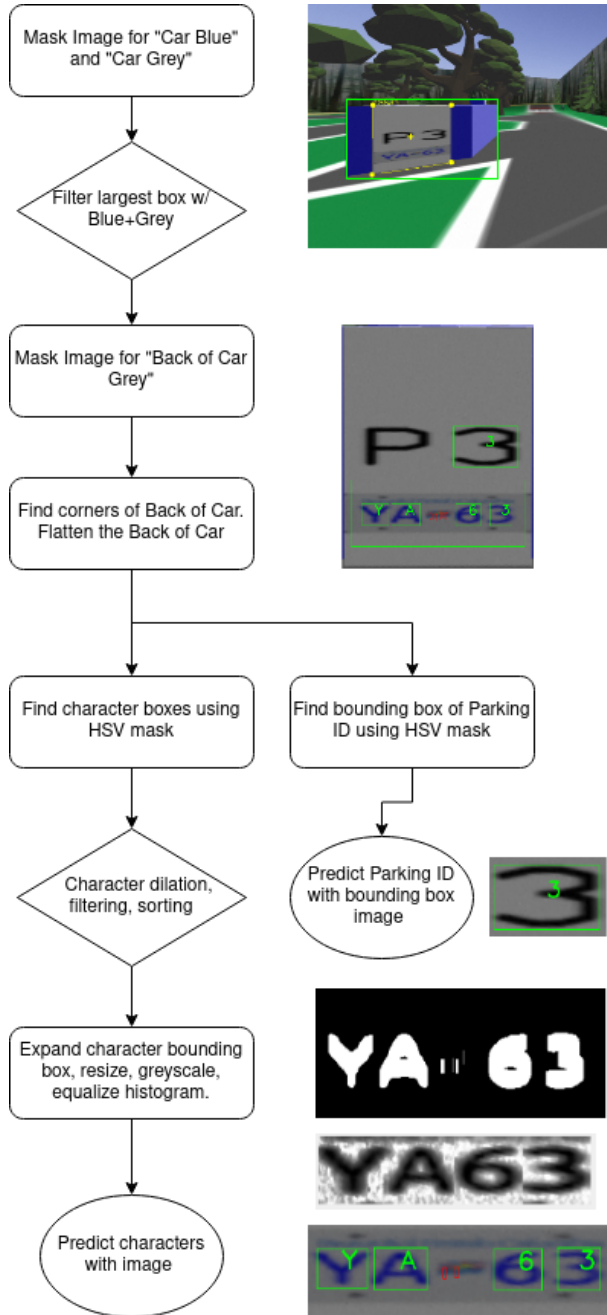


Figure 8: Image Processing Algorithm

4.1 Plate Identification Algorithm

The first step in the pipeline is to mask the image with a blue and a grey HSV filter to isolate the areas that likely contain cars. These two masks are joined and then contoured. Each contour is checked to see if it contains both blue and grey, otherwise it is discarded (this ensures that only actual cars are investigated further, and not sections of the sky or pavement).

The largest contour in the image moves onto the next stage of analysis. The contour's bounding box image is masked with a more specific grey HSV filter to get the top and bottom sections of the back of the car. These contours are combined and then fitted to a low quality polygon. The longest 4 lines of the polygon are selected and the intersections of each of the lines found. The four points that are closest to the centroid of the original contour are selected as the corners of the back of the car. Finally these corners are used to flatten the back of the car and resize it to a standard dimension.

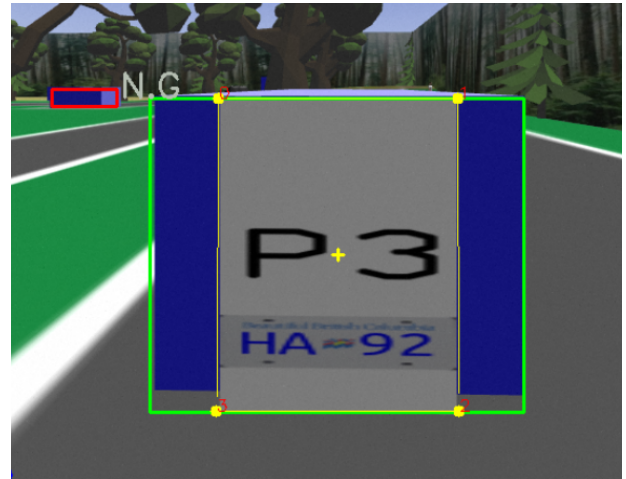


Figure 9: Annotated Back Car. Yellow lines from polygon, yellow dots are the corners, cross is centroid of contour.

The parking ID is extracted using a simple HSV filter, and then greyscaled before passing into the parking ID neural network.

The flattened image is then masked with a blue HSV filter to isolate the license plate characters. This mask undergoes a slight dilation along x and y, then an aggressive dilation only along y. This was chosen to provide a balance of getting the full character without merging large characters when they are next to each other.



Figure 10: Example mask showing non-merging of wide XX combination.

Finally this mask is contoured, bounding boxed, and then the bounding box expanded. The characters are then filtered based on size and aspect ratio. The remaining contours are then sorted by x position and the characters are extracted from the plate. The characters are then resized to a standard dimension and passed to the neural network for classification.



Figure 11: Identified XX00 plate.

4.2 Neural Networks

The plate identification algorithm uses a total of 3 neural networks. One to identify the parking ID, one for the plate letters, and one for the plate numbers. Separate networks were chosen for each of these tasks since it simplified the training process and decreased computation time.

4.2.1 Parking ID model

The parking ID model was trained on a dataset of 400 images of parking IDs extracted from the simulation environment and manually labelled.

Layer (type)	Output Shape	Num. Params.
Conv2D	(None, 48, 48, 3)	30
MaxPool	(None, 24, 24, 3)	0
Conv2D	(None, 22, 22, 64)	1792
MaxPool	(None, 11, 11, 64)	0
Flatten	(None, 7744)	0
Dropout	(None, 7744)	0
Dense	(None, 128)	991360
Dense	(None, 8)	1032

Figure 12: Summary of parking ID neural network

4.2.2 Plate Letters

The plate letters model was trained on a dataset of 17000 images of plate letters extracted from the simulation environment and automatically labelled using the plate list in the environment.

Layer (type)	Output Shape	Num. Params.
Conv2D	(None, 48, 48, 48)	480
MaxPool	(None, 24, 24, 48)	0
Conv2D	(None, 22, 22, 96)	41568
MaxPool	(None, 11, 11, 96)	0
Flatten	(None, 11616)	0
Dropout	(None, 11616)	0
Dense	(None, 512)	5945344
Dense	(None, 26)	13338

Figure 13: Summary of Plate Letters Model

4.2.3 Plate Numbers

The plate numbers model was trained on a dataset of 17000 images of plate numbers extracted from the simulation environment and automatically labelled using the plate list in the environment.

Layer (type)	Output Shape	Num. Params.
Conv2D	(None, 48, 48, 16)	160
MaxPool	(None, 24, 24, 16)	0
Conv2D	(None, 22, 22, 32)	4640
MaxPool	(None, 11, 11, 32)	0
Flatten	(None, 3872)	0
Dropout	(None, 3872)	0
Dense	(None, 256)	991488
Dense	(None, 10)	2570

Figure 14: Summary of Plate Numbers Model

5 User Interface

6 Conclusion