1 Appendix

1.1 Network Structure

Table 1: The proposed contact estimation network architecture. Conv1D stands for 1-dimensional convolution layer, Pool1D means 1-dimensional max pooling, and FC is the fully connected layer.

	Operator	Stride	Filters	Size	Output Shape
	Input				$N \times 54 \times w$
Block 1	Conv1D	1	64	3	$N \times 64 \times w$
	Conv1D	1	64	3	$N \times 64 \times w$
	Pool1D	2	64	2	$N \times 64 \times w/2$
Block 2	Conv1D	1	128	3	$N \times 128 \times w/2$
	Conv1D	1	128	3	$N \times 128 \times w/2$
	Pool1D	2	128	2	$N \times 128 \times w/4$
FC	FC1	-	-	-	2048
	FC2	-	-	-	512
	FC3	-	-	-	16

1.2 Algorithm for Groundturth Contact Generation

The algorithm used to generate the ground truth labels is listed in Algorithm 1. The algorithm connects the local minimum between peaks after applying a low-pass filter to the foot height signal. An example of the ground truth contacts overlapping with foot height is shown in Figure 1. From the plot, we can see the foot bounces up upon touch down, and the algorithm is able to remove it from the ground truth contact list.

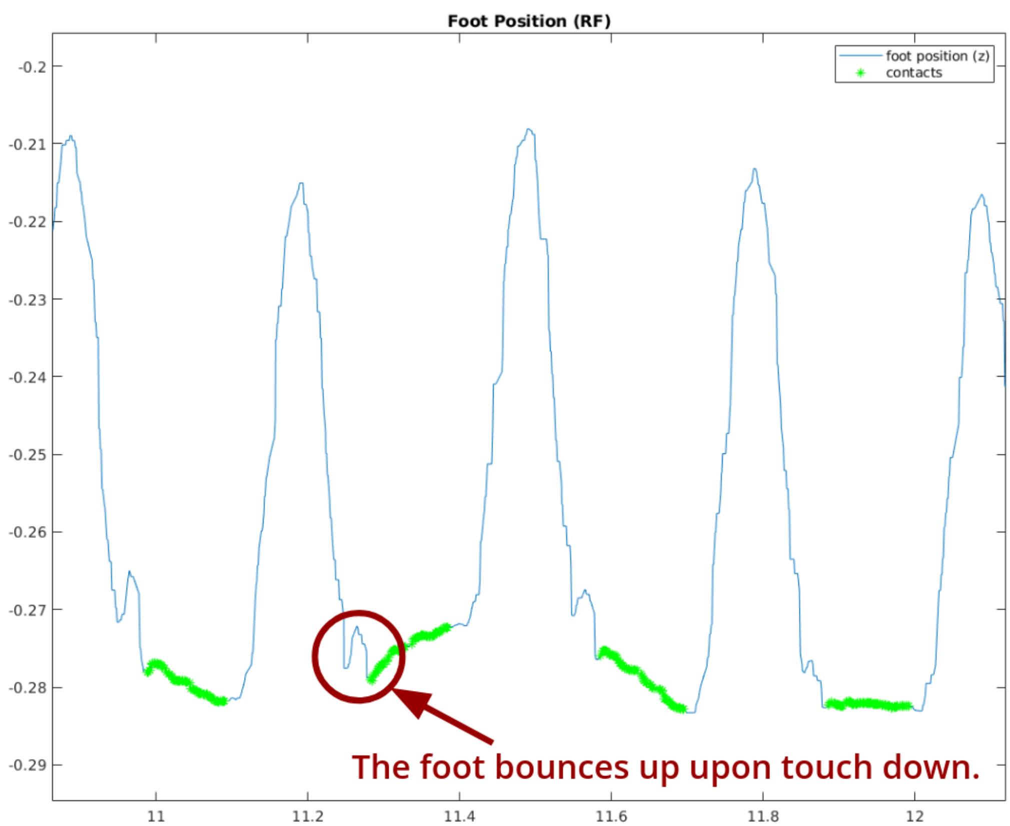

Figure 1: An example of labeled ground truth contacts (in green) overlapped with foot height measurements (in blue) in the hip frame. As seen in the figure, the initial "jump" in the signal is removed from contact ground truth labels.

Algorithm 1 Ground Truth Label Generation

```
contacts = False(num_data, num_legs)
for all legs do

    # Set different half power frequency based on gait
    if gait == trot then
        half_power_freq = 0.04
    else if gait == (pronking or gallop) then
        half_power_freq = 0.08
    end if

    foot_height = low_pass_filter(foot_height, half_power_freq)

    # Extract local max and local min
    local_max = is_local_max(foot_height)
    local_min = is_local_min(foot_height)

    # Extract indices from local max and min
    max_idx = find_idx(local_max)
    min_idx = find_idx(local_min)

    i = 0
    j = 0
    while i < num_min & j < num_max do
        contact_start = min_idx[i]
        next_peak = max_idx[j]

        # Connect all the local minimum before next peak
        count = 0
        while i < num_min & min_idx[i] < next_peak do
            contact_end = min_idx[i]
            i = i + 1
            count = count + 1
        end while

        # If only one local minimum is found between two peaks,
        we set a conservative amount of data beforehand as contact.
        if count == 1 then
            contact_start = contact_end - 30;
            if contact_start<1 then contact_start=1;
            end if
        end if
        contacts[contact_start:contact_end, l] = True
        j = j + 1
    end while
end for
```

Table 2: Accuracy and precision of the proposed network on the test sets. 16 class denotes the accuracy and precision in terms of 16 contact states S. Leg RF to Leg LH list the accuracy and precision of individual leg by comparing the correctness of each leg individually.

Terrain	% Accuracy						% False Positive Rate	% False Negative Rate
	16 class	Leg RF	Leg LF	Leg RH	Leg LH	Leg Avg	Leg Avg	Leg Avg
overall	93.88	97.76	97.66	97.86	98.00	97.82	1.82	2.88
air trotting	100.00	100.00	100.00	100.00	100.00	100.00	0.00	N/A
air pronking	100.00	100.00	100.00	100.00	100.00	100.00	0.00	N/A
asphalt road	94.82	98.01	98.49	98.41	98.59	98.37	1.41	2.02
concrete	93.09	97.30	97.32	97.62	97.57	97.45	2.07	3.25
forest	91.26	96.88	97.47	97.51	97.38	97.31	2.22	3.58
grass	93.60	98.13	97.57	97.73	97.80	97.81	2.33	1.93
middle pebble	93.94	98.08	97.62	97.73	98.22	97.91	1.88	2,47
small pebble	92.62	97.82	97.11	97.67	98.00	97.65	2.55	1.95
rock road	95.45	98.83	98.39	98.46	98.90	98.65	1.12	1.79
sidewalk	95.16	98.20	98.02	98.17	98.41	98.20	1.51	2.33

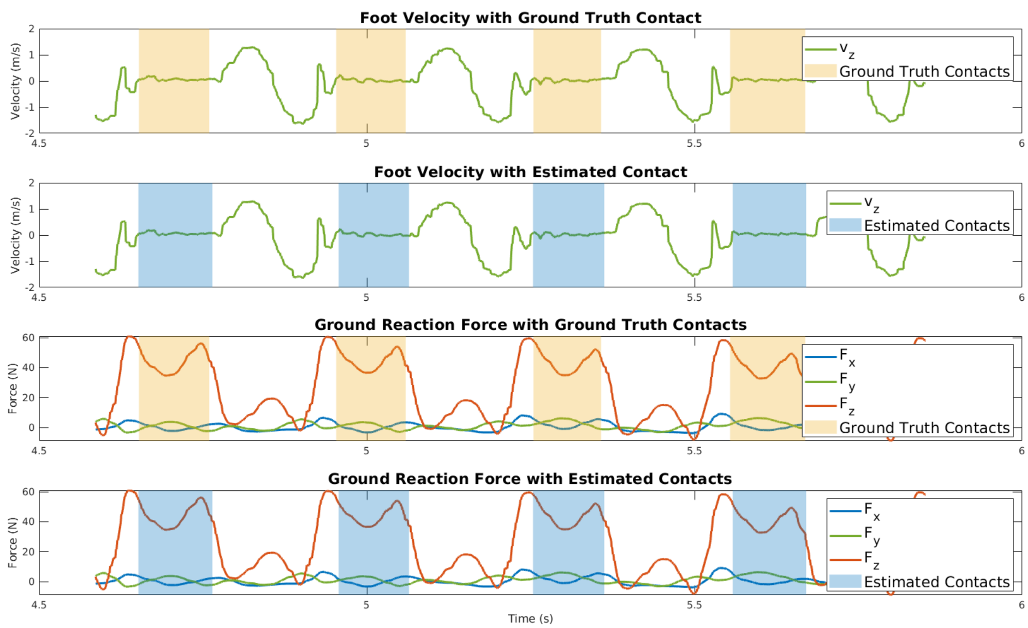

Figure 2: Estimated ground reaction force (After passing through a low-pass filter.) and foot velocity overlapped with estimated contacts and ground truth contacts of one leg in the forest data set. The plots show that the estimated contact phase is consistent with the ground truth contacts. Upon initial contact, the foot bounces up because of the high impact. This plot shows the network can exclude the bouncing phase and correctly estimate contacts where only minimal velocity between the foot and robot body are present.

1.3 Contact Estimation Results

Table 2 lists the accuracy of the trained network on different terrains on the test sets. We evaluate the performance of the network in two ways. First, 16 class denotes the accuracy in terms of 16 contact states S we defined earlier. This case is a harsher way to evaluate the network since it requires all four legs to be correct simultaneously. Second, Leg RF to Leg LH list the accuracy of the individual leg by comparing the estimated contact state of each leg with ground truth contacts individually. We can see that, on average, the trained network achieves 93.88% of accuracy in terms of 16 contact states and 97.82% in each leg individually. Across different terrains, the network keeps an accuracy above 96% for each leg. Furthermore, the air trotting and pronking data sets show that when the robot is walking with a similar kinematic pattern in the air, the network can tell there exist no contact event.

1.4 Contact Estimation with Foot Velocity and Ground Reaction Force

Figure 2 shows the estimated contacts overlapped with the estimated ground reaction force and the foot velocity, as well as the ground truth contacts in the forest test set. The yellow sections indicate the ground truth contact phase, and the blue sections are the estimated contact phase. From the plot, we can see that the estimated contacts are consistent with the ground truth contacts. The bottom

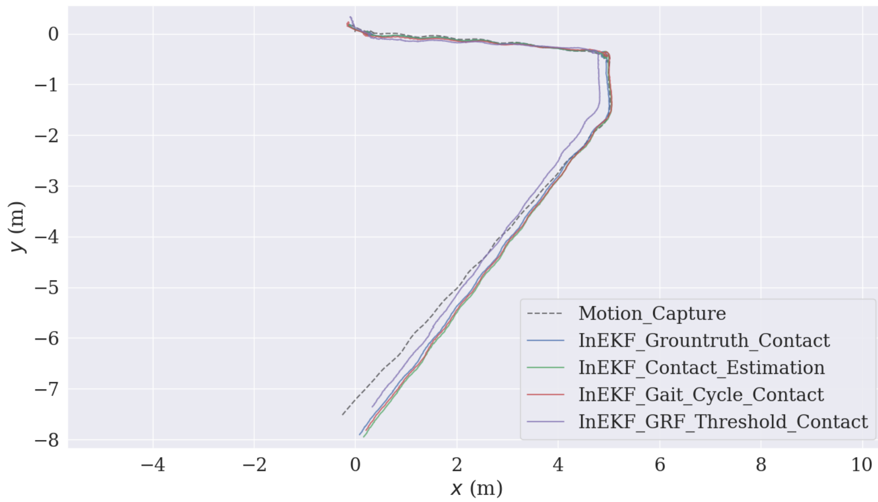

Figure 3: The bird-eye view of the trajectories on the grass test sequence. We can see that the resulting trajectory from GRF thresholding gives a larger drifts in the X-Y plane.

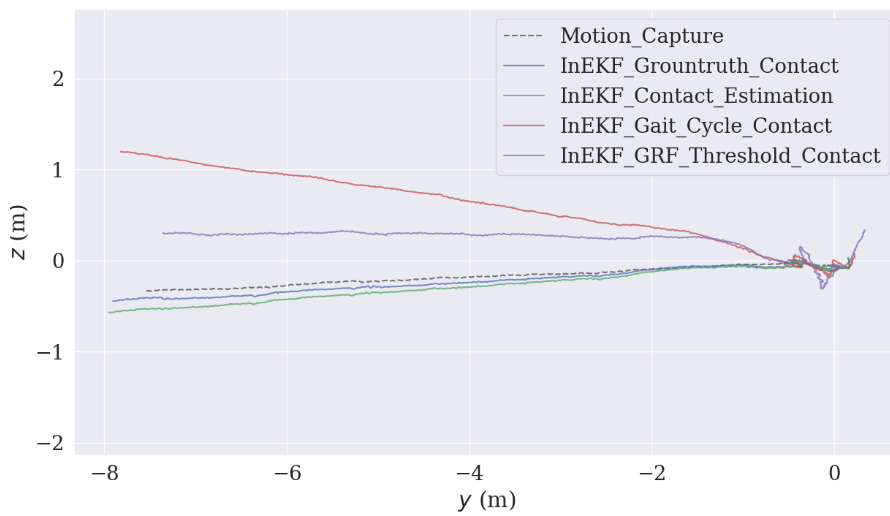

Figure 4: A side-view of the trajectories on the grass test sequence. The InEKF trajectory with gait cycle contact detection methods have a significant drift in the height (Z) axis.

two plots show the estimated ground reaction force overlapped with contacts, and the top two plots demonstrate the foot velocity versus contacts. It is worth noticing that the foot velocity here is represented in the robot hip frame because we are unable to measure the foot velocity in the world frame directly. The plots show that the estimated contact periods match the periods where only minimal vertical velocity is present between the foot and the robot body. Note that when the foot is first in contact with the ground, the foot bounces up and creates a vertical jump in the foot velocity curve. We conjecture this is because Mini Cheetah's feet are made of rubber balls and are bouncy given the high impact at the initial contact. We wish to have a reliable contact estimation for the state estimator. Thus, even though the estimated ground reaction force is high, the initial contact phase is excluded from the ground truth contacts to avoid violating the no-slip condition. As one can see from the plot, the network can exclude the initial bouncing phase and estimate contact events where the foot velocity is closer to 0.

1.5 Comparison Against Other Contact Detection Methods

Figure 3, 4, and 5 show the trajectories of the InEKF with different contact estimation methods in the grass test sequence. We can see that the resulting trajectory from GRF thresholding gives larger

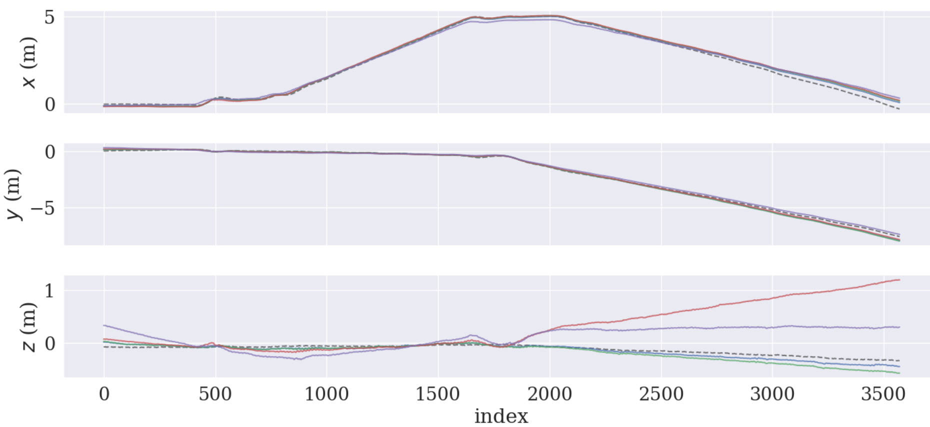

Figure 5: The section view of the trajectories. We can see that both gait cycle and GRF thresholding creates significant height (Z) drift in state estimation.

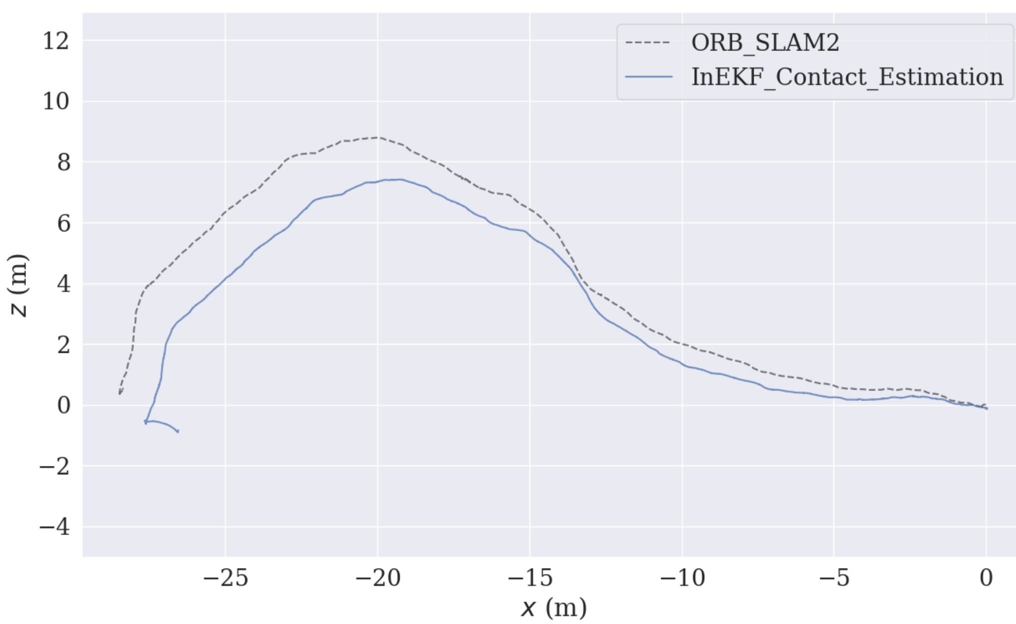

Figure 6: Trajectory of the forest test set. ORB SLAM2 fails close to the end of the sequence. This figure confirms the complementary role of the proposed proprioceptive state estimator.

drifts in the X-Y plane. The result with gait cycle method, although it shows a similar path to the proposed method in the X-Y plane, has a significant drift in the height (Z) axis, as shown in Figure 4 and 5.

1.6 Extra Results on Contact-Aided Invariant Extended Kalman Filtering

We present three extra trajectories on forest, small and middle pebbles, and sidewalk and asphalt road, as shown in Figure 6, 7, and 8, respectively. These three test sets were not included in the training process. The forest test set consists of forest terrain described in the paper. In this experiment, ORB SLAM2 fails close to the end of the sequence due to a sudden sharp turn. In the pebbles test set, the robot walks in a circular shape starting from middle size pebbles to small size pebbles. As for the sidewalk and asphalt test set, the robot starts from the sidewalk, turns, and walks into a sloping asphalt road. The total length of the robot path in this sequence is around $100\,\mathrm{m}$.

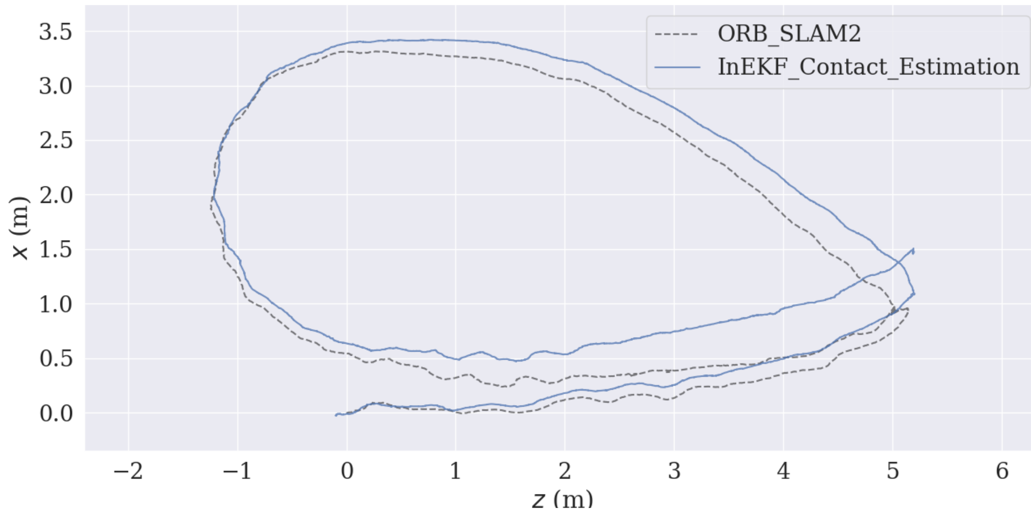

Figure 7: Trajectory of a test set with both small and middle pebbles. The robot walks in a circular shape starting from middle size pebbles to small size pebbles.

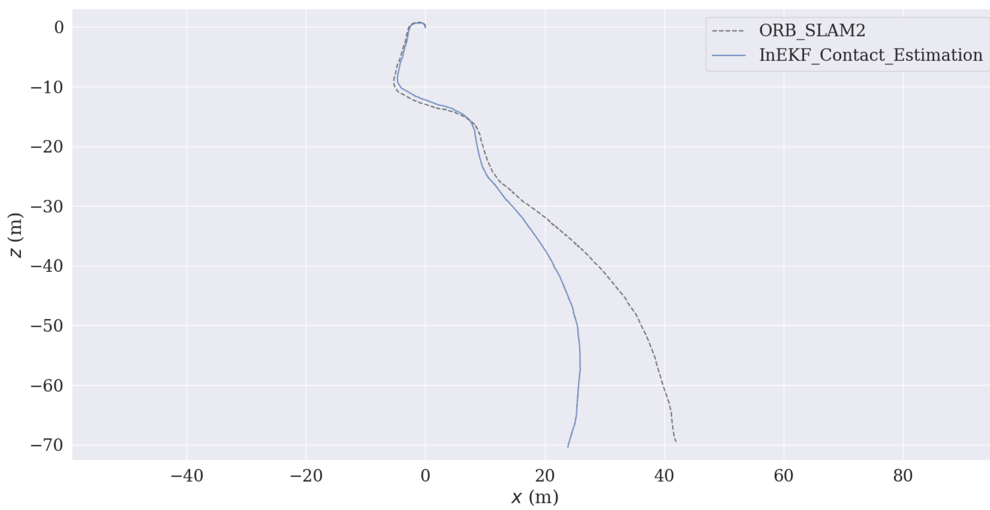

Figure 8: Trajectory of a test set with both sidewalk and asphalt road. the robot starts from the sidewalk, turns and walks into a sloping asphalt road. The total length of this path is around $100\ \mathrm{m}$.

The three plots show that with the proposed contact estimator, the InEKF can successfully estimate the pose across 5 different terrains using only IMU and encoder measurements. It is worth noticing that the InEKF is an odometry-only method, that is, accumulating estimated poses without post-optimization or loop-closure detection. Thus, the error at each estimation is accumulated in the trajectory. The integration of this InEKF odometry into a real-time SLAM system is an interesting future work direction.