Can a Robot Navigate Using Vision Alone Forever?

Tim Barfoot, Professor, <tim.barfoot@utoronto.ca>
Visit to the University of Oxford and Oxbotica
Feb 23-24, 2017
Figure 9. The hardware and test site used for the UTIAS indoor rover test facility data set, as well as the distribution of scans obtained in the terrain. (a) A panoramic image of the UTIAS indoor rover test facility, with the rover used to gather the data identified by the arrow. The terrain consists of gravel spread in a 40-m-diameter circular workspace, emulating scaled planetary hills and ridges. The reflective signs used for ground truth are highlighted by the arrows. (b) The Clearpath Husky A100 rover used to gather the data at UTIAS. The payloads consist of a laser range finder mounted on a pan–tilt unit to provide the 3D laser scans, an inclinometer for pitch and roll correction, and a stereo camera for visual odometry. (c) A plot depicting an overhead view of the reference map of the dome and the rover scan locations. To put this plot into the perspective of (a), the panoramic image was taken looking in from the bottom of the plot.

9.2. Results

In this section, we present the results of the mapping framework when applied to the dataset. The 50 scans were utilized for validation in two ways. In the first section, the robustness of the framework is demonstrated by utilizing subsets of the data to produce a large set of trials. With this quantitative statistical analysis, a more qualitative sense of the algorithm is then provided by utilizing all of the scans to produce a single accurate map. This is accompanied by discussions on the limitations of the framework, as well as rendered scenes for visual validation.
Current Research Snapshot

- visual teach and repeat (VT&R)
- TRex: tethered robot explorer
- continuous-time trajectory estimation

---

**ASRL**
Autonomous Space Robotics Laboratory

- state estimation
- localization and mapping
- cameras and lasers
- planning and control
- fielded systems

---

**fast new planning algorithm: BIT***
(available through OMPL; Jon Gammell)

---

**draft book on state estimation**
(free online)
Abstract — In this paper, we present an infrastructure-free mapping and localization framework for rail vehicles using only a lidar sensor. Our method is designed to handle the pathological environment found in modern underground tunnels: narrow, parallel, and relatively smooth concrete walls with very little infrastructure to break up the empty spaces in the tunnel. By using an RQE-based, point-cloud alignment approach, we are able to implement a sliding-window algorithm, used for both mapping and localization.

We demonstrate the proposed method with datasets gathered on a subway train travelling at high speeds (up to 70 km/h) in an underground tunnel for a total of 20 km across 6 runs. Our method is capable of mapping the tunnel with less than 0.6% error over the total length of the generated map. It is capable of continuously localizing, relative to the generated map, to within 10 cm in stations and at crossovers, and 1.8 m in pathological sections of tunnel. This method improves rail-based localization in a tunnel, which can be used to increase capacity on existing railways and for automated trains.

Keywords — localization; mapping; lidar; train;

I. INTRODUCTION

Knowing the location of rail vehicles, such as trains, trams, and subways, is critical to rail-system management because trains are confined to travel along their railway and have very long stopping distances. This means that, without advanced warning, train safety cannot be guaranteed [1]. To date, all train safety systems rely on track-side infrastructure to determine the location of trains on the railways in that system, which is expensive to install and maintain. Due to the high costs, track-side infrastructure is interspersed along the railway; the spacing can vary from tens of metres for modern radio-frequency-identification-based (RFID) systems to tens of kilometres for conventional commercial transnational railway systems.

The goal of train localization is to ensure that adequate separation between trains is maintained and to provide track determination. Track determination consists of identifying which track a train is travelling on when there are two or more parallel tracks or when passing over a switch, a fork, or merger of railways. Thus, the positioning requirement for a train is relative to nearby infrastructure, which can be accomplished using a locally consistent map. Inspired by recent advances in localization systems for high-speed robotics applications [2, 3], this paper presents a proof of concept using a planar lidar as the sole sensor to estimate position and velocity of a train. Figure 1 shows the experiment setup used to validate our algorithm, which includes a lidar unit mounted on the front of a train in an underground tunnel. We chose to use lidar to avoid the use of track-side infrastructure and due to the nature of the underground railway environment (i.e., dark tunnels), which precludes the use of other common sensors such as cameras and satellite-based systems.

Train position estimation can be viewed as a 1D localization problem as the train is constrained to move on its tracks. This is a simplification of the classic 2D localization problem using planar lidar; however, the speed of the train and the nature of the modern tunnel environment introduce aspects that challenge classic localization techniques. In this context, our contribution is twofold. First, we present a mapping and localization system that addresses the unique challenges of localizing a train travelling at high speeds in a modern underground tunnel. Second, we present observations from a field deployment, described in Figure 2, on a real subway train, during which we collected datasets to develop and test our system.

II. RELATED WORK

Prior work on train localization has focused on two areas: reducing dead-reckoning error and localizing against topological maps [4]. There exists a large body of work detailing the use of various combinations of tachometer, wheel encoders, Inertial Measurement Units (IMUs), Doppler RADAR, and eddy-current sensors for the purpose...
Today’s Outline

- **VT&R 1.0**
  - background on visual teach and repeat (VT&R)
    - what is it?
    - how does it work?
    - what are its limitations?

- **VT&R 2.0**
  - using multiple route-traversal experiences to continually improve
    - **localization**
    - terrain assessment
    - path tracking
VT&R is a Building Block

‘pick and place’

‘visual route following’
Visual Route Following Applications

transportation

mining

military

planetary
VT&R Block Diagram

Teach phase

Robot base

Mobile robot

Stereo camera

Network mapper

Network map

Route planner

Path tracker

Path localizer

Safety monitor

Logger

Repeat phase

Vehicle state

Map input

Images

Images

Images

Speed profile, tracking specs

Relative state estimate

Vehicle commands

E-stop
Operational Concept
Local Metric Accuracy is Enough

“Ascending and Descending” (Escher, 1960) based on the impossible staircase (Penrose and Penrose, 1958)
VT&R: Example

Sudbury 2016
Experiments showed that the altitude at which the VT&R algorithm is performed was crucial. Teaching at a high altitude increases the camera's field of view and results in fewer map match failures. Depending on the height used during the teach pass, different sizes of interest points are used. The higher we fly, the bigger the feature must be in order to be detected by the SURF algorithm. The resolution of the bottom camera is, as mentioned in Section IV-A, very low, and therefore small features are not visible at high altitudes. To summarize, if enough big interest points are available, it is beneficial to fly high. The characteristics of the floor play an important role as well. Highly repetitive or untextured ground is to be avoided.

Since we created a 'fake' depth image by assigning the same altitude to all pixels in the image plane, the localization algorithm cannot deal well with surface height changes. A possible solution for high-altitude flights is to exclusively rely on the pressure sensor for altitude estimation, which has a reasonable accuracy for higher altitudes. Hence, ground height changes would not affect the 'fake' depth image.

VI. FUTURE WORK

This paper described preliminary results obtained from implementing, for the first time, a VT&R algorithm [6] on a quadrocopter. In Section III-B we summarized several assumptions of the current approach. Future work will aim to address those limitations. The goal is to enable route teaching through a human pilot (and not only on the cart) and to allow arbitrary flight paths (and not straight lines at a constant altitude). To do so, the following steps are necessary:

• accounting for roll and pitch when assigning the depth to each pixel;
• extending the controller to be able to track arbitrary paths and to include a vertical controller;
• deriving a feasible desired velocity profile (instead of a constant velocity), which takes vehicle constraints and constraints derived from the VT&R localization algorithm into account;
• increasing the image processing speed, which would enable higher flight speeds;
• extensive testing under various conditions (including lighting changes).

VII. CONCLUSION

We successfully showed that the visual teach and repeat (VT&R) algorithm in [6], which has previously been used for autonomous, long-range navigation of ground vehicles, can be applied to flying vehicles with the 3D sensor being replaced by a monocular, downward-facing camera and an altitude sensor. The use of a manifold map of overlapping submaps allows long-range navigation along a previously explored path without the use of GPS or a globally consistent

Figure 7. Snapshots of the 16-m-long autonomous flight, which is shown in the movie found at http://youtu.be/BRDvK4xD8ZY.
Visual Route Following Pros and Cons

**PROS**

- low-computational-cost point-to-point autonomous driving in GPS-denied environments
- exploits human experience for in-situ path planning
- exploits strong prior on safe terrain during repeating
- exploits strengths of computer vision by keeping viewpoints the same between mapping and localization (crude version of active perception)

**CONS**

- scene appearance can change (e.g., lighting, weather)
- scene geometry can change (e.g., vegetation growth, snow, construction)
- vehicle can change or be poorly modelled
- paths can be obstructed (temporarily or permanently)
Where Can We Exploit Repeat Experience to Improve?

robot base

mobile robot

stereo camera

network mapper

route planner

path tracker

path localizer

logger

vehicle state

map input

images

network map

path map

speed profile, tracking specs

relative state estimate

e-stop

vehicle commands

images

robot base

mobile robot

vehicle state

teach phase

repeat phase
Multi-Experience Localization (MEL)

The approach:
• use several visual experiences to bridge the appearance gap between teach and repeat
Single-Experience Localization Patches

- **appearance-based lidar**
  - McManus et al. (JFR 2013)
- **colour-constant images**
  - Corke et al. (IROS 2013), McManus et al. (ICRA 2014), Maddern et al. (ICRA 2014), Paton et al. (ICRA 2015)
- **multi-stereo**
  - Paton et al. (CRV 2015)

Live (repeat) and privileged (teach)

New pose to be localized using active region
Lighting-Resistant Stereo VT&R using Colour-Constant Images

Montreal 2014

Multi-Experience Localization (MEL)

new pose to be localized using active region

live (repeat)

bridging (repeat)

privileged (teach)


M Paton, K MacTavish, M Warren, and Barfoot T D. “Bridging the Appearance Gap: Multi-Experience Localization for Long-Term Visual Teach and Repeat”, IROS 2016
what’s the difference between multi-experience localization (MEL) and experience-based navigation (EBN)?

1. we add new experiences during autonomous operations
2. we have a privileged path and always want to report localization with respect to this
3. we continually add links between the live experience and the privileged path as we localize
4. we match against landmarks from several experiences in a single solve
5. we don’t decide online whether to save new experiences
6. we select which experiences to use based on what the world looks at the moment
Multi-Experience Localization (MEL)

M Paton, K MacTavish, M Warren, and Barfoot T D. “Bridging the Appearance Gap: Multi-Experience Localization for Long-Term Visual Teach and Repeat”, IROS 2016
Multi-Experience Localization (MEL)

M Paton, K MacTavish, M Warren, and Barfoot T D. “Bridging the Appearance Gap: Multi-Experience Localization for Long-Term Visual Teach and Repeat”, IROS 2016
Multi-Experience Localization (MEL)

M Paton, K MacTavish, M Warren, and Barfoot T D. “Bridging the Appearance Gap: Multi-Experience Localization for Long-Term Visual Teach and Repeat”, IROS 2016
Multi-Experience Localization (MEL)

M Paton, K MacTavish, M Warren, and Barfoot T D. “Bridging the Appearance Gap: Multi-Experience Localization for Long-Term Visual Teach and Repeat”, IROS 2016
Multi-Experience Localization (MEL)

M Paton, K MacTavish, M Warren, and Barfoot T D. “Bridging the Appearance Gap: Multi-Experience Localization for Long-Term Visual Teach and Repeat”, IROS 2016
Multi-Experience Localization (MEL)

M Paton, K MacTavish, M Warren, and Barfoot T D. “Bridging the Appearance Gap: Multi-Experience Localization for Long-Term Visual Teach and Repeat”, IROS 2016
Multi-Experience Localization (MEL)

M Paton, K MacTavish, M Warren, and Barfoot T D. “Bridging the Appearance Gap: Multi-Experience Localization for Long-Term Visual Teach and Repeat”, IROS 2016
Multi-Experience Localization (MEL)

MacTavish K A, Paton M, and Barfoot T D. “Visual Triage: A Bag-of-Words Experience Selector for Long-Term Visual Route Following”, accepted to ICRA 2017
Inlier matches with headlights

Fig. 8: This figure shows the 5th, 50th, and 95th percentile of matched features for each experience. Keyframes are created when either the number of inlier matches drops below 50, or the motion exceeds 0.3 m, or the inlier matches are reduced to maintain (as best it can) our minimum floor estimate.

This paper has presented detailed results that help quantify the impact of different lighting conditions on visual odometry performance. We hypothesized that part of the reduced accuracy at night is indirect and even.

Headlights ‘Snowstorm’ Effect
Winter Trials

M Paton, K MacTavish, M Warren, and Barfoot T D. “Bridging the Appearance Gap: Multi-Experience Localization for Long-Term Visual Teach and Repeat”, IROS 2016
Winter Trials

M Paton, K MacTavish, M Warren, and Barfoot T D. “Bridging the Appearance Gap: Multi-Experience Localization for Long-Term Visual Teach and Repeat”, IROS 2016
Learning to Assess Terrain Better

**The approach:**
- use a place-dependent change detector to compare to several past experiences
Using VT&R, the robot repeats paths previously taught by a human

Berczi L P and Barfoot T D. “Looking High and Low: Learning Place-Dependent Gaussian Mixture Height Models for Terrain Assessment”, to be submitted to IROS 2017
Place-Dependent Change Detector

Berczi L P and Barfoot T D. “Looking High and Low: Learning Place-Dependent Gaussian Mixture Height Models for Terrain Assessment”, to be submitted to IROS 2017
Learning to Track Paths Better

Two approaches:
- iterative learning control
- robust, constrained, learning-based, nonlinear, model-predictive control

```
Two approaches:
- iterative learning control
- robust, constrained, learning-based, nonlinear, model-predictive control
```
Robust, Constrained, Learning-Based, Nonlinear MPC

Overhead view:

1) Lateral constraints
2) 3 sigma confidence region
3) Robot position
4) Desired path vertices

Planning Better Routes

- route selection can be based on:
  - distance, time of day, reliability of path segments
- working towards computing statistics on networks for use in planning
- June 5-18: 120 km of autonomous driving in an old open-pit mine
Sudbury Trials Highlights

VT&R 2.0
Teach: 06/07 19:00
Repeat: 06/16 10:00

Multi-Exp Tracks
VO Tracks
Terrain Assessment
Thanks For Your Attention!