Into Darkness: Visual Navigation Based on a Lidar-Intensity-Image Pipeline

Timothy D. Barfoot, Colin McManus, Sean Anderson, Hang Dong, Erik Beerepoot, Chi Hay Tong, Paul Furgale, Jonathan D. Gammell and John Enright

Abstract Visual navigation of mobile robots has become a core capability that enables many interesting applications from planetary exploration to self-driving cars. While systems built on passive cameras have been shown to be robust in well-lit scenes, they cannot handle the range of conditions associated with a full diurnal cycle. Lidar, which is fairly invariant to ambient lighting conditions, offers one possible remedy to this problem. In this paper, we describe a visual navigation pipeline that exploits lidar's ability to measure both range and intensity (a.k.a., reflectance) information. In particular, we use *lidar intensity images* (from a scanning-laser rangefinder) to carry out tasks such as *visual odometry* (VO) and *visual teach and repeat* (VT&R) in realtime, from full-light to full-dark conditions. This lighting invariance comes at the price of coping with motion distortion, owing to the scanning-while-moving nature of laser-based imagers. We present our results and lessons learned from the last few years of research in this area.

1 Introduction

1.1 Motivation

Visual navigation is an enabling technology for mobile robots operating in challenging, real-world environments. Satellite-based positioning, such as GPS, is often insufficient or unavailable in many interesting situations: indoors, in urban canyons, under forest canopies, underground, underwater, and on other planets. As such, passive cameras and/or lidars are employed to provide as position estimation and also help with path following, hazard detection, and object recognition. Cameras and

T.D. Barfoot (✉) · C. McManus · S. Anderson · H. Dong · E. Beerepoot · C.H. Tong · P. Furgale · J.D. Gammell · J. Enright
University of Toronto Institute for Aerospace Studies, 4925 Dufferin St., Toronto, ON, Canada
e-mail: tim.barfoot@utoronto.ca

M. Inaba and P. Corke (eds.), *Robotics Research*, Springer Tracts
in Advanced Robotics 114, DOI 10.1007/978-3-319-28872-7_28

lidars are typically viewed as separate-yet-complementary sensors. Roughly speaking, (passive) cameras are used to acquire appearance information and geometry while (active) lidars are used to acquire geometry. The active nature of lidars make them well-suited to working in any lighting condition, while passive cameras struggle with lighting change.

An underexploited capability of 3D lidar is its ability to acquire appearance information through *intensity images*. Intensity data is derived from the amount of transmitted light that is reflected back from the scene. Traditionally, the raw output of a lidar sensor is thought to be a 3D point cloud; instead, we consider the output to be a pair of range and intensity images. Figure 1 provides a comparison between passive camera images and lidar intensity images for the same scene.

This paper describes how we use lidar intensity images (derived from a scanning-laser rangefinder) to build a realtime, lighting-invariant visual pipeline that leverages the heritage of the traditional stereo-camera pipeline. We were motivated to do this for two reasons: (i) to navigate in full-dark conditions, and (ii) to recognize places we had seen before, despite drastic changes in lighting. Much of our work is targeted at future planetary exploration missions. For example, permanently shadowed craters near the south pole of the Moon may contain water-ice and other useful volatiles. Missions to these craters will require robots that are able to navigate in full darkness and in

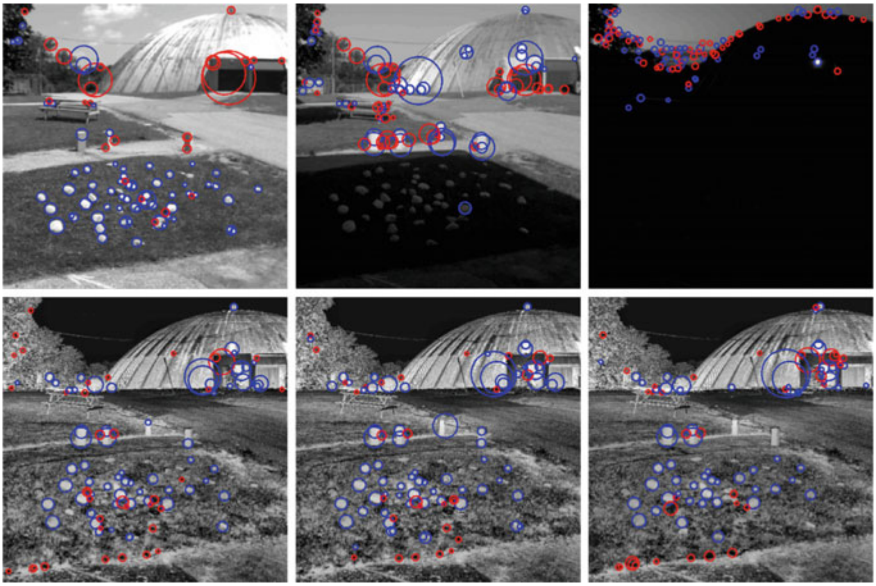

Fig. 1 Examples of passive camera (*top row*) and lidar intensity (*bottom row*) images of the same scene at three different times of day (13h38, 18h12, 05h43). Images acquired using an Optech ILRIS3D survey-grade lidar with built-in passive camera. Raw SURF [5] features are marked in all images. We see that the lidar intensity images and features look very similar regardless of the sunlight conditions, while the passive camera images and features change significantly with lighting

realtime (due to the low-latency communications and short lunar day). However, beyond planetary exploration, robotics generally requires the ability to recognize previously visited places in order to build consistent survey maps and re-drive routes.

We successful built a lidar-intensity-image pipeline suitable for visual odometry (VO) [10, 21, 23] and visual teach and repeat (VT&R) [22, 24]. However, due to the scanning-while-moving nature of laser-based imaging, motion distortion can be significant if the sensor's motion is high relative to its framerate. In order to obtain an accurate motion estimate, we were forced to innovate ways of coping with this distortion [2, 3, 26–29], but we believe that it has been worth the effort to achieve lighting invariance.

The rest of the paper is organized as follows. The remainder of Sect. 1 provides a brief summary of related work and introduces the lidar-intensity-image pipeline. Section 2 discusses motion distortion and our various approaches to overcome it. Section 3 provides the results of some visual-odometry and visual-teach-and-repeat experiments. Section 4 discusses lessons learned and concludes the paper.

1.2 Related Work

We only briefly review other works that have used lidar intensity images (a.k.a., reflectance images) for motion estimation. McManus et al. [23, 24] provide more information. Laser intensity images have been used in the past for surveying applications [8, 18] and some researchers have looked at automated point-cloud registration techniques that use 2D interest points in the intensity images [1, 6].

The SwissRanger sensor, a 'flash lidar', also produces intensity/range images but using a different principle than laser-based scanners. Unlike a laser scanner, the SwissRanger uses an array of 24 LEDs to simultaneously illuminate a scene, offering the advantage of higher framerates. However, the SwissRanger has a limited FOV, short maximum range, and is very sensitive to environmental noise. Weingarten et al. [30] used images from the SwissRanger for robotics applications; however, their method, as well as others that followed [11, 32], only used range data from the sensor and not the intensity data.

May et al. [20], and later Ye and Bruch [31], were the first to develop 3D mapping and appearance-based egomotion techniques using a SwissRanger. May et al. [20] used intensity images to employ two feature-based methods for motion estimation: a Kanade–Lucas–Tomasi (KLT)-tracker and frame-to-frame VO using SIFT features. Their results indicated that the SIFT approach yielded more accurate motion estimates than the KLT approach, but less accurate than the iterative closest point (ICP) method. Although May et al. [20] demonstrated that frame-to-frame VO might be possible with the SwissRanger, the largest environment in which they tested was a 20 m long indoor hallway, with no groundtruth. Furthermore, laser scanners are very different from flash lidars in that they scan the scene with a single light source, introducing new problems such as image formation and image distortion caused by

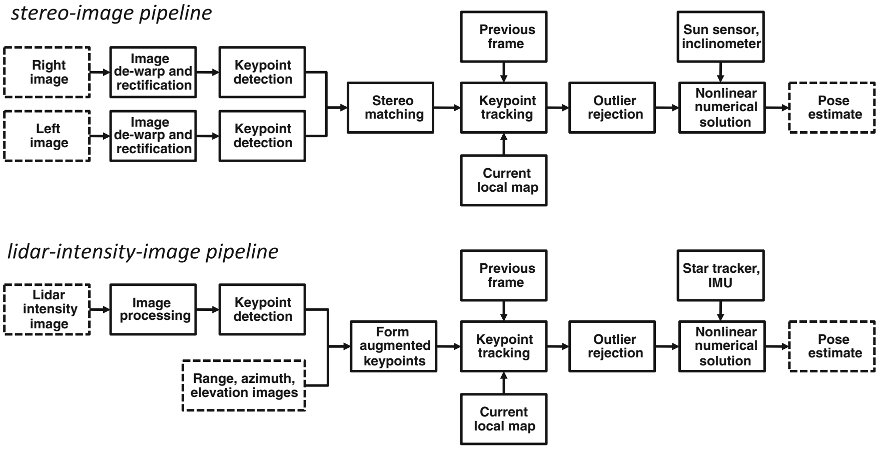

Fig. 2 Stereo-image (*above*) and lidar-intensity-image (*below*) visual pipelines. Both pipelines show the main steps required to go from raw images on the *left* to a pose solution on the *right*. On the surface, we see that switching to lidar intensity images only alters the initial steps. However, due to the scanning-while-moving nature of the lidar-intensity images, most of the blocks require modification to compensate for motion distortion

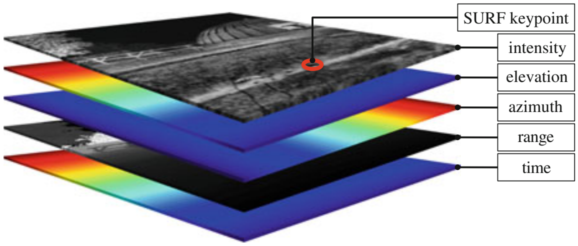

Fig. 3 Image stack concept. We detect sparse keypoints in the intensity image (at subpixel locations) then, for each keypoint, pierce through the stack to look up the associated azimuth, elevation (based on lidar's mirror angles), range, and time (using bilinear interpolation). The result is a keypoint augmented with its 3D position and timestamp

moving and scanning at the same time. We believe our work is the first and only one to use laser-based intensity images in a realtime visual pipeline.

1.3 Lidar-Intensity-Image Pipeline

Figure 2 provides an overview of our lidar-intensity-image pipeline, as well as the typical stereo-image pipeline for comparison. The purpose of these pipelines is to determine the robot's motion from a sequence of images. The main steps of the lidar pipeline are:

(i) acquire a lidar intensity image,
(ii) carry out preprocessing to improve contrast in the image—*either adaptive histogram equalization or a linear range correction*,
(iii) detect sparse keypoints in the intensity image—*we use SURF implemented on a GPU but other methods should work*,
(iv) build an 'image stack' (cf., Fig. 3) and for each keypoint pierce through to form augmented keypoints—*this provides the 3D position (azimuth, elevation, range) and time of each keypoint*,
(v) track keypoints—*we match to the previous frame (VO) and a local map (VT&R)*,
(vi) detect and reject outliers—*we use RANSAC*,
(vii) solve for the robot's motion using a nonlinear, least-squares method—*we can incorporate an attitude sensor such as a star tracker and/or IMU to help determine orientation*.

We carry out these steps every time a new image is acquired.

Figure 4 compares the lidar-intensity-image and stereo-image pipelines on the VO problem. With the robot stopping every time it gathered images (approximately every 0.5 m), we can see that the lidar and stereo pipelines both provide good estimates of the robot motion compared to GPS groundtruth. However, if we allow the lidar to acquire images while in motion, the VO performance degrades quickly as the images becomes distorted. We discuss this motion distortion and our efforts to compensate for it in detail in the next section.

2 Coping with Motion Distortion

2.1 Nature of Distortion

In this paper, we are concerned with laser-based 3D rangefinders that are capable of producing high-quality intensity images. The main sensor we use for realtime oper-

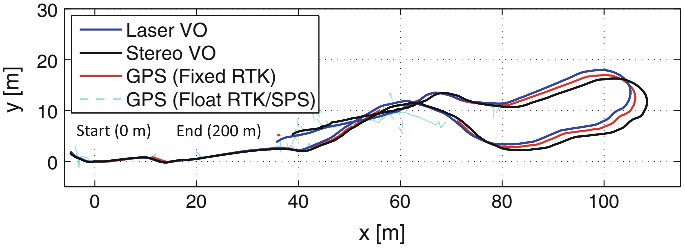

Fig. 4 Comparison of lidar-intensity-image (laser) and stereo-image VO. In this experiment the robot stopped every time images were acquired (approximately every 0.5 m). Both algorithms are able to provide reasonable estimates (compared to GPS groundtruth) in this stop-scan-go paradigm

ations is the Autonosys LVC0702, which uses a combined nodding and polygonal mirror assembly to steer a single laser beam through a raster pattern in order to build intensity/range images (cf., Fig. 5; left). It works out to about 50 m in range.

To our knowledge, the Autonosys unit has the highest pulse repetition frequency (PRF) of the single-laser scanners on the market at 500,000 points/s. We use the unit in a mode that produces 480 × 360 pixel intensity images at 2 Hz. Even for a robot moving at a modest speed, say 0.5 m/s, there is noticeable distortion in the images if they are acquired during motion (cf., Fig. 5; right). This is essentially an exaggerated 'rolling shutter' effect; an image is gathered over a fraction of a second, with every pixel acquired at a unique (known) time.

We considered using a Velodyne HDL64E for our work, but found that the narrow vertical field of view (16°) and resolution (64 pixels) made the resulting intensity images unsuitable for our approach; it would also require careful inter-calibration of intensity values derived from the 64 separate laser sources.

Flash lidar does not suffer from the same motion distortion issues as laser-based rangefinders and eventually may be provide lighting-invariant imagery. Currently, however, inexpensive units such as the SwissRanger SR4000 have limited range capabilities (less than 10 m), struggle to cope with sunlight, and have low resolution (e.g., 176 × 144 pixels). More expensive units, such as the ASC TigerEye, have longer range and work outside but still have limited resolution (e.g., 128 × 128 pixels) and thus are not suitable for our image pipeline as of yet.

The Microsoft Kinect sensor produces similar data products to lidar (i.e., intensity and depth images) but it does not work outside in direct sunlight, its range is limited to approximately 7 m, and the intensity image comes from a passive camera and thus is not lighting-invariant.

Thus, for the time being, we need to be able to handle motion distortion in laser-based lidar images. The next sections describe our efforts to cope with this issue.

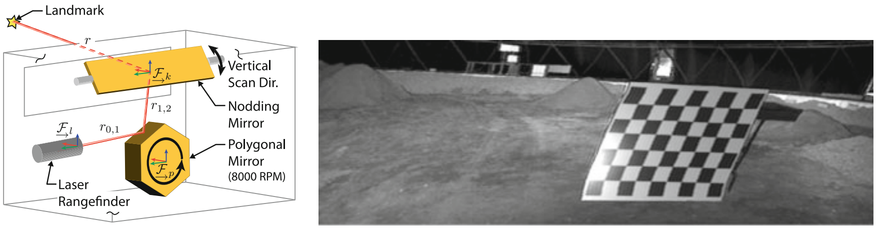

Fig. 5 3D range sensors based on a single laser use a mirror assembly (*left*) to steer the beam through a raster pattern in order to build an image. If the lidar is mounted on a moving robot, the scene can undergo a non-affine transformation during imaging; in the checkerboard example (*right*), some of the straight lines are distorted because the lidar was in motion during acquisition

2.2 Effect on Features

In our pipeline, we extract SURF features from the raw, motion-distorted images and track them on a frame-to-frame basis (cf., Fig. 6). The effect of scanning while moving has not been so severe as to cause feature tracking to fail catastrophically in our experiments so far. Thus, we have avoided carrying out full-image motion compensation. However, if the same scene is imaged twice at two very different vehicle speeds, it is intuitive that feature matching will fail as the images will experience different amounts of motion distortion. We are currently carrying out a study to characterize how large we can make the speed differential and still successfully match features.

2.3 Motion-Compensated RANSAC

The next step is to identify outlying feature tracks and remove them from the pipeline. We originally used the out-of-the-box random sample and consensus (RANSAC) algorithm [12] typically found in the stereo-camera pipeline. This worked reasonably well most of the time, but we found that the threshold used to separate inliers from outliers was difficult to tune. We would typically end up with a lot of false negatives (i.e., throwing away many good feature tracks) or a handful of false positives (i.e., letting some bad feature tracks in). We see these two cases in Fig. 7 (left; middle). Note that when the threshold is tight the inliers are restricted to a horizontal band, implying near-simultaneous capture. This is expected as the lidar scans quickly left-to-right while slowly scanning up-and-down.

We found the reason for RANSAC's difficulty to be our choice of model. Typically, RANSAC in the stereo-camera pipeline seeks to find a rigid, frame-to-frame pose change that explains the most data. Due to the motion distortion present in the

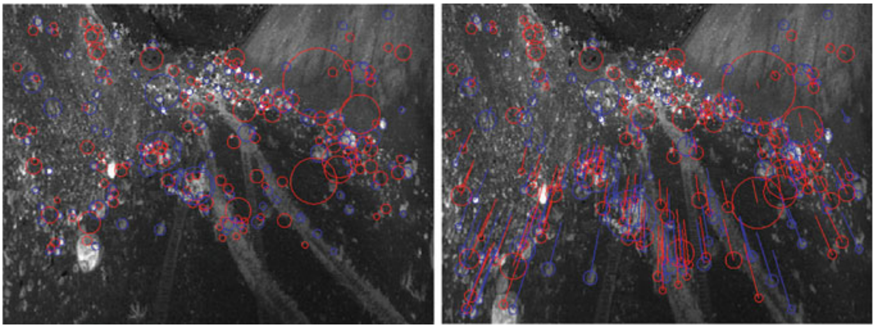

Fig. 6 We extract SURF features from the raw (i.e., motion-distorted) images (*left*) and track them on a frame-to-frame basis (*right*). This allows us to only motion-compensate a sparse number of points rather than the entire image

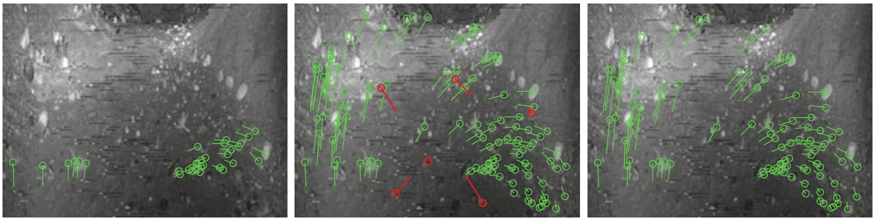

Fig. 7 In a VO pipeline, outliers are usually detected/rejected on a frame-to-frame basis using RANSAC [12], which is used to solve for the rigid pose change between two frames that explains the most features. Unfortunately, for intensity images gathered during motion, this model is insufficient and results in false negatives if the matching threshold is too tight (*left*) or false positives (*red*) if the threshold is too loose (*middle*). To overcome this, we created a motion-compensated version of RANSAC that solves for the 6DOF velocity that explains the most features (*right*) [2]

imagery, we found that it was much more effective to have RANSAC find the *constant velocity* that explains the most data over a two-frame time interval. It turns out that, as with the rigid pose change, the minimum number of feature tracks needed to fit a constant velocity is still three. Thus, RANSAC proceeds as usual, but with this change of model. We see the improved outlier rejection in Fig. 7 (right). We are now able to keep the decision threshold tight and still obtain lots of good feature tracks without false positives. Anderson and Barfoot [2] provide further details.

2.4 Continuous-Time Estimation for Pose

After removing outlying feature tracks, the last major step in the pipeline is to solve for the pose change using an iterative, nonlinear, least-squares method such as *bundle adjustment* [7]. It is again important in this step to account for the proper timestamps of all observed features in order to combat the motion distortion. Traditional approaches represent the robot's trajectory in discrete time (cf., Fig. 8; left). This is sufficient because the exposure times associated with passive-camera image capture are so short that they can be thought of as instantaneous.

Unfortunately, for lidar intensity images, this is not a good assumption. If we simply put a discrete-time pose at nominal image capture times, we cannot account for the actual (and varying) timestamps of the measurement observations and our VO solution ends up being poor. If we put a discrete-time pose at every unique measurement timestamp, the problem is computationally intractable but also underconstrained (without including an additional prior).

We found a better idea was to consider the robot's trajectory to be a continuous function of time (cf., Fig. 8; right) so that we could query it at any particular time at which a feature was observed. Thus, in general, we write the robot's trajectory as $\mathbf{x}(t)$ and then build a measurement reprojection error term as

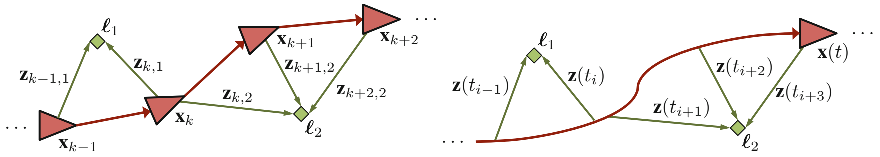

Fig. 8 Typically in robotics, the robot's trajectory is represented in discrete time (*left*). We represent the robot's trajectory in continuous time (*right*), $\mathbf{x}(t)$, which allows us to query the pose at the exact times the landmarks were observed; this is important due to the scanning-while-moving nature of laser-based scanners

$$\mathbf{e}_{k,j} = \mathbf{z}_{k,j} - \mathbf{g}\left(\mathbf{x}(t_k), \boldsymbol{\ell}_j\right), \tag{1}$$

where $\mathbf{z}_{k,j}$ is the observation of landmark j at time t_k, $\mathbf{g}(\cdot,\cdot)$ is the lidar's observation model, and $\boldsymbol{\ell}_j$ is the position of landmark j. We sum over all the measurements to build a nonlinear, least-squares cost function, J, that we seek to minimize with respect to $\mathbf{x}(t)$:

$$J(\mathbf{x}(t), \boldsymbol{\ell}) = \frac{1}{2}\sum_{k,j} \mathbf{e}_{k,j}^T \mathbf{R}_{k,j}^{-1} \mathbf{e}_{k,j}, \tag{2}$$

where $\mathbf{R}_{k,j}$ is the measurement noise covariance associated with $\mathbf{e}_{k,j}$. We then seek to solve the following optimization problem,

$$\left\{\mathbf{x}(t)^\star, \boldsymbol{\ell}^\star\right\} = \underset{\mathbf{x}(t),\, \boldsymbol{\ell}}{\operatorname{argmin}}\, J(\mathbf{x}(t), \boldsymbol{\ell}) \tag{3}$$

for the optimal robot trajectory, $\mathbf{x}(t)^\star$, and landmark positions, $\boldsymbol{\ell}^\star$. We use Gauss-Newton optimization with a robust kernel to find the best trajectory estimate.

Naturally, we still need to discretize $\mathbf{x}(t)$ in some way to make the state estimation problem tractable and have considered a few ways of doing so:

(i) *linear interpolation*: represent the trajectory using discrete-time poses, $\mathbf{x}_i$, with linear pose interpolation in between to evaluate measurement error terms at their appropriate times [10]

(ii) *spline*: represent the trajectory as a weighted sum of a finite number of known basis functions, $\mathbf{x}(t) = \sum_i c_i \boldsymbol{\phi}(t)$, and solve for the optimal coefficients, c_i [13, 15, 25]

(iii) *spline velocity*: represent the trajectory in terms of velocity (i.e., a relative pose trajectory), which we still consider to be a weighted sum of a finite number of known basis functions: $\dot{\mathbf{x}}(t) = \sum_i c_i \boldsymbol{\phi}(t)$, and solve for the optimal coefficients, c_i [3]

(iv) *Gaussian process*: represent the robot trajectory nonparametrically as a Gaussian process, $\mathbf{x}(t) \sim \mathcal{GP}(\boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t, t'))$ and solve for the pose at desired times [27, 29]

Regardless of the method, we solve for only a *finite* number of variables, in each case optimizing the robot trajectory based on the observed feature tracks. Our preferred approach is to do this online in a sliding-window style estimator where we estimate a small temporal section of the robot's trajectory (i.e., several seconds) and then slide the optimization window along to incorporate the next batch of measurements.

2.5 Celestial Attitude Corrections

While the continuous-time trajectory estimation approach is already quite accurate, we can also incorporate absolute attitude (i.e., orientation) corrections into our pose solution, as depicted in Fig. 2. As our motivation has been planetary exploration, we have primarily investigated celestial observations (with ephemeris, coarse location on the planet, and time/date) as a source of absolute attitude data.

In the daytime, we can use a dedicated sun sensor/inclinometer to provide attitude corrections very inexpensively [19]. Alternatively, we have found that it is actually possible with some laser-based imagers to use intensity/range images as a make-shift sun sensor [16]; the sun appears as a blob of points with maximum intensity and zero range (cf., Fig. 9; middle).

At nighttime, a small star tracker/inclinometer (cf., Fig. 9; right) is the preferred source of absolute attitude information. As star measurements can be provided frequently and during motion [17], they seem to be a natural pairing for lidar to support dark navigation.

Regardless of the source of absolute attitude measurements, we typically incorporate them into the VO pipeline by introducing additional error terms in Eq. (2). These celestial attitude sensors can be included at very little additional mass, power, and computational cost and are very beneficial to the accuracy of the VO solution over long distances [19].

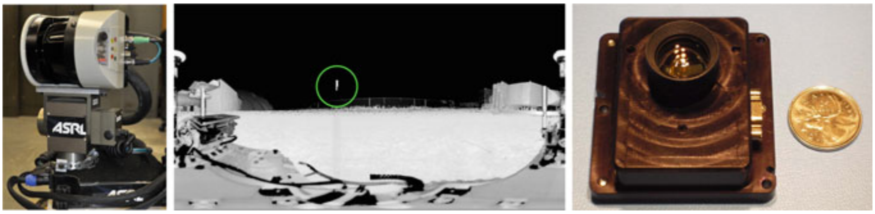

Fig. 9 Celestial/gravity observations can be used to correct rover orientation in a VO pipeline. In the daytime, we use the sun and either a dedicated sun sensor (not shown) or lidar intensity images [16]. For example, a SICK laser (*left*) was swept 360° to produce a panoramic intensity image (*middle*); the sun shows up as an artifact (*circled in green*) with maximum intensity and zero range. At nighttime, we use a small star tracker (*right*), which directly outputs full 3DOF orientation while the robot is motion [17]

3 Experimental Results

3.1 Setup

We gathered a large-scale lidar intensity image dataset at a sand and gravel pit near Sudbury, Ontario, Canada [4]. The ROC6 robot was equipped with the Autonosys lidar and DGPS for groundtruth positioning. The robot travelled the same 1.1 km course 10 times in a diurnal cycle, or approximately every 2.5 h for a 25 h period. Figure 10 depicts the experimental setup and the path the robot took based on DGPS groundtruth. The dataset is available for download from our webpage: http://asrl.utias.utoronto.ca/datasets/abl-sudbury.

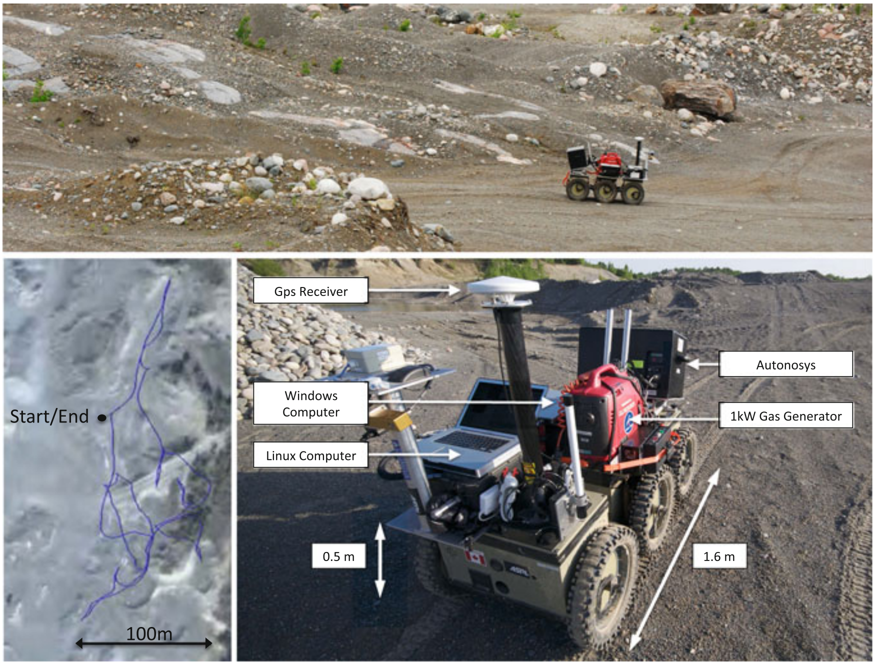

Fig. 10 We gathered 11 km of lidar-intensity-image data and DGPS groundtruth at the Ethier Sand and Gravel Pit near Sudbury, Ontario (*top*). The Autonosys lidar was mounted on the ROC6 field robot (*right*) and the same 1.1 km circuit (*left*) was repeated every 2.5 h for 25 h straight, ensuring data was gathered across an entire diurnal cycle (full-light to full-dark). The entire dataset is available on our webpage: http://asrl.utias.utoronto.ca/datasets/abl-sudbury [4]

3.2 Visual Odometry

Figure 11 shows an example of our motion-compensated visual odometry algorithm running on one of the *full-dark* 1.1 km circuits from our Sudbury dataset. It was a very cloudy night and therefore pitch black during the experiment. It should be noted that while the robot frequently revisited places it had been before, we are not detecting and exploiting loop closures in this experiment (i.e., we are not doing SLAM), merely dead-reckoning from sequential lidar data (in a sliding window).

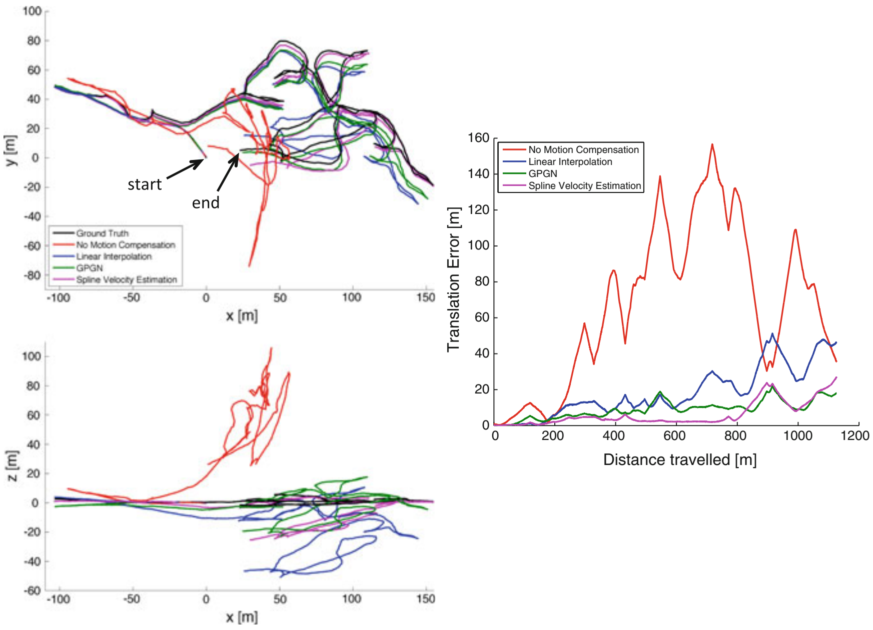

Fig. 11 VO results for one of the *full-dark* 1.1 km gravel-pit circuits, comparing all of the various motion-compensation strategies we have used over the last few years. Roughly speaking, all the methods are a big improvement compared to not compensating for the motion distortion. The spline-velocity estimation of Anderson and Barfoot [3] (*purple*) and the GPGN method of Tong and Barfoot [27] (*green*) do the best. Total Euclidean error (*right*) is much lower for the motion-compensated methods

The plot compares (i) GPS groundtruth, (ii) no motion compensation (i.e., traditional discrete-time VO), (iii) linear interpolation, (iv) Gaussian process Gauss-Newton (GPGN), and (v) spline velocity estimation (integrated after the fact to produce $\mathbf{x}(t)$). All the algorithms use the same VO pipeline, except for the last step involving the nonlinear, numerical pose solution. To provide a fair comparison we used our motion-compensated RANSAC feature tracks for all the algorithms; with traditional feature tracks the performance would be poor for all algorithms, even the discrete-time estimator.

We see some variability in performance across the algorithms, but there are different tuning parameters in each algorithm, making the comparison rough at best. At a high level, the total Euclidean error (cf., Fig. 11; right) shows all that the motion-compensated methods have much lower error than the traditional non-motion-compensated algorithm. On this particular dataset, the GPGN and spline velocity methods fared the best, with linear interpolation performing worse.

The motion compensation in the pose solution clearly helps and comes at little extra computational cost over the discrete-time estimator; we still do nonlinear, iterative least-squares with a robust cost function and estimate a similar number of variables, but the accuracy is higher.

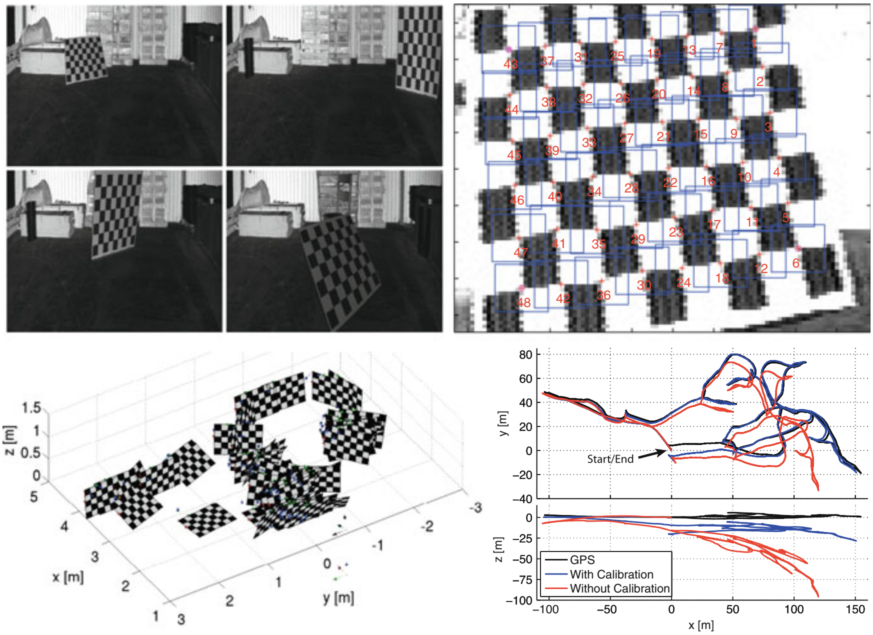

Fig. 12 We developed a lidar calibration tool inspired by the standard passive camera calibration approach. We present a number of views of checkboards to the lidar and capture intensity (*top-left*) and range images (not shown). We then automatically extract the locations of the checkboard corners from the intensity images (*top-right*). We simultaneously solve for the checkerboard poses and intrinsic parameters of the lidar (*bottom-left*). Calibration greatly improves the quality of our VO solution (*bottom-right*); comparison carried out using the spline-velocity approach

3.3 Lidar Calibration

As with any imaging sensor, our lidar-intensity-images require calibration to make the 3D positions of the landmarks accurate (cf., Fig. 12). This procedure is described by Dong et al. [9], but briefly the calibration uses multiple images of checkerboards at known locations to solve for intrinsic parameters by calculating the poses of the checkerboards. We effectively calibrate the azimuth, elevation, and range images in the lidar image stack of Fig. 3. Figure 12 (bottom-right) shows the effect of this calibration on the quality of our VO solution; we see that the calibration is just as important as motion compensation to producing an accurate VO solution.

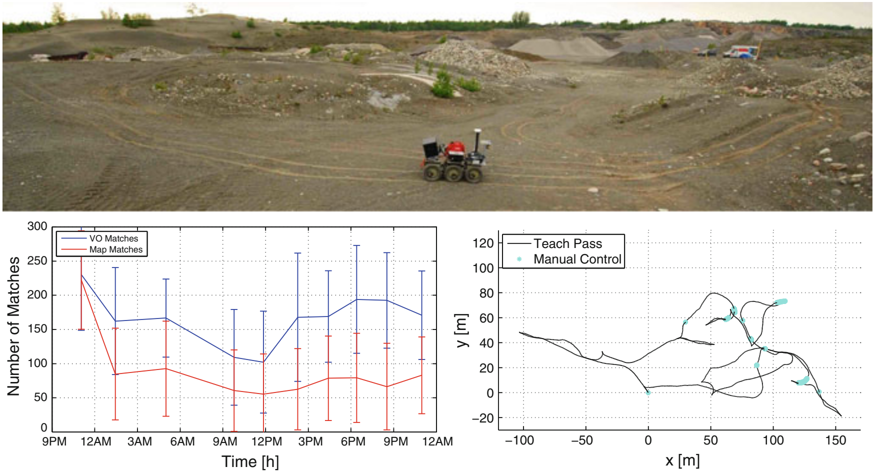

Fig. 13 Our visual-teach-and-repeat method was used to autonomously repeat the 1.1 km gravel-pit circuit 10 times. The path was taught in full daylight and repeated throughout an entire diurnal cycle. The method allowed the ROC6 robot to drive autonomously almost directly in its taught tracks (*top*) for 99.7 % of the distance. To localize relative to the path, the system matched features to the previous frame (VO) and to the map built during the teach pass; we see good numbers of features for all ten repeats (*left*), independent of the time of day. Only when both matching methods failed simultaneously were we required to exert manual control (*right*) to move the robot past small difficult sections (0.3 % of distance)

3.4 Visual Teach and Repeat

The second (and perhaps more important) experiment we will discuss is visual teach and repeat (VT&R). Chronologically, we carried out this experiment before our work to motion-compensate VO and so when we refer to the VO pipeline in this section, it is the basic, discrete-time version without motion compensation. In fact, the lidar dataset [4] described above was actually gathered as by-product of the VT&R experiment. It turns out for path repeating, the non-motion-compensated solution is almost good enough, but we decided to work on motion compensation primarily to improve the robustness of VT&R.

The idea behind VT&R is to pilot a robot manually along a route once to ‘teach’ it, and then to autonomously repeat the route many times. We accomplish this by running the VO pipeline during the teaching phase to estimate motion, but we save all of the features used to estimate VO, relative to the camera view from which they were first observed. During ‘repeat’, we match features from the live camera view to those stored in the map (as well as to the previous live view; cf., Fig. 2). This allows the robot to determine its pose relative to the taught path. A feedback controller then steers the robot to bring the path-tracking errors to zero over time. If the robot cannot match to the map, then VO is used to propagate the previous path-tracking errors forward in time. The end result is a robot that can drive directly in its taught tracks, using only visual feedback (i.e., no GPS).

We originally carried out this work using a stereo camera [14] before switching to lidar. However, we found that if the lighting changed too much between the teach and repeat phases, the robot would be unable to match its live view to the map reliably. This was the main reason we decided to explore using lidar intensity images, which can be matched across a wide variety of lighting conditions.

To demonstrate the lighting-invariant capabilities of our lidar pipeline, we taught a 1.1 km route in daylight (cf., Fig. 13; top) and then repeated it autonomously every 2.5 h for the next 25 h [24]. This meant we were matching light-to-light, light-to-dusk, light-to-dark, and light-to-dawn. Figure 13 (left) shows how many features we were on average able to match to the map (red) and previous image (blue) across all ten repeat runs; both numbers remain fairly constant. By distance, our system was 99.7 % successful, with the remaining 0.3 % requiring some minor manual interventions. Figure 13 (right) shows the union of the few places requiring manual interventions across all 10 repeat runs. Our average path-tracking error was about 8 cm RMS as measured by DGPS.

We found that using a VO pipeline without motion compensation inside our VT&R system meant we could not drive very far without matching to the map. We have yet to put our VO motion-compensation improvements back into VT&R, but believe this will further increase robustness by handling more of the cases where it is difficult to match to the map.

4 Conclusion and Future Work

We have discussed our experiences in building a visual pipeline based on lidar intensity images for both visual odometry and visual teach and repeat. Our major lessons learned along the way are:

(i) lidar intensity images offer excellent lighting invariance and can be used successfully in a visual pipeline,
(ii) lidar image stacks require careful calibration to achieve high-quality VO results (i.e., comparable to the stereo-camera pipeline),
(iii) the scanning-while-moving nature of laser-based imagers results in motion distortion that affects the accuracy of VO if left unchecked,
(iv) visual teach and repeat is possible even without motion compensation but will be more robust with it,
(v) it is possible to compensate for motion distortion in the RANSAC and pose solution steps of the VO pipeline,
(vi) it is possible to extract features from the raw intensity images, but this may no longer work if the motion distortion becomes too high,
(vii) absolute attitude corrections can be used to correct the robot's orientation and further improve the accuracy of the pipeline.

We believe our work shows not only that it is possible to build a VO pipeline that will work in the dark (and any other lighting condition), but that we can successfully match features across lighting conditions. We used this matching ability to build a lighting-invariant, visual-teach-and-repeat system, but we see this enabling other lighting-invariant robotics capabilities as well. For example, our next step is to do place recognition across lighting conditions. We hope that an affordable version of the lidar we used in this work becomes available within a few years, as we believe this could have a big impact in enabling real-world applications.

References

1. Abymar, T., Hartl, F., Hirzinger, G., Burschka, D., Frohlich, C.: Automatic registration of panoramic 2.5D scans and color images. In: Proceedings of the International Calibration and Orientation Workshop EuroCOW, vol. 54, Castelldefels, Spain (2007)
2. Anderson, S., Barfoot, T.D.: RANSAC for motion-distorted 3D visual sensors. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan (2013)
3. Anderson, S., Barfoot, T.D.: Towards relative continuous-time SLAM. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 1025–1032, Karlsruhe, Germany (2013)
4. Anderson, S., McManus, C., Dong, H., Beerepoot, E., Barfoot, T.D.: The gravel pit lidar-intensity imagery datase. Technical report ASRL-2012-ABL001, University of Toronto (2012)
5. Bay, H., Ess, A., Tuytelaars, T., Gool, L.: SURF: speeded up robust features. Comput. Vis. Image Underst. (CVIU) **110**(3), 346–359 (2008)
6. Bohm, J., Becker, S.: Automatic marker-free registration of terrestrial laser scans using reflectance features. In: Proceedings of the 8th Conference on Optical 3D Measurement Techniques, pp. 338–344, Zurich, Switzerland (2007)
7. Brown, D.C.: A solution to the general problem of multiple station analytical stereotriangulation. Rca-mtp data reduction technical report no. 43, Patrick Airforce Base, Florida (1958)
8. Dold, C., Brenner, C.: Registration of terrestrial laser scanning data using planar patches and image data. In: Proceedings of the ISPRS Commission V Symposium 'Image Engineering and Vision Metrology', vol. XXXVI, Dresden, Germany (2006)

9. Dong, H., Anderson, S., Barfoot, T.D.: Two-axis scanning lidar geometric calibration using intensity imagery and distortion mapping. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 3657–3663, Karlsruhe, Germany (2013)
10. Dong, H.J., Barfoot, T.D.: Lighting-invariant visual odometry using lidar intensity imagery and pose interpolation. In: Proceedings of the International Conference on Field and Service Robotics (FSR), Matsushima, Japan (2012)
11. Droeschel, D., Holz, D., Stuckler, J., Behnke, S.: Using time-of-flight cameras with active gaze control for 3D collision avoidance. In: Proceedings of the IEEE International Conference on Robotics and Automation, Anchorage, Alaska, United States (2010)
12. Fischler, M., Bolles, R.: Random sample and consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
13. Furgale, P., Rehder, J., Siegwart, R.: Unified temporal and spatial calibration for multi-sensor systems. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan (2013)
14. Furgale, P.T., Barfoot, T.D.: Visual teach and repeat for long-range rover autonomy. J. Field Robot., Special issue on visual mapping and navigation outdoors, **27**(5):534–560 (2010)(video1), (video2), (video3)
15. Furgale, P.T., Barfoot, T.D., Sibley, G.: Continuous-time batch estimation using temporal basis functions. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 2088–2095, St. Paul, USA (2012)
16. Gammell, J.D., Tong, C.H., Barfoot, T.D.: Blinded by the light: exploiting the deficiencies of a laser rangefinder for rover attitude estimation. In: Proceedings of the 10th Conference on Computer and Robot Vision (CRV), pp. 144–150, Regina, Canada (2013)
17. Gammell, J.D., Tong, C.H., Berczi, P., Anderson, S., Barfoot, T.D., Enright, J.: Rover odometry aided by a star tracker. In: Proceedings of the IEEE Aerospace Conference, pp. 1–10, Big Sky, MT (2013)
18. Kretschmer, U., Abymar, T., Thies, M., Frohlich, C.: Traffic construction analysis by use of terrestrial laser scanning. In: Proceedings of the ISPRS WG VIII/2 Laser-Scanners for Forest and Landscape Assessment, vol. XXXVI, pp. 232–236, Freiburg, Germany (2004)
19. Lambert, A., Furgale, P.T., Barfoot, T.D., Enright, J.: Field testing of visual odometry aided by a sun sensor and inclinometer. J. Field Robot., Special issue on Space Robotics, **29**(3):426–444 (2012) (video)
20. May, S., Fuchs, S., Malis, E., Nuchter, A., Hertzberg, J.: Three-dimensional mapping with time-of-flight cameras. J. Field Robot. **26**(11–12), 934–965 (2009)
21. McManus, C., Furgale, P.T., Barfoot, T.D.: Towards appearance-based methods for lidar sensors. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 1930–1935, Shanghai, China (2011)
22. McManus, C., Furgale, P.T., Stenning, B.E., Barfoot, T.D.: Visual teach and repeat using appearance-based lidar. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 389–396, St. Paul, USA (2012)
23. McManus, C., Furgale, P.T., Barfoot, T.D.: Towards lighting-invariant visual navigation: an appearance-based approach using scanning laser-rangefinders. Robot. Auton. Syst. **61**, 836–852 (2013)
24. McManus, C., Furgale, P.T., Stenning, B.E., Barfoot, T.D.: Lighting-invariant visual teach and repeat using appearance-based lidar. J. Field Robot. **30**(2), 254–287 (2013)
25. Oth, L., Furgale, P., Kneip, L., Siegwart, R.: Rolling shutter camera calibration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1360–1367 (2013)
26. Tong, C., Dong, H., Barfoot, T.D.: Pose interpolation for laser-based visual odometry. Submitted to the J. Field Robot., 27 May 2013. Manuscript # ROB-13-0044
27. Tong, C.H., Barfoot, T.D.: Gaussian process Gauss-Newton for 3D laser-based visual odometry. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 5184–5191, Karlsruhe, Germany (2013)

28. Tong, C.H., Furgale, P.T., Barfoot, T.D.: Gaussian process Gauss-Newton: non-parametric state estimation. In: Proceedings of the 9th Conference on Computer and Robot Vision (CRV), pp. 206–213, Toronto, Canada (2012)
29. Tong, C.H., Furgale, P.T., Barfoot, T.D.: Gaussian process Gauss-Newton for non-parametric simultaneous localization and mapping. Int. J. Robot. Res. **32**(5), 507–525 (2013)
30. Weingarten, J.W., Gruner, G., Siegwart, R.: A state-of-the-art 3D sensor for robot navigation. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robotics and Systems, vol. 3, pp. 2155–2160, Lasuanne, Switzerland (2004)
31. Ye, C., Bruch, M.: A visual odometry method based on the SwissRanger SR4000. In: Proceedings of the SPIE—Unmanned Systems Technology XII, vol. 7692 (2010)
32. Yuan, F., Swadzba, A., Philippsen, R., Engin, O., Hanheide, M., Wachsmuth, S.: Laser-based navigation enhanced with 3D time-of-flight data. In: Proceedings of the IEEE International Conference on Robotics and Automation, Kobe, Japan (2009)