Leveraging Experience for Large-Scale LIDAR Localisation in Changing Cities

Will Maddern, Geoffrey Pascoe and Paul Newman

Abstract—Recent successful approaches to autonomous vehicle localisation and navigation typically involve 3D LIDAR scanners and a static, curated 3D map, both of which are expensive to acquire and maintain. In this paper we propose an experience-based approach to matching a local 3D swathe built using a push-broom 2D LIDAR to a number of prior 3D maps, each of which has been collected during normal driving in different conditions. Local swathes are converted to a combined 2D height and reflectance representation, and we exploit the GPU rendering pipeline to densely sample the localisation cost function to provide robustness and a wide basin of convergence. Prior maps are incrementally built into an experience-based framework from multiple traversals of the same environment, capturing changes in environment structure and appearance over time. The LIDAR localisation solutions from each prior map are fused with vehicle odometry in a probabilistic framework to provide a single pose solution suitable for automated driving. Using this framework we demonstrate real-time centimetre-level localisation using LIDAR data collected in a dynamic city environment over a period of a year.

I. INTRODUCTION

Autonomous consumer vehicles have progressed from a distant goal a decade ago at the first DARPA Grand Challenge [1] to a common sight in some parts of the world, with Google and other research groups demonstrating fully autonomous prototypes that have driven hundreds of thousands of miles [2], [3], [4]. The most successful of these prototypes combine 3D LIDAR scanners with high-accuracy GPS+INS systems to localise with centimetre-precision in curated, globally consistent 3D prior maps [5].

A major obstacle preventing widespread deployment of autonomous road vehicles is the prohibitive cost of the sensor systems. At over £160,000 at the time of writing, a typical autonomous vehicle sensor suite consisting of a Velodyne 3D LIDAR scanner [6] and an Applanix POS-LV GPS+INS system [7] remains far beyond the price range of sensors for consumer vehicles. This high cost has lead researchers to investigate cheaper sensor alternatives such as cameras; however, despite recent advances in exploiting consumer cameras for autonomous vehicle localisation [8], [9], [10], [11], many challenges remain before a truly reliable vision-based localisation system is available for widespread use.

A secondary cost not often considered in the domain of automated driving is the costs involved in building and maintaining maps suitable for autonomous vehicle operation. While Google Street View [12] serves as proof that it is possible to produce 3D street-level maps on a truly global scale with dedicated survey vehicles, these surveys only capture the structure and appearance of the environment once every few years. Surveying environments of this scale at a rate required to capture scene variation relevant for autonomous driving (such as changes in road surface, buildings, parked vehicles, local vegetation and seasonal variation) would be prohibitively expensive for a map provider, and any economic benefits of selling automated driving systems may be offset by the constant need to re-survey the environment that these autonomous vehicles operate in.

Instead, we propose that an “experience-based” approach...
[13], as shown in Fig. 1, is a natural method of dealing with environmental change over time in the domain of autonomous vehicles. Rather than defining a single static global map that all vehicles must use (and that must be updated frequently), each vehicle independently builds and maintains a set of “experiences”, capturing representations specific to its operating environment. As the vehicle traverses a route a number of times, it will intrinsically capture richer representations of more dynamic locations as required - since map creation is driven by the task-oriented metric of localisation performance, it will record only the information necessary to maintain consistent accuracy across all locations. Once sufficient experiences are accumulated, the autonomous vehicle may be sufficiently confident to, for example, offer to take over control on certain sections of a frequently-traversed route.

In this paper we present a probabilistic experience-based approach to localisation with 2D push-broom LIDAR sensors. We adapt the approach of [14] and build locally metric 3D “swathes” using odometry information, eliminating our requirement for a 3D LIDAR. We use low-cost GPS observations as a “weak localiser” to find the most relevant experiences for the current swathe, then localise the swathe within the experience with a robust sample-based method. If the current swathe is not matched to any experience with sufficient accuracy, a new experience is created to capture more detail of the environment in difficult localisation conditions. We present a GPU implementation that can provide real-time localisation at 5Hz and pose estimates at 40Hz, and demonstrate large-scale localisation using over 50km of data collected in a dynamic city environment over a period of a year.

II. RELATED WORK

LIDAR-based localisation for autonomous vehicle applications has been addressed by several authors in the past decade. Early approaches using a traditional horizontal planar 2D LIDAR configuration [15], [16] produced impressive localisation performance at medium scales, but suffered from out-of-plane effects such as ground-strike. Later approaches [17] actuated the 2D laser to produce local 3D pointclouds for localisation, to avoid the limitations of a fixed planar LIDAR configuration.

The availability of 3D LIDAR sensors for the DARPA Urban Challenge [18] enabled larger-scale robust localisation approaches. In [19], a Velodyne LIDAR was used to build a 2D orthographic reflectivity map of road surfaces, and centimeter-accurate localisation was achieved using a particle filter. This approach was extended in [20] to incorporate reflectivity variance across multiple scans in a probabilistic framework. A similar approach presented in [21] used a 2D height-map representation for navigating in a multi-level parking garage.

To reduce the sensing cost of a LIDAR-based localisation system for autonomous vehicles, [14] presented an approach based on a 2D LIDAR in a push-broom configuration. By integrating vehicle odometry over a short window, multiple 2D laser scans were combined to form a 3D representation of path the vehicle has recently traversed. This local 3D map was compared to a global 3D map using a grid-based histogram approach, yielding centimetre-level accuracy and increased robustness in comparison to a GPS+INS system. This approach was extended in [22] to use 2D LIDARs for both vehicle odometry and localisation relative to a prior map.

The methods mentioned so far all localise relative to a single static global map, either incrementally constructed using a SLAM framework or offline using an optimisation approach. However, environments for autonomous vehicles contain highly dynamic objects (vehicles, pedestrians) as well as features that change gradually over time (parked vehicles, construction, seasonal changes), and a static global map will fail to capture this variation. Rather than attempting to combine multiple distinct representations of the same environment into a single static map, the experience-based framework of [13], [23] simply stores these representations, dubbed “experiences”, and attempts to localise relative to multiple experiences simultaneously. This approach produced impressive localisation performance over a 3-month period at different times of day and in different weather conditions, however still suffers from the limitations of stereo-camera-based approaches, namely strong dependence on scene illumination and a narrow convergence basin [24].

III. MAP AND SWATHE CONSTRUCTION

In this section we present our approach for local 3D pointcloud construction from a 2D LIDAR sensor for the purposes of localisation against a prior 3D map. We define a 2D LIDAR scan $s(t)$ at time $t$ as follows:

$$s(t) = \{d_1, \ldots, d_m, r_1, \ldots, r_m\}$$

where $d_i$ is the laser distance measurement (in metres) for beam $i$, and $r_i$ is the corresponding reflectance measurement in the infrared spectrum. $S = \{s(t_0), \ldots, s(t)\}$ describes a collection of such scans over the time period $[t_0, t]$. To compose the 2D LIDAR scans $S$ into a local 3D representation, we must estimate the trajectory of the vehicle over a short time period as follows:

$$\dot{x}(t) = \begin{bmatrix}
\cos\left(\int_{t_0}^{t} \dot{w}_z(t) \, dt\right) \\
\sin\left(\int_{t_0}^{t} \dot{w}_z(t) \, dt\right) \\
\sin\left(\int_{t_0}^{t} \dot{w}_x(t) \, dt\right)
\end{bmatrix}$$

where $\dot{v}(t), \dot{w}(t)$ are the translational and rotational velocities of the vehicle estimated using inertial sensors, wheel odometry or visual odometry. By integrating $\dot{x}(t)$ over the period $[t_0, t]$, we produce the estimated continuous $SE(3)$ pose $x(t)$. We then define the “swathe” $Q$ as the local 3D pointcloud produced by projecting 2D scans $S$ along the continuous-time trajectory $x(t)$ as follows:

$$Q_{t_0}^t = g(\dot{v}, \dot{w}, S)|_{t_0}^t$$
where the operator $g(.)$ defines the pointcloud projection function along the trajectory. Fig. 2 illustrates the process of building the 3D pointcloud from 2D scans along the vehicle trajectory $\hat{x}(t)$.

For localisation, we require a prior 3D pointcloud against which we can compare the swathe $Q_i$. In contrast to other methods [20], [11] which construct a globally consistent 3D map of the environment, we only require locally consistent maps to compare the swathe against. Hence, we can apply the same process as above to construct swathes using historical data. We denote each of these historical swathes as an individual map $m^i = (Q^i, z^i)$, where $z^i$ is the GPS observations recorded along the trajectory used to construct $Q^i$. The collection of all maps $M$ is as follows:

$$M = \{m^1 \ldots m^N\}$$

(4)

By storing multiple maps of the same environment gathered in successive traversals, we can simultaneously localise against multiple “experiences” of the environment under different conditions, increasing robustness at the cost of additional storage space. Although each map is not globally consistent, swathes built from short time periods $[t_0, t]$ are sufficient for relative localisation against prior trajectories, which enables teach-and-repeat [24] and relative-frame planning approaches.

IV. LOCALISATION

Given a swathe $Q_k$ built at the discrete update interval $k$ and a set of prior maps $M$, we can frame the process of estimating the pose of the vehicle $\hat{x}_k$ as a maximum-a-posteriori estimation problem as follows:

$$\hat{x}_k = \arg \max_{x_k} p(x_k | x_{k-1}, u_k, z_k, M, Q_k)$$

(5)

where $x_{k-1}$ is the pose of the vehicle at the previous interval $k-1$, $u_k$ is the motion experienced between update intervals and $z_k$ is the GPS observation (if available) at interval $k$. Applying Bayes’ rule yields the following relationship after removing unnecessary conditionals:

$$p(x_k | x_{k-1}, u_k, z_k, M, Q_k) \propto p(x_k | x_{k-1}, u_k, z_k) \prod_i p(Q_k | x^i_k, m^i)$$.

(6)

The first term $p(x_k | x_{k-1}, u_k, z_k)$ is the location prior, which models the likelihood of a pose using incremental odometry and GPS observations. The second term $p(Q_k | x^i_k, m^i)$ is the observation likelihood, which models the likelihood of producing the observed swathe $Q_k$ given a hypothesised pose $x^i_k$ within the prior map $m^i$. The product of the observation likelihood term for all $i$ permits the simultaneous comparison to multiple prior maps $m^i$. The two terms will be examined in detail below.

A. Location Prior

The location prior term from Equation (6) can be further expanded as follows:

$$p(x_k | x_{k-1}, u_k, z_k) \propto p(x_k | x_{k-1}, u_k) \prod_i p(z_k | x^i_k)$$

(7)

where $p(x_k | x_{k-1}, u_k)$ is the motion model of the vehicle and $p(z_k | x^i_{k-1})$ is the GPS observation likelihood for each previous pose $x^i_{k-1}$ in the local frame of map $m^i$ (which will be discussed further in the following section). The odometry update $u_k$ can be computed by integrating the continuous-time vehicle velocity from Equation (2) as follows:

$$u_k = \int_{t_{k-1}}^{t_k} \dot{x}(t) \, dt$$

(8)

where $t_{k-1}$ and $t_k$ are the timestamps at update intervals $k - 1$ and $k$ respectively.

B. Weak GPS Localisation

The GPS observation likelihood term in Equation (7) is not intended to provide centimetre-accurate location priors; rather, it is used as a weak localiser, providing information on which maps $m^i$ will be relevant for the current swathe $Q_k$. By making use of occasional, inaccurate GPS observations when they are available, we sidestep the need for loop closure algorithms to deal with global initialisation and the “kidnapped robot” problem, but equally we do not rely on GPS for accurate location estimates for path planning and vehicle behaviour.

As our maps are only locally metric, both localisation accuracy and map quality decrease as the distance between the current pose estimate in the map $x^i_k$ and the location...
of the GPS observation $z_k$ increases. To represent this, we approximate the GPS observation as a Gaussian with covariance matrix $\Sigma_z$, then further inflate the covariance based on the measurement as follows:

$$\hat{\Sigma}^i_{z_k} = \Sigma_z + \left[ z_k - h(x^i_{k-1}, m^i) \right] \left[ z_k - h(x^i_{k-1}, m^i) \right]^T$$

(9)

where $\hat{\Sigma}^i_{z_k}$ is the new covariance of the GPS location estimate for map $m^i$ and the function $h(x^i_{k-1}, m^i)$ produces the expected GPS observation at location $x^i_{k-1}$ in map $m^i$. Therefore, only maps $m^i$ containing prior GPS observations $z^i$ close to the current GPS observation $z_k$ will yield acceptably low location prior uncertainties, making them relevant for swathe-based localisation as discussed below.

C. Swathe Localisation

While the location prior and weak GPS localisation alone may yield an acceptable localisation solution for some applications, true centimetre-level accuracy can be obtained by making use of 3D map and swathe information. However, existing methods for comparing 3D point clouds are often computationally expensive [25], highly susceptible to initialisation noise and data association errors [26], [27], or a combination of the two. Instead, we follow the approach of [19], [21], [14] and project our 3D point cloud onto a 2D plane, yielding a 2.5D height and reflectance representation.

The x-y plane on which to perform the 2D projection can be extracted in a number of ways: by fitting a ground plane to laser data, transforming the point cloud relative to the local gravity vector using an inertial measurement unit or simply assuming the vehicle is locally horizontal. More important is the consistency of the method, such that the same ground plane is used each time the vehicle revisits a location.

We can factor the observation likelihood term from Equation 4 by sampling from the swathe $Q_k$ as follows:

$$p(Q_k | x^i_{k}, m^i) = \prod_j p(q_j | x^i_j, m^i)$$

(10)

where each sample $q_j = (h_j, r_j)$ and $h_j, r_j$ are the height and reflectance sampled from location $j$. The sample likelihood term $p(q_j | x^i_j, m^i)$ can then be directly computed in log-likelihood form as follows:

$$-\log \left [ p(q_j | x^i_j, m^i) \right ] \propto \frac{h_j - H_j(x^i_j, m^i)}{\sigma_h} + \frac{r_j - R_j(x^i_j, m^i)}{\sigma_r} \right ]^2$$

(11)

where the functions $H_j(\cdot)$ and $R_j(\cdot)$ produce the expected height and reflectance values at sample location $j$ in the map, and $\sigma_h, \sigma_r$ are the standard deviations of height and reflectance respectively, based on intrinsic sensor noise from the LIDAR scanner. Note that in contrast to methods that use either reflectance only [19], [20] or height only [21], [14], we combine both sources of information into a single cost function to improve robustness.

D. Localisation Covariance

While the localisation problem in Equation 5 could be solved using an optimisation-based approach to yield a maximum likelihood estimate for the vehicle pose $\hat{x}_k$, for autonomous driving applications it is often desirable to also know the uncertainty of the localisation estimate in order to influence vehicle behaviour (e.g., reducing speed in areas of high uncertainty, or reverting control to the driver of the vehicle). Inspired by [28], we further sample the observation likelihood term $p(Q_k | x^i_{k}, m^i)$ at a series of poses $x^i_{(j)}$ to produce a mean offset pose $\bar{x}_k$ and uncertainty $\hat{\Sigma}_k$ as follows:

$$K = \sum_j x^i_{(j)} x^i_{(j)}^T p(Q_k | x^i_{(j)}, m^i)$$

$$u = \sum_j x^i_{(j)} p(Q_k | x^i_{(j)}, m^i)$$

$$s = \sum_j p(Q_k | x^i_{(j)}, m^i)$$

$$\bar{x}_k = \frac{1}{s} u, \hat{\Sigma}_k = \frac{1}{s} K - \frac{1}{s^2} uu^T$$

(12)

The resulting covariance matrix $\hat{\Sigma}_k$ captures both the intrinsic noise of the sensor and swathe representation as well as the uncertainty in data association between swathe and map. This serves to provide a realistic estimate of the localisation uncertainty of the vehicle and a wide basin of convergence, but at the cost of increased computational complexity during sampling. The computational cost is addressed in Section V.

E. Map Update

Although the use of multiple maps allows us to robustly localise against multiple representations of the same environment under different conditions, there will come occasions when no map is sufficiently similar to the current swathe to provide an acceptable localisation estimate. Indeed, this will occur whenever the vehicle traverses a previously unmapped location. Rather than attempting to generalise from existing maps, we simply add the current swathe and GPS observations to the set of maps $\mathbf{M}$ according to the following condition:

$$\det(\hat{\Sigma}_k) > \rho_{\text{max}} : m^{n+1} \rightarrow (Q_k, z_k)$$

(13)

where $\hat{\Sigma}_k$ is the current localisation uncertainty and $\rho_{\text{max}}$ is a threshold for the maximum acceptable uncertainty. By dynamically adding live sensor data to the map only when localisation performance is insufficient, we are performing “experience-based” navigation [13]; this ensures we capture sufficiently rich representations of dynamic environments without map storage requirements that are linear with operation time.
V. GPU IMPLEMENTATION

While the sampling methods presented in Equations 10 and 12 produce high-quality likelihood and uncertainty estimates, they require a significant computational cost to evaluate. However, since the samples are independent, it is possible to exploit modern GPU processors to simultaneously evaluate a large number of observation likelihoods in parallel. In particular we exploit the mipmap reduction pipeline [29] to efficiently reduce large tiled cost images with as many as 2M pixels to low-resolution reduced cost images (typically 16 x 16 pixels) in under a millisecond, where each pixel represents the cost of a specific \((x, y, \theta)\) offset. The following steps outline the process of efficiently computing \(\mathbf{S}_k^t\) and \(\Sigma_k^t\):

1. Compute 3\(\sigma\) bounds for \(x, y, \theta\) from \(\Sigma_{k-1}\).
2. Convert the sparse map and swathe pointcloud to a dense mesh representation.
3. Project the swathe mesh at a series of orientations \(\theta_1 \ldots \theta_n\) covering the 3\(\sigma\) bound.
4. Rasterise the map and swathe to form 2D orthographic height/reflectance images.
5. Compute the likelihood \(p(q_j | x_k^t, m^t)\) for each pixel in the height/reflectance images for a series of index offsets \((x_1, y_1) \ldots (x_n, y_n)\) covering the 3\(\sigma\) bound, forming a tiled cost image.
6. Reduce the tiled cost image to a stacked reduced cost image using an efficient mipmap reduction.
7. Fit a mean and covariance to the tiled cost image as per Equation 12 to yield \(\mathbf{S}_k^t\) and \(\Sigma_k^t\).

This process is illustrated in Fig. 3. For the above process, implemented in OpenGL and running on a 2012 MacBook Pro with an Nvidia GT650M GPU, swathe localisation can be performed at 5Hz for each map in parallel, providing a corrected odometry signal at 40Hz.

VI. EXPERIMENTAL SETUP

In this section we present our experimental approach to demonstrating long-term localisation with an experience-based 2D push-broom LIDAR approach.

A. Experimental Data

Our experimental platform is the Oxford University Robotcar, an autonomous Nissan LEAF, depicted in Fig. 1. The LEAF is equipped with a SICK LMS-151 laser scanner in push-broom configuration mounted unobtrusively on the rear bumper. Vehicle odometry is provided by shaft encoders on each wheel, providing velocity estimates at 40Hz. Low-cost GPS observations were simulated by querying the NovAtel SPAN-CPT GPS+INS system for the raw, unfiltered GPS-only position estimates at 1Hz.

The experimental dataset consists of six traversals of a an approximately 8km route through central Oxford collected between July 2013 and August 2014, illustrated in the map section of Fig. 7. The traversals were made on public roads at different times of day (from 5am to 7pm) and in different traffic conditions. GPS signals were significantly degraded in areas of tree cover and narrow streets. There is some variation in the route over time, as different routes became available or were closed due to construction work.

Ground truth position estimates for experiments of this duration is challenging even for a tightly-coupled GPS+INS system, as noted in [20]. Instead we make use of a stereo camera mounted on the vehicle, and perform an offline multi-session pose graph optimisation combining visual odometry
and loop closures provided by FAB-MAP [30] to form a globally consistent ground-truth metric map. While this produces superior results to the GPS+INS system as illustrated in Fig. 4, it is an expensive offline process requiring long swaths and therefore is only suitable for generating ground-truth poses for benchmarking an online localisation algorithm.

B. Localisation Algorithm Details

For localisation, we use the push-broom SICK LMS-151 LIDAR in combination with wheel odometry (to estimate vehicle velocities \( \dot{v}(t), \dot{\varphi}(t) \)) and GPS observations derived from the SPAN-CPT. Table 1 lists the parameters used for the localisation algorithms:

### Table 1: Localisation Algorithm Parameters

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t - t_0 )</td>
<td>Swathe Period</td>
<td>10s</td>
</tr>
<tr>
<td>( \Sigma_{x_k} )</td>
<td>GPS Covariance</td>
<td>5m in ( x, y )</td>
</tr>
<tr>
<td>( \sigma_h )</td>
<td>Height Std Dev</td>
<td>0.5m</td>
</tr>
<tr>
<td>( \sigma_r )</td>
<td>Reflectance Std Dev</td>
<td>100</td>
</tr>
<tr>
<td>( n )</td>
<td>Number of ( x, y, \varphi ) offsets</td>
<td>16</td>
</tr>
<tr>
<td>( \rho_{max} )</td>
<td>Max Uncertainty Threshold</td>
<td>0.1</td>
</tr>
</tbody>
</table>

VII. RESULTS

In this section we examine the localisation performance and number of experiences recorded while driving in a changing city environment over a period of a year.

A. Localisation Accuracy

Of chief importance for autonomous driving is the accuracy of the localisation estimate. Fig. 5 shows the localisation error distribution relative to metric ground truth for the final traverse, demonstrating low longitudinal (0.38m), lateral (0.07m) and heading (0.43°) RMS errors. Additionally, the localisation performance of the GPS+INS system (with inertial filter) was evaluated relative to metric ground truth, illustrating that even high-cost inertial navigation systems are subject to gradual drift over long periods of time, rendering them unsuitable for long-term autonomy over multiple years.

For autonomous driving applications it also important that the localisation uncertainty estimate does not underestimate the true error, as this may lead to over-confident vehicle behaviour in uncertain environments. To evaluate the covariance estimates of the swathe-based localisation algorithm, we use the normalised estimation error squared (NEES) [31], which characterises the consistency of a state estimator. The NEES score \( \epsilon_k \) is computed as follows:

\[
\epsilon_k = (\hat{x}_k - x_k)^T \hat{\Sigma}_k^{-1} (\hat{x}_k - x_k)
\]

where \( \hat{x}_k \) and \( \hat{\Sigma}_k \) are the estimated location and uncertainty at update \( k \) and \( x_k \) is the true location from ground truth. Over all \( k \) the set of NEES scores \( \epsilon \) will follow a chi-squared distribution, and \( \hat{\Sigma} \) can be deemed a conservative estimate of the uncertainty if the following condition is satisfied:

\[
E[\epsilon] < \dim(x)
\]

For the SE(2) localisation problem, the expected NEES score \( E[\epsilon] \) must fall below the state vector dimension of 3 for it to yield conservative estimates of the uncertainty. Fig. 4 shows the distribution of NEES scores for the swathe localisation experiment. The expected NEES score of 1.13 indicates that the localisation uncertainty estimates provided by the filtering framework slightly overestimate the true error, and therefore are conservative.
Additionally, since the solutions from multiple maps are combined in the filtering framework, the RMS localisation uncertainty estimate remains conservative and does not become over-confident over time; a crucial requirement for long-term operation in dynamic environments.

VIII. CONCLUSIONS

Dealing with structural change in an environment is a critical requirement for long-term operation of autonomous vehicles in cities. The range of variation, from pedestrians, vehicles, parked cars, seasonal change and even construction of new buildings is something no static map approach could adequately represent. In this paper we demonstrated real-time, continuously-improving, centimetre-accurate localisation with low-cost 2D LIDAR sensors, with large-scale experiments spanning a year of operation in a dynamic city environment. We presented a probabilistic method of combining localisations from multiple prior maps to improve accuracy, and an efficient implementation that runs in real-time on commodity GPU hardware. We believe that learning from experience is the key to enabling true life-long autonomy for mobile robots in complex, changing cities.
REFERENCES


