LAPS-II: 6-DoF Day and Night Visual Localisation with Prior 3D Structure for Autonomous Road Vehicles

Will Maddern, Alexander D. Stewart and Paul Newman

***Abstract*— Robust and reliable visual localisation at any time of day is an essential component towards low-cost autonomy for road vehicles. We present a method to perform online 6-DoF visual localisation across a wide range of outdoor illumination conditions throughout the day and night using a 3D scene prior collected by a survey vehicle. We propose the use of a one-dimensional illumination invariant colour space which stems from modelling the spectral properties of the camera and scene illumination in conjunction. We combine our previous work on Localisation with Appearance of Prior Structure (LAPS) with this illumination invariant colour space to demonstrate a marked improvement in our ability to localise throughout the day compared to using a conventional RGB colour space. Our ultimate goal is robust and reliable any-time localisation - an attractive proposition for low-cost autonomy for road vehicles. Accordingly, we demonstrate our technique using 32km of data collected over a full 24-hour period from a road vehicle.**

I. Introduction

We want to build a low cost (around £100) 6-DoF navigation system for autonomous road vehicles using monocular cameras, which provides reliable metric localisation outdoors even under greatly varying lighting conditions including day and night. This is a challenging problem, and this paper describes our progress towards that goal.

We frame our task in the following way: first we use a survey vehicle equipped with a single pushbroom laser and colour cameras to make a coloured 3D point cloud "prior" of the environment. Then at run-time, we make this prior available to the economy vehicle to be localised which is equipped with only monocular cameras. We believe this is a viable model for mobile autonomy - costly surveys being leveraged by any number of users with only cheap sensors.

The way we exploit the survey can be simply put and it was described in its nascent form in [1]. Consider first, two images of a scene taken from nearby poses; the distribution of colours within them will be similar because they are pictures of the same things. Imagine now we place a camera at a known place relative to the survey. By the same argument, the projection of the survey points into the camera image should yield a distribution of colours similar to the survey. A simple mental picture is helpful here - "survey points of a bright red door should again project to a red door in the runtime image if the extrinsic parameters are right". We can now simply pose the localisation problem as maximising the similarity (we will finesse this term later) between runtime and survey appearance by varying the camera extrinsics - aka the camera location. Thus we are Localising using the Appearance of Prior Structure. We call this LAPS.

Authors are from the Mobile Robotics Group, Dept. Engineering Science, University of Oxford, UK. {wm,alexs,pnewman}@robots.ox.ac.uk

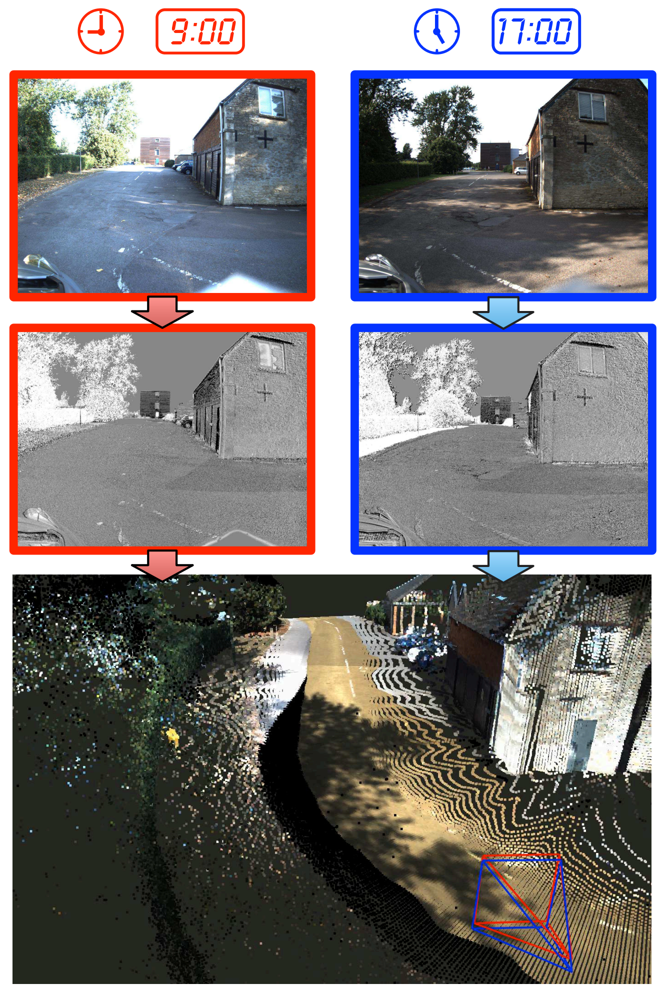

Fig. 1. Using illumination invariant images to localise in 6-DoF relative to a 3D pointcloud at different times of day. RGB images (top row) are converted to an illumination invariant colour space (middle row) using knowledge of the camera spectral response. This mostly removes variation due to sunlight and shadow. The 6-DoF position and viewpoint of the cameras can then be recovered relative to a coloured 3D point cloud (bottom, red/blue view frustrum) using illumination invariant appearance information alone. Note that what appears to be a shadow under the building is in fact a section of resurfaced tarmac, distinguished due to its material properties despite significant change in illumination.

A major challenge facing visual localisation and navigation approaches in outdoor environments is change in appearance across a wide range of illumination conditions, in particular those encountered during a typical 24-hour day-night cycle. One of the contributions of this paper, above showing LAPS working on 32km of outdoor data, is the way in which we deal with this challenge.

Robust and reliable localisation regardless of weather conditions and time of day is a critical requirement for vision-based autonomous road vehicles [2]. The computer vision community has provided much of the current work in visual feature and scene recognition techniques that are robust to changes in scene geometry and illumination. Visual navigation systems that build upon robust features such as SIFT [3] have produced impressive results in recent years [4]. However, as demonstrated in [5], these robust features do not provide true invariance to the illumination variation encountered in typical outdoor environments.

Much of the motivation behind scale and illumination invariant feature development comes from the field of large-scale image search and retrieval [6], where knowledge of the source image sensor properties is typically unavailable. In a robotics context, however, we can exploit full knowledge of the image sensor characteristics to infer true physical quantities about the scene. The process of inferring physical properties of objects from imagery is often referred to as passive remote sensing, and is a common in the field of satellite observation [7]. In a similar vein, research in the field of colour constancy [8] has produced a number of attempts to determine image features that identify the spectral properties of objects present in an image, regardless of the spectrum or intensity of the source illuminant. These approaches typically benefit from knowledge of the spectral properties of the camera, and often require a photometric calibration routine [9]. Colour constancy has been recently applied to the visual localisation problem in [10], [11] for the specific case of removing shadows from images, but not over a full day and night period.

In this paper we seek to exploit the full knowledge of our camera to improve visual localisation performance across a 24-hour period. We make use of a so-called "illumination invariant" colour space to minimise the variations in image sensor response caused by viewpoint and illumination conditions. We present simplifications to our prior work in visual localisation [1] to incorporate the invariant feature space and reduce computation time, and present visual localisation results on over 32km of data over a full 24-hour period.

II. Related Work

Recent successful approaches to robust image patch features [3], [12], [13] compute descriptors based on combinations of gradients and histograms of local greyscale patches around the point of interest. This results in a degree of invariance to absolute illumination levels. However, these techniques are not invariant to changes in the spectrum or direction of the light source, as these do not manifest as a simple uniform scaling of the mean greyscale intensities [14]. Approaches to incorporate colour histogram information into descriptors have produced limited improvements in performance [15].

An alternate approach to visual localisation under changing conditions is to store and maintain multiple representations of the same location. [16] presents an extension to a Visual Odometry (VO) pipeline which stores frame sequences as 'experiences', and continuously attempts to localise against prior trajectories while recording new experiences if necessary. Experience Based Navigation (EBN) demonstrated successful localisation over 37km of data collected over a period of three months, with significant illumination and weather condition variation. However, for the use case presented in [1] of a survey vehicle that infrequently revisits an environment, it is unlikely that a sufficient number of experiences will be collected to completely cover the variation in appearance of an environment; as EBN relies on the underlying data association provided by BRIEF descriptors [13] it is subject to the same limitations illustrated in [5] for appearance change over longer periods of time. Similar methods that attempt to model change in image descriptors over time include [17].

The importance of image sequence information when localising across changes in appearance is demonstrated by a number of approaches [18], [19]. In Seq-SLAM [19] a visual dataset collected on an 8km road network at midday is successfully matched to the same location at midnight during a rainstorm. Downscaling, patch normalisation and normalised cross-correlation are employed to reduce the variance due to scene illumination intensity and spectrum, but localisation performance is limited to topological place recognition. Further investigation in [20], [21] reveal a critical dependence on viewpoint direction when revisiting locations, making sequence-based approaches unsuitable as the sole localisation source for autonomous road vehicles.

Perhaps the work that is most similar to the approach presented in this paper is [22]. The authors combine a chrominance-space-based HDR imaging system [23] with a sparse 3D edge map of the environment, and perform 6-DoF localisation using a particle filter. Successful localisation was demonstrated during wide illumination changes in intervals between 7am and 5pm, and even during rain. An approach to generate the required sparse edge maps from 3D point cloud data was presented in [24], removing the requirement for a dedicated surveyor. While effective, this approach has to our knowledge only been demonstrated in urban environments with strong edges, and does not make use of appearance information to differentiate between objects in the scene.

III. Illumination Invariant Colour Space

In this section we summarise the illuminant invariant colour space presented in [25], [26] to improve the consistency of scene appearance over a range of outdoor illumination conditions. For a recent review on state-of-the-art approaches to illumination invariant imaging, otherwise known as colour constancy, we refer the reader to [8].

The following equation describes the relationship between the response of a linear image sensor R with spectral sensitivity $F(\lambda)$ to an illumination source with emitted spectral power distribution $E(\lambda)$ incident on an object with surface reflectivity $S(\lambda)$ [27]:

$$R^{x,E} = \underline{a}^x \cdot \underline{n}^x I^x \int S^x(\lambda) E^x(\lambda) F(\lambda) d\lambda \qquad (1)$$

where the unit vectors $\underline{a}^x$ and $\underline{n}^x$ represent the direction of the light source and the direction of the surface normal, and I^x represents the intensity of the illuminant on point x in the scene. From eq. 1 we wish to obtain an image feature $\mathcal{I}$ that depends on the material properties $S^x(\lambda)$ of the surface at point x, while minimising the effect of illumination source spectrum $E^x(\lambda)$ and intensity I^x. We follow the approach in [27] and assume the spectral sensitivity function $F(\lambda)$ can be modelled as a Dirac delta function centred on wavelength λ_i, which yields the following response function:

$$R^{x,E} = \underline{a}^x \cdot \underline{n}^x I^x S^x(\lambda_i) E^x(\lambda_i) \tag{2}$$

Although an infinitely narrow band spectral response assumption is unrealistic for most practical image sensors, results in [25] indicate colour constancy performance is maintained under this assumption with realistic 60-100nm full width at half-maximum (FWHM) sensor responses.

We take the logarithm of both sides of eq. 2 to separate the components as follows:

$$\log\left(R^{x,E}\right) = \log\left\{G^x I^x\right\} + \log\left\{S^x(\lambda_i)\right\} + \log\left\{E^x(\lambda_i)\right\} \tag{3}$$

where $G^x = \underline{a}^x \cdot \underline{n}^x$ is the relative geometry between illuminant and scene. This yields a linear combination of three components: a scene geometry and intensity component; an illuminant spectrum component; and a surface reflectance component. According to [27], for outdoor scenes illuminated by natural lighting it is reasonable to approximate the illuminant spectrum as a black-body source, and as such we can substitute the Wien approximation to a black-body source for the illuminant spectrum term in eq. 3:

$$\log(R_i) = \log\{GI\} + \log\left\{2hc^2\lambda_i^{-5}S_i\right\} - \frac{hc}{k_B T \lambda_i} \tag{4}$$

where h is Planck's constant, c is the speed of light, k_B is the Boltzmann constant and T is the correlated colour temperature of the black-body source. Note that for all references to the term "illumination invariant" in this paper, we are referring to a colour space that makes this assumption; that the source illuminant is approximately black-body. As demonstrated in [28], a dirac-delta sensor response and black-body source assumption provides good results for colour discrimination in outdoor scenes illuminated primarily by natural lighting. The first and third terms of eq. 4 can be eliminated by incorporating sensor responses at different wavelengths λ_i. In contrast with the approach proposed in [25], we use only a one-dimensional colour space $\mathcal{I}$ consisting of three sensor responses R_1, R_2, R_3 corresponding to peak sensitivities at ordered wavelengths $\lambda_1 < \lambda_2 < \lambda_3$:

$$\mathcal{I} = \log(R_2) - \alpha\log(R_1) - (1-\alpha)\log(R_3) \tag{5}$$

The one-dimensional colour space $\mathcal{I}$, conceptually similar to a robust greyscale space, will be independent of the correlated colour temperature T if the parameter α satisfies the following constraint:

$$\frac{hc}{k_B T\lambda_2} - \frac{\alpha hc}{k_B T\lambda_1} - \frac{(1-\alpha)hc}{k_B T\lambda_3} = 0 \tag{6}$$

which simplifies to

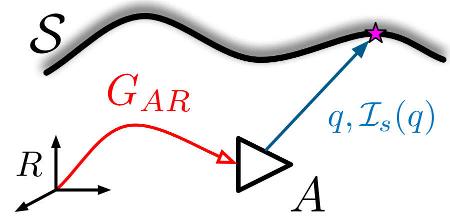

Fig. 2. A robot equipped with a camera observes the scene $\mathcal{S}$, defined in reference coordinate system R at position A at which it captures image $\mathcal{I}_A$. For each new image acquired, the transform G_{AR} is estimated by harmonising the appearance between image $\mathcal{I}_A$ and prior appearance $\mathcal{I}_\mathcal{S}$.

$$\frac{1}{\lambda_2} = \frac{\alpha}{\lambda_1} + \frac{(1-\alpha)}{\lambda_3}, \tag{7}$$

therefore we can uniquely determine α for a given three-channel camera simply with knowledge of the peak spectral responses of the Bayer filter. This formulation avoids the requirement for a set of training images in contrast to [11].

Note that a single illumination invariant feature is usually insufficient to uniquely identify a particular colour - as stated in [25] a minimum of four spectral responses is required to determine true colour. However, we propose that the one-dimensional space is sufficient to differentiate between most materials in natural scenes. An example of the resulting illumination invariant colour space is illustrated in Fig. 1. Despite large changes in sun angle, shadow pattern and illumination spectrum between images captured at 9am and 5pm, both illumination invariant images exhibit minimal variation, and even capture material differences in the road surface which are not immediately apparent from the raw images due to strong shadows. The use of only three spectral responses allows us to exploit commodity colour cameras with standard Bayer filters, reducing the cost of the localisation sensor suite.

IV. Localisation using Prior 3D Structure

In this section we present a simplified version of the 6-DoF localisation approach presented in [1] which incorporates the illumination invariant colour space representation.

For a robot at position A in the known 3D scene $\mathcal{S}$ with local co-ordinate frame R, we seek the transform G_{AR} using only a single illuminant invariant image $\mathcal{I}_A$ captured at position A, as illustrated in Fig. 2. We assume the known 3D scene $\mathcal{S}$ consists of a point cloud sampled by a survey vehicle, where each point $\boldsymbol{q} \in \mathbb{R}^3$ has an associated prior illumination invariant feature $\mathcal{I}_\mathcal{S}(\boldsymbol{q}) \in \mathbb{R}^1$ sampled at survey time.

The appearance $\mathcal{I}_A$ of a point $\boldsymbol{q}$ viewed from position A is found by reprojecting $\boldsymbol{q}$ onto the image plane $\boldsymbol{x}$ using the camera projection parameters $\boldsymbol{\kappa}$ as follows:

$$\boldsymbol{x}_A \equiv \mathcal{P}(\boldsymbol{q}, G_{AR}, \boldsymbol{\kappa}) \tag{8}$$

To recover the transform G_{AR} we seek to harmonise the information between the prior appearance $\mathcal{I}_\mathcal{S}$ and the appearance $\mathcal{I}_A$ as viewed from position A. We define an objective function f which measures the discrepancy between the visual appearance of the subset of points $\mathcal{S}_A$ from position A and the prior appearance of the points $\mathcal{I}_S$ as follows:

$$
\begin{aligned}
& f\left(\overbrace{\mathcal{I}_A(\mathcal{P}(\boldsymbol{q}, \widehat{G}_{AR}, \boldsymbol{\kappa}))}^{\text{Appearance of } \mathcal{S}_A \text{ from A}}, \quad \overbrace{\mathcal{I}_S(\boldsymbol{q})}^{\text{Prior appearance of } \mathcal{S}_A} \;\middle|\; \overbrace{\boldsymbol{q} \in \mathcal{S}_A}^{\text{Scene viewed by A}} \right) : \mathbb{R}^{2\times|\mathcal{S}_A|} \mapsto \mathbb{R}^1 \\
& \equiv f\left(\mathcal{I}_A(\boldsymbol{x}_A),\, \mathcal{I}_S(\boldsymbol{q}) \;\middle|\; \boldsymbol{q} \in \mathcal{S}_A\right) \qquad (9)
\end{aligned}
$$

As in [1] we choose Normalised Information Distance (NID) as the objective function, as it provides a true metric that is robust to local illumination change and occlusions. Given two discrete random variables $\{X, Y\}$, the NID is defined as follows:

$$
NID(X, Y) \equiv \frac{H(X, Y) - I(X; Y)}{H(X, Y)} \qquad (10)
$$

where $H(X, Y)$ denotes the joint entropy and $I(X; Y)$ denotes the mutual information. Substituting NID for our objective function from eq. 9 yields the following:

$$
f \equiv NID(\mathcal{I}_A(\boldsymbol{x}_A),\, \mathcal{I}_S(\boldsymbol{q}) \mid \boldsymbol{q} \in \mathcal{S}_A) \qquad (11)
$$

We can now frame the localisation problem as a minimisation of eq. 11 as follows:

$$
\widehat{G}_{AR} : \underset{G_{AR}}{\arg\min}\, NID(\mathcal{I}_A(\boldsymbol{x}_A),\, \mathcal{I}_S(\boldsymbol{q}) \mid \boldsymbol{q} \in \mathcal{S}_A) \qquad (12)
$$

The initial estimate $\widehat{G}_{AR}|_0$ can be set to the previous position of the sensor, or can incorporate incremental motion information provided by wheel encoders, visual odometry or another source.

In contrast to the original approach presented in [1], we do not require a sequential pair of images at locations A and B, nor a joint optimisation to solve for transforms $\widehat{G}_{AR}$ and $\widehat{G}_{BA}$. The use of a strong illumination invariant appearance prior $\mathcal{I}_S$ allows us to forego the need for an image sequence and joint optimisation (and thus significantly reduce computational requirements) without reducing localisation performance, as will be illustrated in Section VI.

V. Experimental Setup

In this section we present our experimental approach to demonstrating day and night localisation with monocular cameras and a 3D scene prior.

A. Experimental Data

The experimental dataset consists of 48 loops of a 670m urban environment, corresponding to 2 loops per hour across a period of 24 hours and a total distance of 32km. The experimental location, pictured in Fig. 3, is the site of many previous navigation experiments [1], [29], [16].

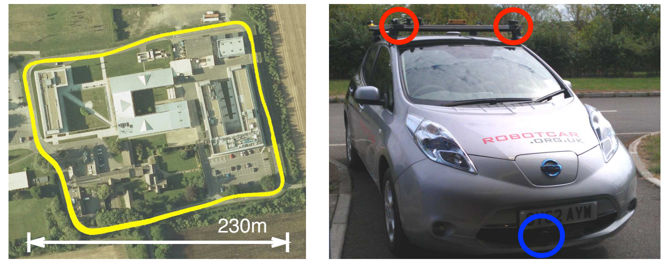

Fig. 3. Experimental test site and platform. The experimental data consists of 48 loops around an urban environment for a total distance of 32km over a period of 24 hours. The experimental platform is the Oxford University Robotcar, a modified Nissan LEAF. The Robotcar is equipped with a pair of Grasshopper2 cameras on either side of the roofrack for localisation (circled in red), and a vertical SICK LMS-151 mounted to the front bumper to build the 3D scene prior (circled in blue). Ground truth is provided by a NovAtel SPAN-CPT inertial navigation system.

Our experimental platform is the Oxford University Robotcar, an autonomous Nissan LEAF, depicted in Fig. 3. The LEAF is equipped with a vertical SICK LMS-151 laser scanner at the front and two side-facing Point Grey Grasshopper2 monocular cameras mounted to the roof. Odometry is provided by shaft encoders on each wheel, and ground truth position is provided by a NovAtel SPAN-CPT Align real-time-kinematic inertial navigation system (INS) mounted above the rear axle.

B. Localisation Algorithm Details

For localisation we use the two Point Grey Grasshopper2 colour CCD cameras mounted to the roof of the vehicle. The Sony ICX285AQ image sensor in the Grasshopper2 incorporates a Bayer filter with peak sensitivities at 470nm, 540nm and 620nm for B, G and R channels respectively [30], and therefore we use α = 0.4642 to form illumination invariant images from the three colour channels. The extrinsic calibration between cameras, lasers and the INS are recovered using the approach in [31].

The minimisation problem in eq. 12 is solved with the quasi-Newton BFGS method [32] implemented in Ceres [33] using the analytical derivatives presented in [1] obtained using B-spline interpolation. The cost function is implemented in the OpenCL[1] language and solved using an Nvidia GTX TITAN GPU, requiring approximately 3ms per evaluation, which enables real-time operation.

For each traverse of the test environment, LAPS is initialised using the first pose of the INS, but then performs successive localisations for each camera image regardless of the INS solution. This initial global localisation could be equally well performed using a visual feature-based loop closure algorithm such as [4].

For comparison with the original implementation in [1], we perform localisation with both greyscale images against a greyscale 3D prior along with illumination invariant images against an illumination invariant 3D prior.

C. Scene Prior

To investigate the importance of scene prior collection time for localisation over the period of a full day, we collect a scene prior at both midday (12:00) and midnight (24:00). The

[1] `http://www.khronos.org/opencl/`

3D scene prior is generated using a "push-broom" approach with the vertical 2D laser scanner as in [29]. RGB, greyscale and illumination invariant values for each point were derived by reprojecting the 3D location of each point at the time of capture into the sensor frame of the Grasshopper2 cameras mounted on the roof. The positions of the scene prior are registered to the global UTM frame provided by the INS to facilitate global localisation.

VI. Results

In this section we present the results of our 24-hour localisation experiment and demonstrate the effectiveness of the illumination invariant colour space across a wide range of illumination conditions.

A. Global Localisation

Fig. 4 presents global localisation results for LAPS using illumination invariant images in comparison to the INS positions. The relationship between the local structure of the environment and the consistency of the LAPS localisation solution can be examined with insets *A* to *F*. It is clear that LAPS provides localisation results that are consistent with the far more expensive NovAtel INS system.

Areas with features and obstacles close to the road (trees, fences and buildings) such as insets *B*, *D* and *E* provide the highest local consistency, since the local structure and appearance is sufficiently rich, yielding a clear minimum entropy solution to the NID cost function. Areas which are mostly planar, such as insets *A* and *C*, or with structure on only one side of the road, such as inset *F*, do not provide the same rich local structure and therefore the localisation consistency is correspondingly poorer.

B. Localisation in Daylight

Fig. 5 presents RGB and illumination invariant images during daylight hours along with error statistics for global localisation using greyscale and illumination invariant images. It is clear that while the raw RGB images vary significantly between 8am and 7pm, the illumination invariant colour space remains mostly immune to changes in illumination and shadows. Using the illumination invariant colour space during daylight hours, LAPS localises with mean absolute translational positions error under 0.4m in each axis, and mean absolute orientation errors under 1.8 degrees about each axis.

While the illumination invariant images provide better stability over time, they exhibit a smaller dynamic range than an equivalent greyscale image. This will serve to compact the histograms used to compute the NID value and therefore may reduce the sensitivity of the solution to small changes - this would explain the translational error distribution, where small changes in position have less of an effect on the cost function as small changes in orientation. However, over the course of the day LAPS with illumination invariant images provides equal or lower translational error in comparison to LAPS with greyscale images, and significantly lower orientation error.

Fig. 7. Sample sections from the coloured 3D point cloud scene priors built using the midday and midnight datasets for the same physical location, showing the dramatic change in the appearance of the scene between day and night. The blue-white streaks in the verges off the road in the midnight prior are due to the LEAF's headlights, and the orange orb at the top of the building is a streetlight. Note that although the pointcloud is coloured here with RGB for ease of presentation, we store both RGB and illumination invariant appearances for each point.

C. Localisation at Night

As the test site is illuminated at night by narrow-spectrum sodium and mercury streetlights, the primary assumption that the source illuminant can be modelled as a black-body radiator is violated. This leads to poor localisation performance when using the illumination invariant colour space, since the appearance no longer matches the illumination invariant prior captured during the day. Fig. 6 illustrates the poor quality of the illumination invariant images captured at night, in comparison to one captured during the day.

However, we make the observation that the absolute illumination levels do not vary significantly over the course of the night, and therefore a greyscale 3D prior captured after nightfall would provide a stable appearance reference for localisation after dark with artificial lighting. Fig. 7 illustrates the difference between the appearance of the 3D prior collected at midday and at midnight at a location in the dataset.

Using greyscale images and a 3D prior captured at midnight, LAPS localises with mean absolute translational position error under 0.5m in each axis, and mean absolute orientation error under 2.8 degrees about each axis, as shown in Fig. 6. These results approach the localisation accuracy during the day with an illumination invariant prior - this implies that for reliable localisation at any time of day, only one illumination-invariant daytime prior and one greyscale nighttime prior are necessary.

VII. Conclusions

In this paper we presented successful 24-hour 6-DoF localisation using mono cameras, accurate to a fraction of a meter and a few degrees over 32km of driving, both during daylight hours (with an illumination invariant 3D prior captured at

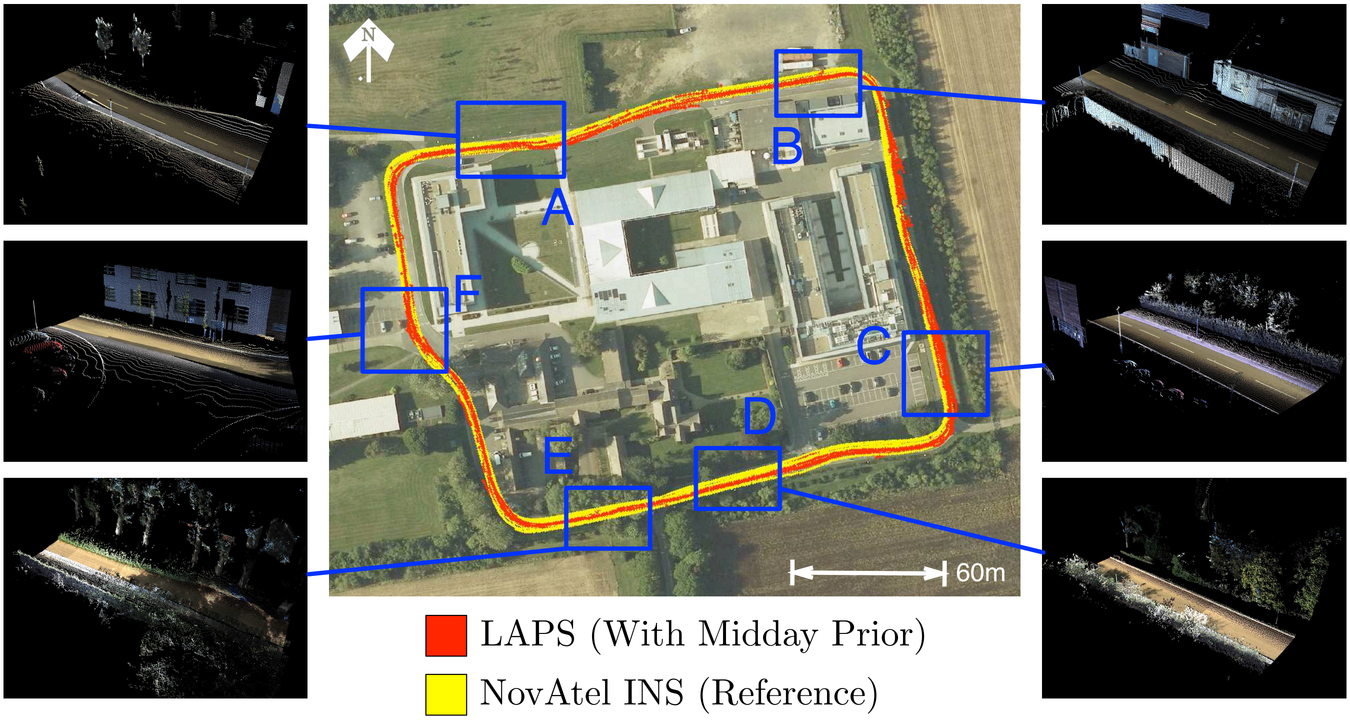

Fig. 4. Overhead view of the test site with with the estimated positions from the reference NovAtel INS, and the illumination invariant LAPS algorithm detailed in this paper, for the full 24 hour period. First consider insets *E* & *D*: LAPS is notably more consistent in these areas due to the confining 3D structure of the scene, whilst this confinement is detrimental to the INS as it occludes the sky. This relationship is also evident in inset *B*. The corollary to this, that in open areas with minimal confining structure, or in areas poorly and/or unevenly covered by the prior, where the path of the vehicle results in a camera viewpoint that consists of only a single plane, the performance of LAPS degrades. This can be seen in insets *A, C* & *F* respectively.

midday) and during the night (with a greyscale prior captured at midnight). By exploiting knowledge about the physical properties of the camera and outdoor illumination, we extend the usefulness of an expensive survey beyond mere minutes or hours to an entire day or night cycle, reducing the number of surveys (and therefore cost) required for a given environment. We have presented a simplified version of an illumination invariant colour space, which is applicable to consumer RGB colour cameras with Bayer filters, combined with a computationally inexpensive variant of LAPS, which provides real-time visual localisation using consumer graphics hardware. Coupled with the use of inexpensive monocular cameras, we believe these results pave the way to ubiquitous low-cost autonomy for road vehicles in the future.

A. Future Work

We are currently integrating LAPS-II into a sliding window filtering framework with vehicle odometry provided by encoders on each wheel of the LEAF, to provide a high-rate smooth pose solution suitable for closed-loop vision-based autonomous driving at all hours of the day.

VIII. Acknowledgements

This work was generously supported by EPSRC grants supporting Will Maddern and Alex Stewart and an EPSRC Leadership Fellowship EP/I005021/1 supporting Paul Newman. We would also like to thank Steve Collins for his assistance with colour constancy and camera calibration, Chris Prahacs for maintaining and driving the robotic platform and Peter Corke for setting our thinking about colour in motion in 2012.

References

[1] A. D. Stewart and P. Newman, "Laps-localisation using appearance of prior structure: 6-dof monocular camera localisation using prior pointclouds," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 2625–2632.

[2] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 3354–3361.

[3] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2. Ieee, 1999, pp. 1150–1157.

[4] M. Cummins and P. Newman, "Appearance-only SLAM at large scale with FAB-MAP 2.0," *The International Journal of Robotics Research*, vol. 30, no. 9, pp. 1100–1123, 2011.

[5] A. J. Glover, W. P. Maddern, M. J. Milford, and G. F. Wyeth, "FAB-MAP+ RatSLAM: appearance-based SLAM for multiple times of day," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 3507–3512.

[6] H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *Computer Vision–ECCV 2008*. Springer, 2008, pp. 304–317.

[7] T. M. Lillesand, R. W. Kiefer, J. W. Chipman, *et al.*, *Remote sensing and image interpretation.* John Wiley & Sons Ltd, 2004, no. Ed. 5.

[8] D. H. Foster, "Color constancy," *Vision research*, vol. 51, no. 7, pp. 674–700, 2011.

[9] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data," *Image Processing, IEEE Transactions on*, vol. 11, no. 9, pp. 972–984, 2002.

[10] J. M. Alvarez, A. López, and R. Baldrich, "Illuminant-invariant model-based road segmentation," in *Intelligent Vehicles Symposium, 2008 IEEE*. IEEE, 2008, pp. 1175–1180.

[11] P. Corke, R. Paul, W. Churchill, and P. Newman, "Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 2085–2092.

[12] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer Vision–ECCV 2006*. Springer, 2006, pp. 404–417.

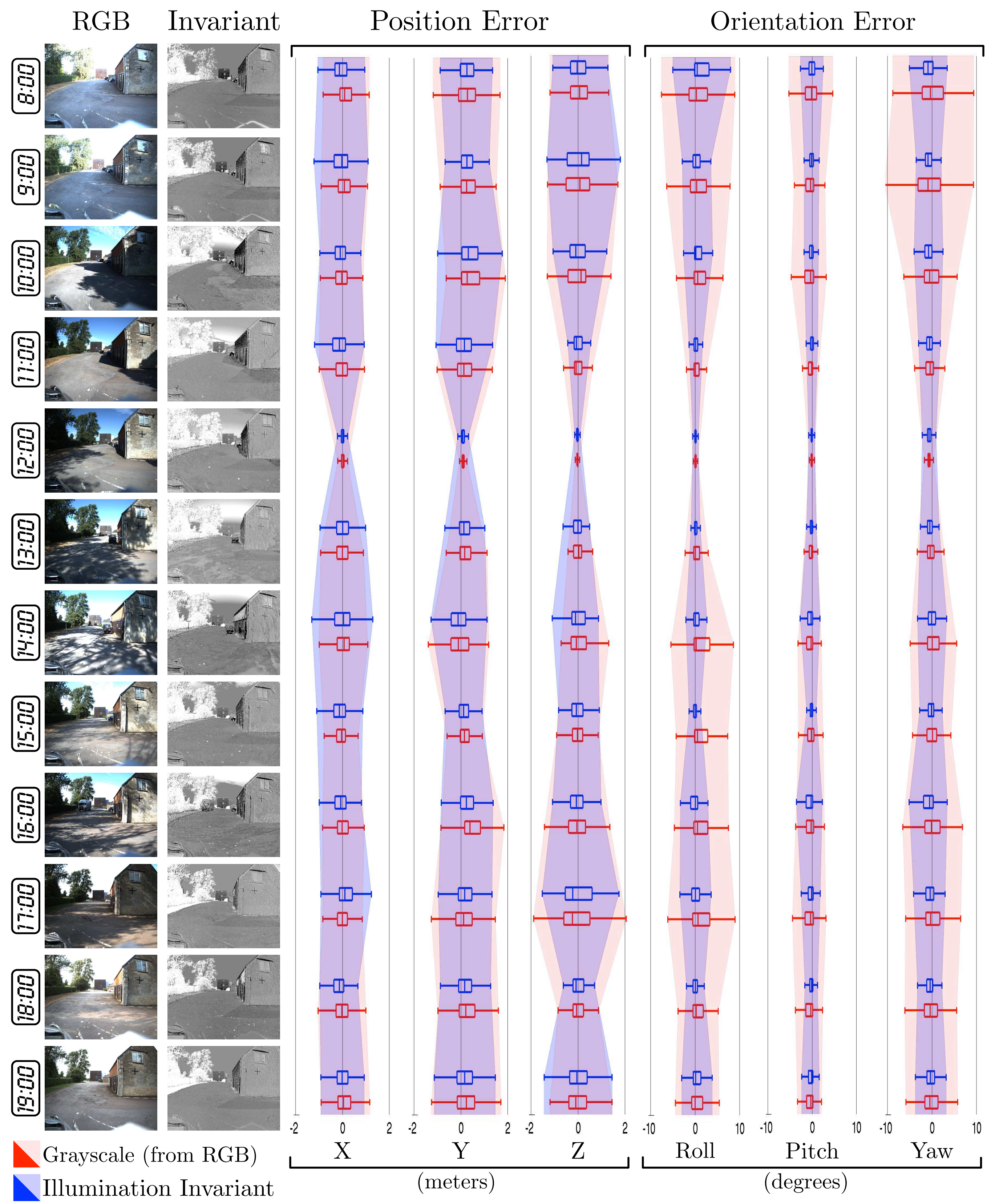

Fig. 5. LAPS localisation error analysis against a midday 3D scene prior during daylight hours (8am to 7pm) for both greyscale and illumination invariant images. For a typical location in the test environment, the raw RGB images vary significantly in appearance over the course of the day, but the illumination invariant colour space remains mostly static. Position and orientation error for both greyscale (red) and illumination invariant (blue) images decrease when temporally closer to the midday prior, however the illumination invariant images provide significantly lower orientation error over the course of the day. As expected, localising relative to the midday prior at midday provides the lowest error for both images. All errors are reported relative to the INS solution.

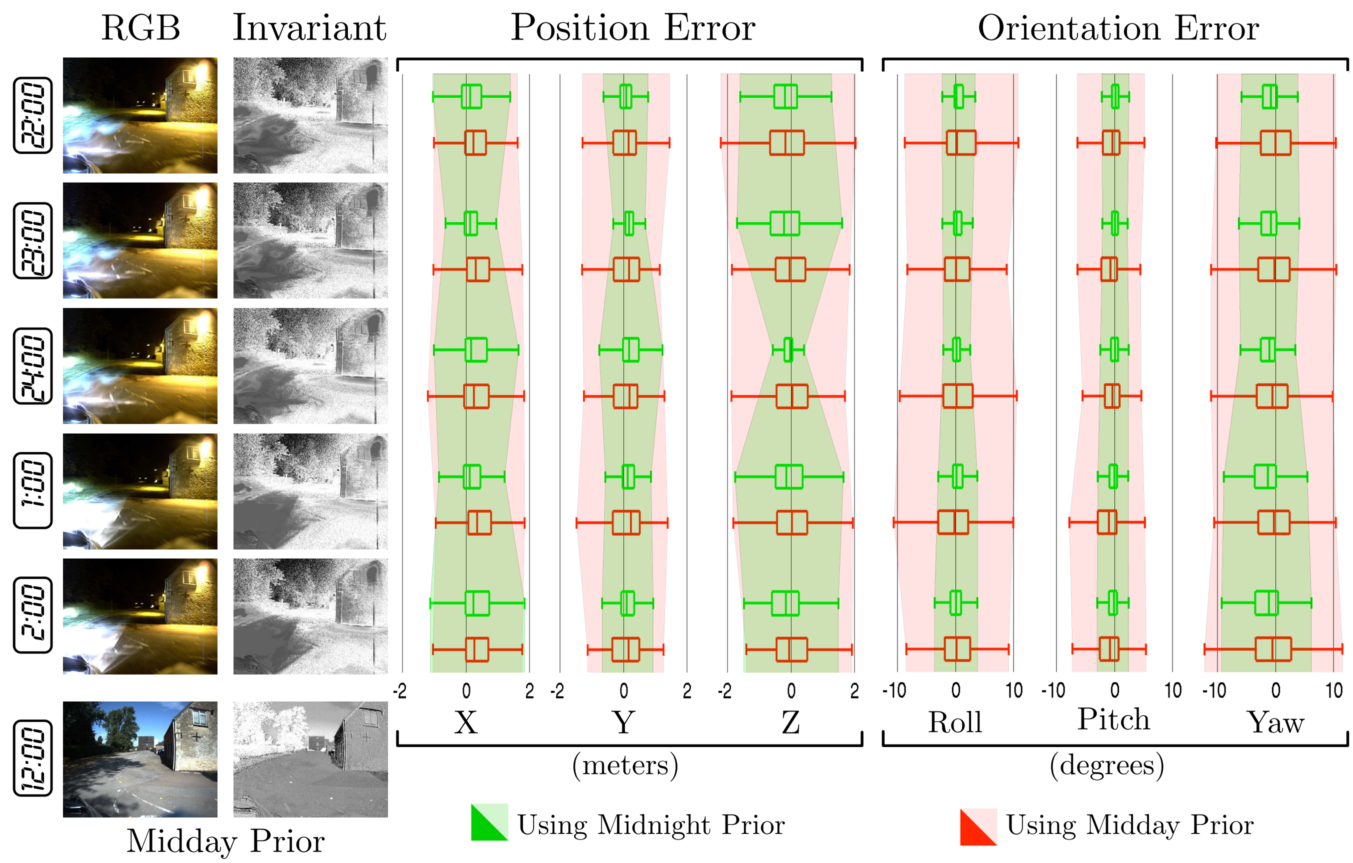

Fig. 6. LAPS localisation error analysis during the night (8pm to 7am; examples from 10pm to 2am shown) using a 3D scene prior gathered at midday compared to a prior gathered at midnight. For the same location in the test environment, the raw RGB images do not vary significantly over the course of the night, but the illumination invariant colour space yields very poor results due to non-black-body sodium lamp and LED headlight illumination, and do not resemble images captured in daylight (see Midday Prior images for reference). Due to the robustness of the NID metric, LAPS is still capable of localising at night relative to the midday prior using greyscale images, but with reduced accuracy. Using greyscale images and a midnight prior, localisation performance approaches that achieved with illuminant invariant images during the day. All errors are reported relative to the INS solution.

[13] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *Computer Vision–ECCV 2010*. Springer, 2010, pp. 778–792.

[14] C. Valgren and A. J. Lilienthal, "SIFT, SURF and seasons: Long-term outdoor localization using local features." in *EMCR*, 2007.

[15] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: A SIFT descriptor with color invariant characteristics," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2. IEEE, 2006, pp. 1978–1983.

[16] W. Churchill and P. Newman, "Practice makes perfect? managing and leveraging visual experiences for lifelong navigation," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 4525–4532.

[17] A. Ranganathan, S. Matsumoto, and D. Ilstrup, "Towards illumination invariance for visual localization," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013.

[18] K. L. Ho and P. Newman, "Detecting loop closure with scene sequences," *International Journal of Computer Vision*, vol. 74, no. 3, pp. 261–286, 2007.

[19] M. J. Milford and G. F. Wyeth, "Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1643–1649.

[20] P. Furgale and T. D. Barfoot, "Visual teach and repeat for long-range rover autonomy," *Journal of Field Robotics*, vol. 27, no. 5, pp. 534–560, 2010.

[21] N. Sünderhauf, P. Neubert, and P. Protzel, "Are we there yet? challenging SeqSLAM on a 3000 km journey across all four seasons," in *Proc. of Workshop on Long-Term Autonomy, IEEE International Conference on Robotics and Automation (ICRA)*, 2013.

[22] S. Nuske, J. Roberts, and G. Wyeth, "Robust outdoor visual localization using a three-dimensional-edge map," *Journal of Field Robotics*, vol. 26, no. 9, pp. 728–756, 2009.

[23] ——, "Extending the dynamic range of robotic vision," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*. IEEE, 2006, pp. 162–167.

[24] P. Borges, R. Zlot, M. Bosse, S. Nuske, and A. Tews, "Vision-based localization using an edge map extracted from 3d laser range data," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 4902–4909.

[25] S. Ratnasingam and S. Collins, "Study of the photodetector characteristics of a camera for color constancy in natural scenes," *JOSA A*, vol. 27, no. 2, pp. 286–294, 2010.

[26] W. Maddern, A. D. Stewart, C. McManus, B. Upcroft, W. Churchill, and P. Newman, "Illumination invariant imaging: Applications in robust vision-based localisation, mapping and classification for autonomous vehicles," in *Proc. of Workshop on Visual Place Recognition in Changing Environments, IEEE International Conference on Robotics and Automation (ICRA)*, 2014.

[27] G. D. Finlayson and S. D. Hordley, "Color constancy at a pixel," *JOSA A*, vol. 18, no. 2, pp. 253–264, 2001.

[28] S. Ratnasingam and T. M. McGinnity, "Chromaticity space for illuminant invariant recognition," *Image Processing, IEEE Transactions on*, vol. 21, no. 8, pp. 3612–3623, 2012.

[29] I. Baldwin and P. Newman, "Road vehicle localization with 2d push-broom lidar and 3d priors," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 2611–2617.

[30] "Grasshopper2 GS2-FW technical reference manual," Point Grey Research, 2011.

[31] A. Napier, P. Corke, and P. Newman, "Cross-calbration of push-broom 2d lidars and cameras in natural scenes," in *Proceedings of 2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013.

[32] N. Jorge and J. W. Stephen, "Numerical optimization," *Springerverlang, USA*, 1999.

[33] S. Agarwal, K. Mierle, and Others, "Ceres solver," https://code.google.com/p/ceres-solver/.