Illumination Invariant Imaging: Applications in Robust Vision-based Localisation, Mapping and Classification for Autonomous Vehicles

Will Maddern[1], Alexander D. Stewart[1], Colin McManus[1], Ben Upcroft[2], Winston Churchill[1] and Paul Newman[1]

Abstract—In this paper we propose the use of an illumination invariant transform to improve many aspects of visual localisation, mapping and scene classification for autonomous road vehicles. The illumination invariant colour space stems from modelling the spectral properties of the camera and scene illumination in conjunction, and requires only a single parameter derived from the image sensor specifications. We present results using a 24-hour dataset collected using an autonomous road vehicle, demonstrating increased consistency of the illumination invariant images in comparison to raw RGB images during daylight hours. We then present three example applications of how illumination invariant imaging can improve performance in the context of vision-based autonomous vehicles: 6-DoF metric localisation using monocular cameras over a 24-hour period, life-long visual localisation and mapping using stereo, and urban scene classification in changing environments. Our ultimate goal is robust and reliable vision-based perception and navigation - an attractive proposition for low-cost autonomy for road vehicles.

I. INTRODUCTION

Robust and reliable operation regardless of weather conditions and time of day is a critical requirement for vision-based autonomous road vehicles [1]. A major challenge facing visual localisation, navigation and scene classification approaches in outdoor environments is the change in appearance across a wide range of illumination conditions, in particular those encountered during a typical 24-hour day-night cycle. Visual navigation systems that build upon robust features such as SIFT [2] have produced impressive results in recent years [3]; however, as demonstrated in [4], 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 [5], 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 [6]. In a similar vein, research in the field of colour constancy [7] has produced a number of attempts to determine image features that identify the spectral

Fig. 1. Using an illumination invariant colour space to improve image similarity at different times of day. RGB images (top row) are converted to an illumination invariant colour space (bottom row) using knowledge of the camera spectral response. This significantly reduces variation due to sunlight and shadow, yielding a greyscale image where grey values depend primarily on the material properties of the objects in the scene. 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 spectrum and shadow.

properties of objects present in an image, regardless of the spectrum or intensity of the source illuminant.

In this paper we seek to exploit full knowledge of the spectral properties of the cameras on an autonomous vehicle platform, and present an "illumination invariant" colour space that reduces the effects of sunlight and shadow present in raw RGB images, as illustrated in Fig. 1. The next section presents related approaches, and Section III summarises relevant portions of the colour constancy literature to derive the illumination invariant transform. The illumination invariant images are evaluated over a 24-hour period in Section IV, and results are presented for example applications of metric monocular localisation in Section V, life-long stereo localisation in Section VI, and urban scene classification in Section VII. Section VIII concludes the paper.

II. RELATED WORK

Robust image-point features such as SIFT [2], SURF [8] and BRIEF [9] 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

[1]Will Maddern, Alex Stewart, Colin McManus, Winston Churchill and Paul Newman are with the Mobile Robotics Group, University of Oxford, UK. {wm,alexs,colin,winston,pnewman}@robots.ox.ac.uk
[2]Ben Upcroft is with the School of Electrical Engineering and Computer Science, Queensland University of Technology, Australia. ben.upcroft@qut.edu.au

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 [10]. The "Leuven/Light" benchmark dataset for point feature illumination invariance [11] consists only of images taken from the same location with varying exposure, which does not represent true variation in outdoor illumination due to sunlight and shadow.

Approaches to incorporate colour histogram information into descriptors have produced limited improvements in performance [12]. Approaches to learning true illumination invariant point features have been presented [13], but these rely on training images that capture the full range of illumination variation in the environment. Successful point-feature-based visual place recognition over a 24 hour period was accomplished using a thermal imaging camera in [14] and an imaging LIDAR sensor in [15], but both of these approaches require expensive specialist hardware.

Recently, strong progress in visual localisation under challenging conditions has been demonstrated by whole-image sequence-based approaches [16]. In Seq-SLAM [17] a visual dataset collected on an 8 km road network at midday is successfully matched to the same location at midnight during a rainstorm. However, localisation performance is limited to topological place recognition, and is critically dependent on viewpoint direction when revisiting locations [18], making sequence-based approaches unsuitable as the sole localisation source for autonomous road vehicles.

Perhaps the work that is the most similar to the approaches presented in this paper is [19] and [20]. Both make use of the single-channel illumination invariant imaging approach presented in [21] for the purposes of removing shadows from images taken from a road vehicle. However, both approaches require the collection of training data from the test environment in order to calibrate the spectral properties of the camera.

III. ILLUMINATION INVARIANT COLOUR SPACE

In this section we present an illuminant invariant colour space adapted from [22], [23] to improve the consistency of scene appearance over a range of outdoor illumination conditions. Fig. 2 and the following equation describe 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 $S(\lambda)$ [24]:

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

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 [24] 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:
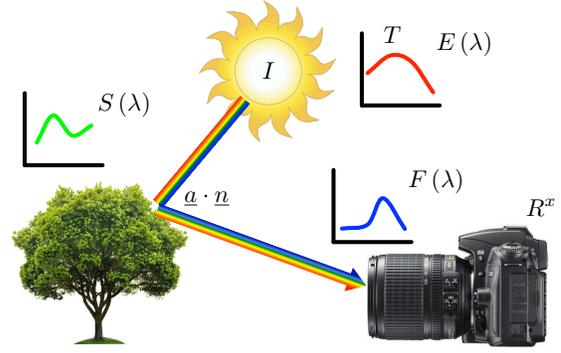

Fig. 2. Illustration of image sensor response in an outdoor scene. During daylight hours, scene lighting typically consists of sunlight or skydome illumination of intensity I and spectral power distribution $E(\lambda)$, which in this paper is modelled as a black-body illuminant of unknown correlated colour temperature T. Objects in the scene reflect light with unknown surface reflectivity $S(\lambda)$ and geometry term $\underline{a} \cdot \underline{n}$, which depend on the material properties of the object and the relative angle of the light source and image sensor. The reflected light is incident upon the image sensor, which produces response R^x with spectral sensitivity $F(\lambda)$ at location x on the imaging plane. The challenge of colour constancy is to produce a colour space $\mathcal{I}$ that depends only on object material properties $S(\lambda)$ using a combination of sensor responses R_i^x with known spectral sensitivities $F_i(\lambda)$, without explicit knowledge of light source intensity, spectrum or geometry.

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

Although an infinitely narrow-band spectral response assumption is unrealistic for most practical image sensors, results in [22] indicate colour constancy performance is maintained under this assumption with realistic 60-100 nm full-width at half-maximum (FWHM) sensor responses, and a spectral sharpening process such as [25] can improve seperation between multiple overlapping sensor responses.

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

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

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.

Several long-term outdoor spectrophotometry experiments [26], [27] have demonstrated that natural lighting closely follows the Planckian locus in the visible spectrum, with correlated colour temperatures (CCT) ranging between 4,000K and 25,000K, which covers the full range of atmospheric conditions from deep sunset to bright sun to cloud cover to blue skydome. Therefore we follow [22] and approximate the illuminant spectrum as a Planckian source, and as such we can substitute the Wien approximation to a Planckian source into eq. 3:

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

where h is Planck's constant, c is the speed of light, k_B is the Boltzmann constant and T is the CCT of the black-body source. 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 [22], 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 $\lambda_1 < \lambda_2 < \lambda_3$:

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

Note that a single illumination invariant channel is usually insufficient to uniquely identify a particular colour - as stated in [22] 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. By substituting eq. 4 into eq. 5, we can show the one-dimensional colour space $\mathcal{I}$ will be independent of the correlated colour temperature T if the parameter α satisfies the following constraint:

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

which simplifies to

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

Therefore we can uniquely determine α for a given three-channel camera simply with knowledge of the peak spectral responses of each sensor channel. The use of only three spectral responses allows us to exploit commodity colour cameras with standard Bayer filters, reducing the cost of the sensor suite. By estimating these peaks using the data sheet for the Bayer filter of the camera sensor, we avoid the requirement for a set of training images, in contrast to [19].

Table I lists the peak spectral responses estimated from the datasheets of the three Point Grey cameras used in this paper along with the corresponding α parameter. Algorithm 1 shows the single line of MATLAB code required to convert a 3-channel floating-point RGB image into the corresponding illumination invariant image using the α parameter. An offset of 0.5 is applied to bring responses into the [0, 1] range required for MATLAB images.

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.

IV. 24-HOUR ILLUMINATION INVARIANT IMAGING

To evaluate the performance of the illumination invariant colour space, we collected a 24-hour visual dataset consisting of 24 670m loops of an urban testing environment (one loop for each hour of the day). The experimental test site and platform are pictured in Fig. 3. Each loop was divided

TABLE I
CHOICE OF α PARAMETER FOR POINT GREY CAMERAS WITH SONY IMAGE SENSORS

Camera	Image Sensor	λ_1	λ_2	λ_3	α
Grasshopper2	ICX285AQ	470nm	540nm	620nm	0.4642
Bumblebee2	ICX204AK	460nm	540nm	610nm	0.3975
Flea2	ICX267AK	470nm	535nm	610nm	0.4706

Algorithm 1 MATLAB code for 3-channel floating-point RGB image to Illumination Invariant image conversion.

```
1  function [ ii_image ] = ...
     RGB2IlluminationInvariant( image, alpha )
2  ii_image = 0.5 + log(image(:,:,2)) - ...
     alpha*log(image(:,:,3)) - ...
     (1-alpha)*log(image(:,:,1));
```

into 71 distinct locations, which were spatially aligned across datasets using a NovAtel SPAN-CPT GPS-aided INS mounted to the vehicle, forming 24 images of each location from near-identical viewpoints for each hour over the 24-hour period. Fig. 4 shows all 24 images from one location over the 24-hour period, using both the raw RGB colour space as well as the illumination invariant colour space.

To quantitatively evaluate the consistency of the illumination invariant images over the 24-hour period, we adopt the approach used in [19] and examine the zero-mean normalised cross correlation (ZNCC) between images collected from the same place at different times. Images from the left and right Grasshopper2 cameras were downsampled to reduce viewpoint dependence and concatenated together, forming a 278 x 104 pixel resolution image for each location at each time of day. The mean ZNCC between all images at the same location at different times of day was computed for both RGB images (on a per-channel basis) and for illumination invariant images, and the results are shown in Fig. 5. Illumination invariant images are notably more consistent during daylight hours, as evident in the block-like correlation structure between 07:00 and 20:00. However, raw
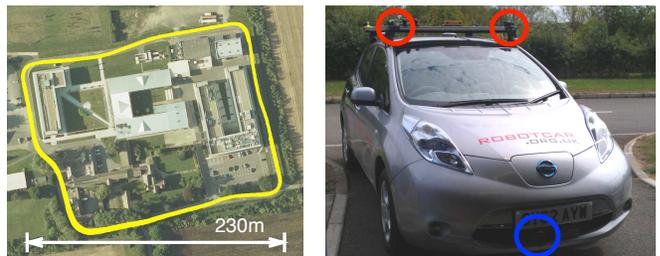

Fig. 3. Experimental test site and platform. The experimental data consists of 24 loops around an urban environment for a total distance of 16km over a period of 24 hours, with images captured at 71 distinct locations along the route. The experimental platform is the Oxford University Robotcar, a modified 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 (circled in blue) to build the 3D scene prior for metric localisation. Ground truth for localisation is provided by a NovAtel SPAN-CPT inertial navigation system.

(a) RGB images at one location for every hour of a 24-hour period.
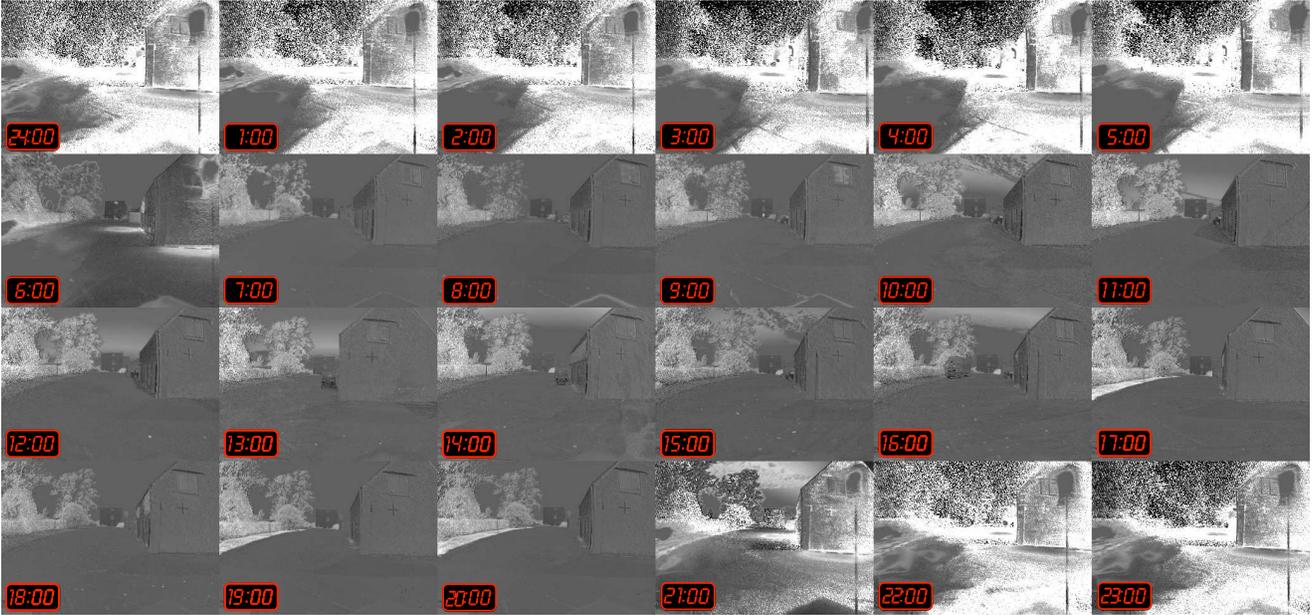
(b) Illumination invariant images at one location for every hour of a 24-hour period.

Fig. 4. Time lapse montage of images captured using a Grasshopper2 camera at one of the 71 locations for every hour of a 24-hour period. Raw RGB images (a) exhibit significant variation during daylight hours (from 7:00 to 20:00) due to changes in sunlight intensity, direction and spectrum. Illumination invariant images (b) are notably more consistant over this time period, and the material properties of different types of objects (road surfaces, vegetation, stone, bricks, wooden surfaces, metal vehicles) lead to distinct grey level values in the images. The major source of artefacts in the illumination invariant images during daylight hours is under- or over-exposed pixels in the raw RGB images, particularly at boundaries between bright sunlight and deep shadow; these artefacts are easily removed by only considering pixels for which all three sensor responses lie within the linear detection region of the camera. At night (22:00 to 5:00) the scene lighting is dominated by sodium-vapour streetlights and white LED headlamps, neither of which are black-body sources, and therefore the illumination invariant images yield poor results. However, since the artificial illumination at night does not vary over time, the raw RGB images remain consistent over this period.

RGB images are more consistent at night when the necessary assumptions of black-body illumination are violated. Therefore, the best results for visual localisation over 24 hours would be obtained using a combined system of illumination invariant images during daylight hours and raw RGB images at night.

In addition to evaluating illumination invariant image consistency over 24 hours, an exhaustive search was performed to determine the optimal value of α for the Grasshopper2 camera, by varying the parameter to maximise the mean ZNCC during daylight hours. This resulted in an optimal α value of approximately 0.49, which is within 10% of the value predicted by simply reading the peak spectral response values for each channel from the image sensor datasheet. In contrast to [19] which required manually-annotated training

images of materials in shadows and a spectral sharpening optimisation, our approach yields close-to-optimal illumination invariant images using only a single parameter derived directly from the image sensor specifications.

V. 24-Hour Metric Monocular Localisation

An application that benefits directly from the increased consistency of illumination invariant images during daylight hours is the 6-DoF metric monocular localisation presented in [28]. This approach builds upon the localisation using appearance of 3D prior structure (LAPS) presented in [29], and incorporates the illumination invariant colour space representation during daylight hours.

The metric localisation process is as follows: a survey vehicle collects a 3D scene prior $\mathcal{S}$ with local coordinate frame R consisting of a point cloud, where each point $\boldsymbol{q} \in \mathbb{R}^3$ has an associated prior RGB and illumination invariant value $\mathcal{I}_\mathcal{S}(\boldsymbol{q}) \in \mathbb{R}^1$ sampled at survey time. At runtime, the vehicle at position A captures an RGB or illumination invariant image $\mathcal{I}_A$ using a monocular camera. To recover the position of the vehicle G_{AR} within the 3D point cloud, 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 as follows:

$$\widehat{G}_{AR} : \underset{G_{AR}}{\arg\min}\, NID(\mathcal{I}_A(\boldsymbol{x}_A),\, \mathcal{I}_S(\boldsymbol{q}) \mid \boldsymbol{q} \in \mathcal{S}_A) \quad (8)$$

where NID is the Normalised Information Distance between the prior appearance and the live appearance. By solving the above optimisation, we can recover the 6-DoF position of the vehicle relative to the 3D scene prior.

Fig. 6 presents the results from a 24-hour localisation experiment using LAPS. Two 3D scene priors were collected: an illumination invariant prior collected at midday for use during daylight hours, and an RGB prior collected at midnight for use at night. During the day, LAPS provided 6-DoF metric localisation using illumination invariant images accurate to 0.4 m in translation and 1.8 degrees in rotation, and at night, using RGB images, results accurate to 0.5 m in translation and 2.8 degrees in rotation.

This is an important result in the context of long-term robot operation, since it provides 24-hour day and night metric localisation with comparable accuracy to a INS system, using only low-cost monocular cameras with standard Bayer filters and a 3D point cloud prior collected offline by a survey vehicle. Additionally, it reduces the cost of surveying the environment prior to localisation, since only one survey during daylight hours and one survey at night is required to cover the variation in scene appearance due to the day-night cycle.

VI. Life-long Stereo Localisation

Another application that benefits from the use of an illumination invariant colour space is the life-long stereo localisation framework presented in [30]. This approach builds upon the experience-based navigation framework presented in [31], which represents the map as a series of stereo visual odometry sequences, dubbed *experiences*, along with metric localisations between different experiences of the same loca-

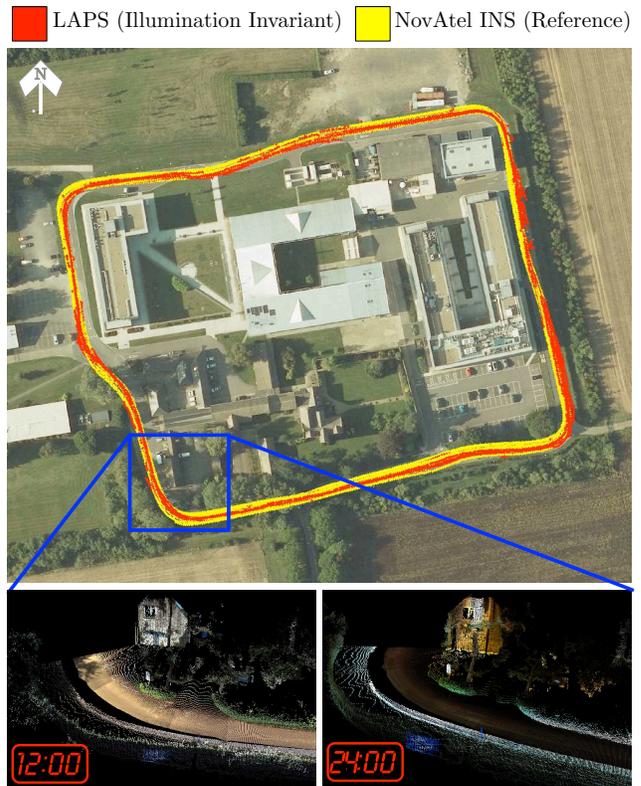LAPS (Illumination Invariant)   NovAtel INS (Reference)

Fig. 6. Metric global localisation results over a 24 hour period using LAPS. Using two Grasshopper2 monocular cameras and two 3D coloured pointcloud scene prior surveys (one illumination invariant prior collected at midday for use during daylight hours, pictured bottom left, and one RGB prior collected at midnight for use at night, pictured bottom right), LAPS provided global metric localisation with accuracy comparable to a NovAtel GPS-aided inertial navigation system over a full 24 hour period.

tion. As a robot traverses an environment, it simultaneously stores new experiences from the live stereo imagery, along with attempting to localise relative to prior experiences to link multiple representations of the same environment. This approach was found to work well for gradual scene change over time, but not if the appearance of the scene changes significantly between each visit.

The method in which illumination invariant images are incorporated into the experience-based navigation framework is illustrated in Fig. 7. A second localiser makes use of illumination invariant images in parallel with the main localisation system. Although the metric relative poses calculated between illumination invariant images tends to be more noisy than that calculated between greyscale (luminance) images, the localiser is less likely to fail if the scene appearance change is due to sunlight intensity, direction or spectrum variation. Fig. 8 shows two example localisations, illustrating the advantages of each colour space for matching locations over time. The combined system significantly reduced the mean distance between successful localisations, decreasing the probability of travelling 100 m without a successful localisation from 40% using the baseline greyscale-only system to only 5% using the combined greyscale and illumination invariant system. This is an important result in the context of life-long robot operation, as it significantly increases the mean-time-before-failure of the navigation system.
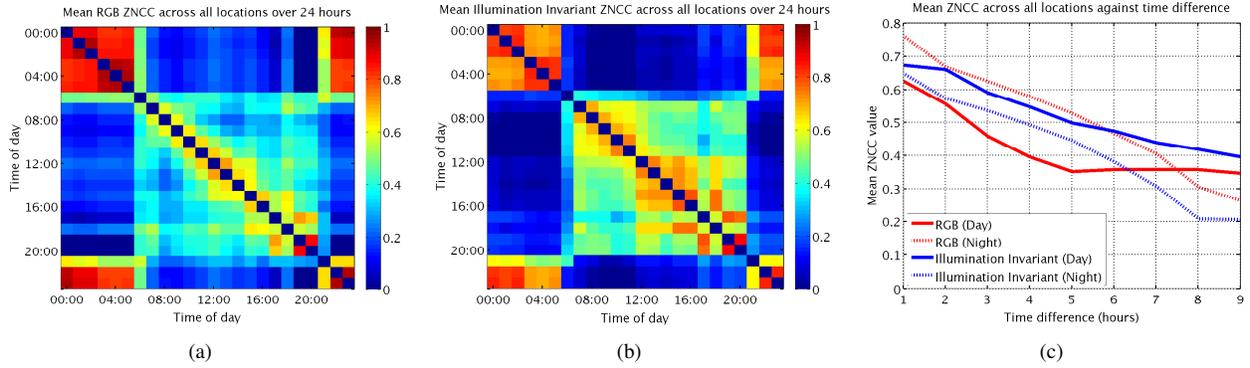
Fig. 5. Mean ZNCC between 71 locations at different times of day (higher values indicate better correlation). (a) and (b) show the mean ZNCC between locations across a 24-hour period. The mean correlation between RGB images during daylight hours (07:00 to 20:00) is low; the illumination invariant images exhibit more consistent block-like correlation over this period. However, the RGB images are more consistent at night, and to a lesser extent at dawn (06:00) and dusk (21:00). (c) shows the mean ZNCC between locations against the difference in time between traverses. Illumination invariant images are notably more consistent over time during daylight hours in comparison to RGB images. RGB images are more consistent over time at night, but the performance is comparable to illumination images during the day. Therefore, the best policy for visual localisation over a 24 hour period is to use illumination invariant images during daylight hours and raw RGB images at night.
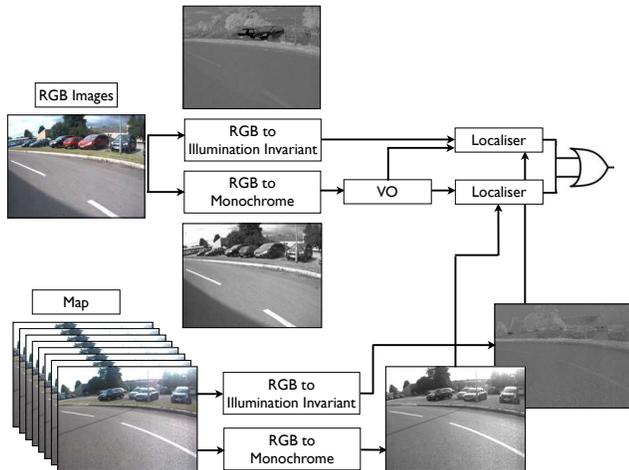
Fig. 7. Block-flow diagram of the combined stereo localisation approach. Live RGB images are converted to both monochrome (luminance) and illumination invariant colour spaces. Stereo visual odometry is used to generate metric pose estimates between sequential monochrome images, yielding a map in the form of a pose graph. In parallel, the system attempts to find metric estimates between the current live image and previous images stored in prior maps, using both monochrome and illumination invariant colour spaces. Although the monochrome images produce higher quality pose and velocity estimates between sequentially captured frames, the illumination invariant images are more consistent over time, increasing the likelihood of localising to a previously visited location under different lighting conditions. Localising in parallel combines the advantages of both colour space representations, increasing the robustness of the system over the lifetime of the vehicle.

VII. URBAN SCENE CLASSIFICATION

The primary goal of the illumination invariant transform is to produce an image where pixel values correspond to the material properties of objects viewed in the scene, regardless of scene illumination intensity, direction or spectrum. This makes it a natural front-end processing step for scene classification algorithms, which attempt to segment the image into clusters corresponding to objects [32], [33] and assign appropriate labels to the clusters [34], [35], in order to gain a high-level semantic understanding of the scene. Illumination invariant images provide both consistent appearance for objects made of the same material in different images, as well as clear differences in appearance between objects made of different materials in the same image.

In [36], the data-transfer approach of [33] was applied to both RGB and illumination invariant images from urban scenes gathered in Oxford as well as scenes from the public benchmark KITTI dataset [1]. Fig. 9 shows example classifications from both datasets, illustrating the significant increase in the quality of object classification when using illumination invariant images in the presence of shadows. Illumination invariant image classification provided higher recall and F1 scores for all object classes, with the exception of vehicles and sidewalks (due to reflective surfaces on vehicles and proximity to similar road surfaces respectively). This is an important result in the context of autonomous road vehicles, which must successfully navigate the road in the presence of a large range of static and dynamic object classes under a wide range of illumination conditions, and make use of higher-level semantic information for planning, decision-making and obstacle avoidance.

VIII. CONCLUSIONS

In this paper we have presented compelling results for the use of illumination invariant imaging to improve the performance and robustness of vision-based autonomous road vehicles in typical outdoor environments. The illumination invariant transform presented in this paper is no more complex than a typical RGB to greyscale conversion, and requires only a single parameter derived from the image sensor datasheet. However, the effects of this transform on images gathered over a 24 hour period are profound, removing almost all variation due to sunlight intensity, direction, spectrum and shadow present in the raw RGB images. We believe the improvements demonstrated in the three sample applications of metric monocular localisation, life-long stereo navigation and urban scene classification pave the way to ubiquitous low-cost autonomy for vision-based road vehicles in the future.
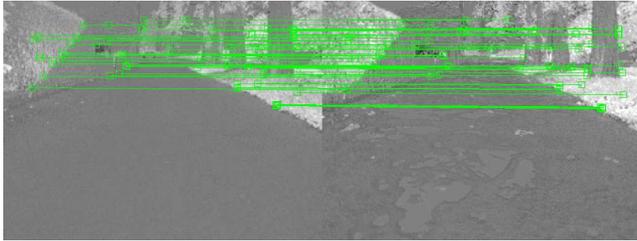
(a) Localisation under the trees.


(b) Localisation near a car park.

Fig. 8. Examples where the illumination invariant images helped the system localise under heavy shadows (a) and where the illumination invariant images failed to localise (b); individual feature matches are shown in green. Artefacts and noise introduced by the illumination invariant transform can interfere with point feature extraction and matching, which can sometimes result in fewer matches. However, the benefit of illumination invariant images becomes clear when looking at regions with high visual variability caused by outdoor illumination changes as in (a).
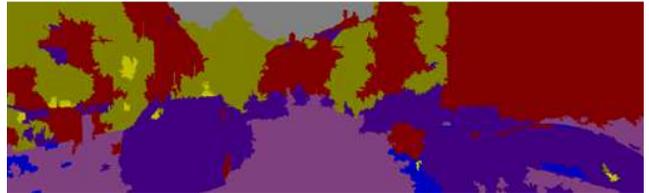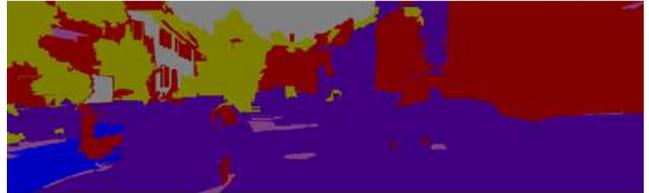
IX. ACKNOWLEDEGMENTS

REFERENCES

[1] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Computer Vision and*
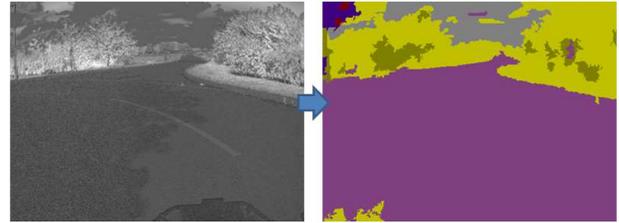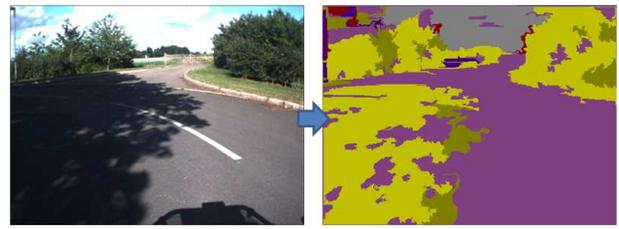
(a) Oxford classification results.

(b) KITTI classification results.

Fig. 9. A comparison of classification results using RGB and illumination invariant images. (a) shows results in Oxford using a Bumblebee2 camera and (b) shows results using the benchmark KITTI dataset, which uses a Flea2 camera. Note how in both cases the shadow on the road is correctly classified when using the illumination invariant image, and in (b) the illumination invariant classified buildings, road, vehicles, vegetation and sky closely match the true objects in the scene.

Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 3354–3361.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

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

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

[9] 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.

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

[11] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *International journal of computer vision*, vol. 65, no. 1-2, pp. 43–72, 2005.

[12] 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.

[13] H. Lategahn, J. Beck, B. Kitt, and C. Stiller, "How to learn an illumination robust image feature for place recognition," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, 2013, pp. 285–291.

[14] W. Maddern and S. Vidas, "Towards robust night and day place recognition using visible and thermal imaging," *RSS 2012: Beyond laser and vision: Alternative sensing techniques for robotic perception*, 2012.

[15] C. McManus, P. Furgale, B. Stenning, and T. D. Barfoot, "Visual teach and repeat using appearance-based lidar," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 389–396.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, "On the removal of shadows from images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 1, pp. 59–68, 2006.

[22] 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.

[23] S. Ratnasingam, S. Collins, and J. Hernández-Andrés, "Optimum sensors for color constancy in scenes illuminated by daylight," *JOSA A*, vol. 27, no. 10, pp. 2198–2207, 2010.

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

[25] M. S. Drew and H. R. V. Joze, "Sharpening from shadows: Sensor transforms for removing shadows using a single image," in *Color and Imaging Conference*, vol. 2009, no. 1. Society for Imaging Science and Technology, 2009, pp. 267–271.

[26] S. Henderson and D. Hodgkiss, "The spectral energy distribution of daylight," *British Journal of Applied Physics*, vol. 14, no. 3, p. 125, 1963.

[27] J. Hernández-Andrés, J. Romero, J. L. Nieves, R. L. Lee Jr, *et al.*, "Color and spectral analysis of daylight in southern europe," *JOSA A*, vol. 18, no. 6, pp. 1325–1335, 2001.

[28] W. Maddern, A. D. Stewart, and P. Newman, "LAPS-II: 6-DoF day and night visual localisation with prior 3D structure for autonomous road vehicles," in *Intelligent Vehicles Symposium, 2014 IEEE*. IEEE, 2014.

[29] 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.

[30] C. McManus, W. Churchill, W. Maddern, A. D. Stewart, and P. Newman, "Shady Dealings: Robust, long-term visual localisation using illumination invariance," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014.

[31] 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.

[32] L. Ladicky, C. Russell, P. Kohli, and P. H. Torr, "Associative hierarchical CRFs for object class image segmentation," in *International Conference in Computer Vision*, 2009.

[33] J. Tighe and S. Lazebnik, "SuperParsing: Scalable nonparametric image parsing with superpixels," in *European Conference on Computer Vision*, 2010.

[34] I. Posner, M. Cummins, and P. Newman, "A generative framework for fast urban labeling using spatial and temporal context," *Journal of Autonomous Robots*, vol. 26, p. 153, 2009.

[35] A. Geiger, M. Lauer, and R. Urtasun, "A generative model for 3D urban scene understanding from movable platforms," in *International Conference on Computer Vision and Pattern Recognition*, 2011.

[36] B. Upcroft, C. McManus, W. Churchill, W. Maddern, and P. Newman, "Lighting invariant urban street classification," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014.