2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Dealing with Shadows: Capturing Intrinsic Scene Appearance for Image-based Outdoor Localisation

Peter Corke Rohan Paul Winston Churchill Paul Newman

***Abstract*— In outdoor environments shadows are common. These typically strong visual features cause considerable change in the appearance of a place, and therefore confound vision-based localisation approaches. In this paper we describe how to convert a colour image of the scene to a greyscale *invariant* image where pixel values are a function of underlying material property not lighting. We summarise the theory of shadow invariant images and discuss the modelling and calibration issues which are important for non-ideal off-the-shelf colour cameras. We evaluate the technique with a commonly used robotic camera and an autonomous car operating in an outdoor environment, and show that it can outperform the use of ordinary greyscale images for the task of visual localisation.**

I. Introduction

> *"Shadows are everywhere! Yet, the human visual system is so adept at filtering them out, that we never give shadows a second thought; that is until we need to deal with them in our algorithms. Since the very beginning of computer vision, the presence of shadows has been responsible for wreaking havoc on a variety of applications."* [1].

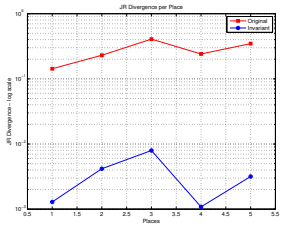
Vision-based localisation systems rely on place models based on scene appearance recorded as an image. However image formation is an interplay of both scene structure and the current lighting conditions. Ideally the same place would always produce the same image, but observations are strongly influenced by illumination. For place recognition we are not interested in modelling the lighting variations between observations, these only act as distractors in the image. Indeed, if we could always just observe the intrinsic scene appearance, uncorrupted by illumination changes, localisation would be a less messy business.

As a motivating example consider Figure 1(a), the road appearance is dominated by transient shadows. Now consider the transformed image in Figure 1(b). Clearly the strong shadowing effects on the road surface have been removed and the road appears as a grey tone — its true underlying appearance. Applying this transformation to all our shadow effected observations removes the illumination distractors and makes localisation considerably easier — similar places actually look similar, Figures 1(c) and 1(d).

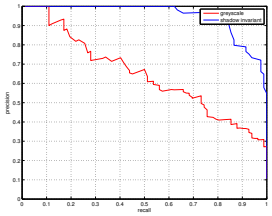
The ability to deal with confounding shadows and vast illumination change is important for any localisation algorithm hoping to achieve long-term success. One example is our recent Experience Based Navigation [2] where we model the world as a set of independent but related experiences. The number of experiences stored for a place is a function of its visual diversity: highly varying places require many experiences, while relatively staid regions are described by a handful. We found that sections like Figure 1(a) spawned many experiences, not because the underlying scene had changed, but due to the large illumination changes it undergoes. We are therefore motivated to find an invariant representation that removes these ephemeral shadow effects.

(a) Image with shadows

(b) Shadow invariant image



(c) Intrinsic JR divergence



(d) Precision-recall curve

Fig. 1. Shadows create confounding effects in images (a). The shadow invariant image eliminates the effect of lighting (b) by capturing the intrinsic material properties of the scene, invariant to shadow effects (c). This has a effect on significantly improving place recognition based on image similarity metric (d). Figure best viewed in colour. In (c-d) the *red* and *blue* plots are used for the original grayscale and shadow invariant images respectively.

In this paper we present a novel approach to appearance-based localisation for an outdoor robot, based on *intrinsic* rather than *visual* scene appearance. We apply recent results on shadow removal to eliminate the effects of scene illumination and recover an image that more truly reflects the intrinsic, or material, characteristics of the environment. To make this work with a real colour camera we have developed a new approach to computing the sharpening matrix which is used to partly decouple the colour channels. The overall approach is evaluated using images from a real outdoor robot acquired at different times of day, at a number of locations, and across several days. The data includes major illumination changes and shadow effects. Using this dataset we demonstrate a significant improvement in precision and

Peter Corke is with the School of Electrical Engineering and Computer Science, Queensland University of Technology, Australia. e-mail: Peter.Corke@qut.edu.au.

Rohan Paul, Winston Churchill and Paul Newman are with the Mobile Robotics Group, University of Oxford, UK. e-mail: {rohanp,winston,pnewman}@robots.ox.ac.uk.

recall over standard vision-based image retrieval.

The next section presents related works and Section III introduces the theory of shadow removal — a much studied problem in the computer vision community — which sets the scene for the rest of the paper. Section IV applies the technique to real images with a non-ideal camera as required for our problem of image-based robot localisation. Section V demonstrates the approach on imagery collected by a mobile platform and finally Section VI concludes the paper.

II. Related Works

Researchers in the computer vision community have explored shadow detection and removal in images, with approaches falling under two categories. The first relies on learning classifiers based on intensity and colour cues, combined with an energy minimisation step to ensure global consistency. For example, Zhu *et al.* [3] use boosting and CRFs to classify shadows in single monochromatic natural images. Guo *et al.* [4] compare pairwise patches of the same texture under similar and different lighting conditions in a graph-cut framework to produce a binary labelling. Shadow removal is performed using a soft matting technique to reduce edge effects of the binary classification.

The second class of techniques attempt to model the physical process of image formation: the illumination and the underlying scene reflectance properties. In a series of papers, Finlayson *et al.* [5] [6] developed the idea of invariant images. Modeling the image formation process, they compute a transformation from RGB to an invariant log-chromaticity space. Further, Narasimham *et al.* [7] demonstrated that the radiance of the same scene observed under large illumination changes can be used to classify material type and produce invariant images. Nayar *et al.* [8] [9] present a physics-based approach for describing scene appearance when imaged under non-ideal weather conditions like haze and fog. The model allows recovery of pertinent scene properties like 3-D structure and clear day contrasts given one or more images.

Recently Kwatra *et al.* [10] presented an information-theoretic approach, which can be considered as a hybrid of the two categories of approaches described above. Leveraging the fact that the entropy is always greater in the observed image than the reflectance and illumination fields, they use an energy minimisation scheme with texture and smoothness priors to remove shadows. Their solution is continuous and works effectively on diffuse shadows such as those cast by clouds in aerial imagery with processing times of 10-20 seconds for typical image sizes on a very high end machine.

Within mobile robotics the problem has received recent attention in the context of detection shadows cast on the road by vehicles and surrounding structure. Park and Lim [11] estimate the lit-road texture by sampling a small window in front of the host cars bumper, which they assume is unobstructed, and then apply background subtraction techniques. Boosting is applied at the pixel level followed by global smoothing. They incorporate heuristic and contextual cues specific to the shadows cast by cars, and are unable to cope with natural shadows.

In this paper we apply the model based approach described by Finlayson *et al.* [5] [6] to produce invariant images. We use these to improve our localisation performance for scenes that exhibit strong shadowing effects.

III. Theory of Invariant Images

In this section we briefly recapitulate the theory related to shadow removal, that is, how to create a greyscale invariant image from a colour input image [5] [6]. The key concepts in the colour perception process are the source of illumination (the illuminant) with a spectral power density of $E(\lambda)\,\mathrm{W/m^2/m}$; a point in the scene with reflectance $R(\lambda) \in [0,1]$ which produces a luminance $L(\lambda)\,\mathrm{W/m^2/m}$ that induces a stimulus in the sensor possessing a spectral response given by $M(\lambda)$.

Referring to Figure 1(a) we see bright parts of the scene which are illuminated directly by the sun and the darker shadow regions which are illuminated, not by the sun, but from the sky whether blue or cloudy. Our first assumption is that points in the scene are illuminated by a blackbody radiator, that is, one whose radiant power spectral density is given by Planck's law

$$E(\lambda) = \frac{2hc^2}{\lambda^5(e^{hc/k\lambda T}-1)}\mathrm{W/m^2/m} \tag{1}$$

where h, k and c are Planck's constant, Boltzmann's constant and the speed of light respectively. T is the temperature of the radiator but in this context can be considered the colour temperature of the illuminant which varies from 2,000-3,000K for dawn or dusk, 5,000-5,400K for noon-day sun, 8,000-10,000K for an overcast sky, and 10,000-12,000K for blue sky. Shadows therefore have two identifying characteristics: they are dark and they are more blue than the same material under direct sun illumination.

For trichromatic vision the response of the colour channels is given by

$$\begin{aligned} R &= \int_\lambda E(\lambda)R(\lambda)M_R(\lambda)d\lambda \\ G &= \int_\lambda E(\lambda)R(\lambda)M_G(\lambda)d\lambda \\ B &= \int_\lambda E(\lambda)R(\lambda)M_B(\lambda)d\lambda \end{aligned} \tag{2}$$

where $M_x(\lambda)$ is the spectral response of the sensor for the channel $x \in \{R,G,B\}$. If we consider the sensor responses as the ultimate narrow band filters, Dirac functions $M_x(\lambda) = \delta(\lambda - \lambda_x)$ we can simplify the integrals of (2) and write

$$\begin{aligned} R &= E(\lambda_R)R(\lambda_R)M_R(\lambda_R) \\ G &= E(\lambda_G)R(\lambda_G)M_G(\lambda_G) \\ B &= E(\lambda_B)R(\lambda_B)M_B(\lambda_B) \end{aligned} \tag{3}$$

where $M_x(\lambda_x)$ is the peak response of the sensor at wavelength λ_x.

The next step is to reduce dimensionality by removing the effect of changes in illumination magnitude. We achieve this

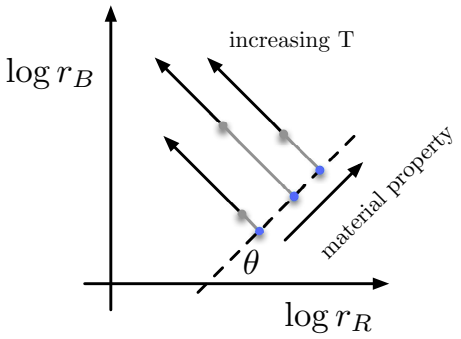


Fig. 2. In the log-chromaticity space the colour appearance of each material appears along parallel lines as a function of illuminant temperature T. A shadow invariant greyscale image is obtained by projecting the points on to the direction (defined by θ) orthogonal to the illumination change direction.

by, for each pixel, computing chromaticity coordinates

$$r = \frac{R}{G},\ b = \frac{B}{G}. \tag{4}$$

For the case of r we substitute (1) and (3) and write

$$\begin{aligned} r &= \frac{E(\lambda_R)R(\lambda_R)M_R(\lambda_R)}{E(\lambda_G)R(\lambda_G)M_R(\lambda_G)} \\ &= \frac{\frac{2hc^2}{\lambda^5(e^{hc/k\lambda_R T}-1)}R(\lambda_R)M_R(\lambda_R)}{\frac{2hc^2}{\lambda^5(e^{hc/k\lambda_G T}-1)}R(\lambda_G)M_R(\lambda_G)} \end{aligned}$$

and similarly for b. The particular choice of chromaticity coordinates and alternatives is discussed in Section IV-A. A helpful approximation is to eliminate the -1 term, which is reasonable for colour temperatures in the range under consideration, and this allows us to simplify

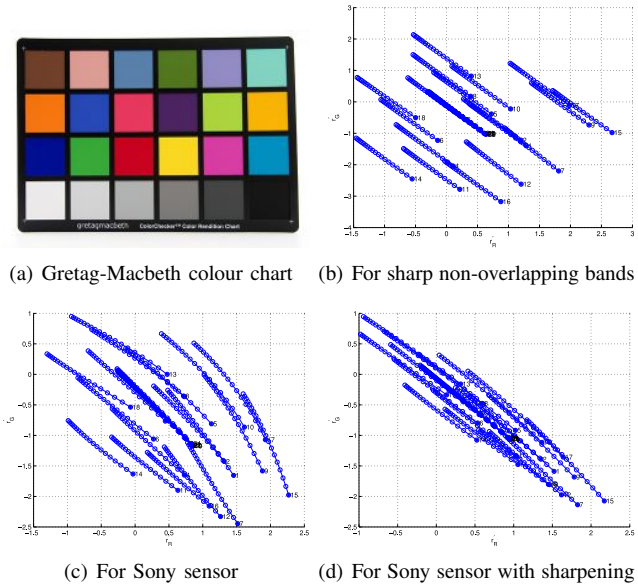
$$\begin{aligned} r &\approx \frac{e^{hc/k\lambda_G T}R(\lambda_R)M_R(\lambda_R)}{e^{hc/k\lambda_R T}R(\lambda_G)M_G(\lambda_G)} \\ &\approx e^{hc(1/\lambda_G - 1/\lambda_R)/kT}\frac{M_R(\lambda_R)}{M_G(\lambda_G)}\frac{R(\lambda_R)}{R(\lambda_G)} \end{aligned}$$

which is a function of the colour temperature T and various constants: natural constants, sensor response wavelength λ_x and magnitude $M_x(\lambda_x)$, and material properties $R(\lambda_x)$. We treat b similarly. Taking the logarithm leads to the very simple form for red and blue chromaticity:

$$\log r = c_1 - \frac{c_2}{T},\ \log b = c_1' - \frac{c_2'}{T}. \tag{5}$$

Plotting $\log b$ against $\log r$ then each pixel is mapped to a point, but as the colour temperature changes the coordinate of each pixel will move along a line with a slope of c_2'/c_2, see Figure 2. Importantly this means that as illumination changes all points move in the same direction, so a projection onto the orthogonal direction results in a 1-dimensional quantity invariant to colour temperature and a function of material reflectance and camera sensor properties. We exponentiate the projected value to improve the dynamic range and return a positive number.

To verify the performance of this algorithm we ran a simple simulation. We considered the standard Gretag-Macbeth colour checker test chart, see Figure 3(a), for which $R_i(\lambda)$ has been measured and tabulated for each of the $i = 1 \cdots 24$



(a) Gretag-Macbeth colour chart (b) For sharp non-overlapping bands

(c) For Sony sensor (d) For Sony sensor with sharpening

Fig. 3. Locus of Gretag-Macbeth colour checker test squares (shown in a) as colour temperature varies from 2,500–10,000 K. Numbers indicate the corresponding tile on the colour checker chart (numbered left-right and top-bottom). (b) For narrow-band, non-overlapping, spectral bands. (c) For Sony sensor as per Figure 4. (d) For Sony sensor with sharpening transform and geometric mean.

tiles[1]. For a range of blackbody illuminants with colour temperature spanning the range of $2,500 - 10,000\,\mathrm{K}$ we can plot the locus of the points in the log-chromacity space, see Figure 3(b) and they are indeed straight lines as theory predicts.

To summarise, we have mapped a 3-dimensional pixel value which is a complex mixture of material and illumination properties at the world point to a 1-dimensional pixel value which is a function of only material reflectance properties. To achieve this we have made a number of simplifying assumptions: that points in the scene are illuminated by a Planckian source, that the camera sensor magnitude response is a linear function of scene luminance and its spectral response is a Dirac function. Importantly, a mixture of two Planckian sources is not Planckian and many artificial light sources are not Planckian.

IV. Invariant Images from Real Cameras

In practice the assumptions just stated do not hold. Real colour cameras have overlapping spectral responses and we verify the negative impact of this in simulation, and then investigate some strategies to rectify it.

A. Real camera spectral response

In this work we use the well known Point Grey Bumblebee 2 stereo colour camera, although just the left image is used for the results in this paper. This camera has a linear magnitude response, that is, it has $\gamma = 1$, but the spectral response for its Sony ICX204 sensor as shown in Figure 4 is far from three Dirac functions — there is considerable

[1]The tiles in the bottom row of the chart are shades of grey and therefore have the same shaped spectral response, differing only in magnitude. They are all tones of the colour white.

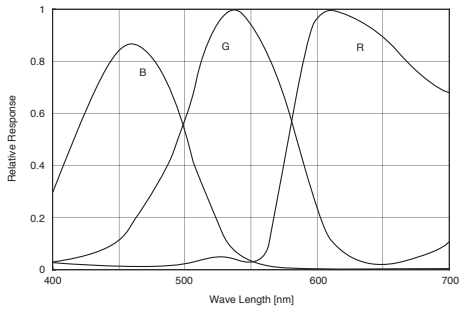


Fig. 4. Spectral response of the Sony ICX204 sensor as used in the Point Grey Bumblebee 2 camera.

spectral overlap between all three colour channels. If we repeat the previous simulation exercise for the case of this camera the locus of the points on the log-chromaticity plane is shown in Figure 3(c) and far from straight. Projecting the points onto a line would not result in an illumination invariant value, the tiles would map to overlapping intervals not points.

There are a number of solutions that can ameliorate this result. The first is to revisit the chromaticity coordinates that we chose earlier. There is no particular reason to have chosen the green value for the denominator, green is not privileged in any way. It certainly will lead to problems when the green value is zero or even small and dominated by photosite noise (shot and thermal noise). Instead of (4) we choose to normalise by the geometric mean

$$r = \frac{R}{(RGB)^{1/3}},\ b = \frac{B}{(RGB)^{1/3}} \tag{6}$$

which improves the straightness of the loci on the log-chromaticity plane.

The second is to introduce a sharpening transform [12] which is applied to camera tristimulus output values such that

$$[R', G', B']^T = \mathbf{M}[R, G, B]^T$$

where $\mathbf{M} \in \mathbb{R}^{3\times 3}$ and the columns have unit norms. We define a cost function which penalises all pairwise channel overlaps (R-G, B-G and R-B)

$$J = \int_\lambda (R'(\lambda)-G'(\lambda))^2+(B'(\lambda)-G'(\lambda))^2+(R'(\lambda)-B'(\lambda))^2 d\lambda \tag{7}$$

and adjust $\mathbf{M}$ to maximise this. In order to ensure unit column norms we consider each column as a point on the surface of a unit-sphere which we parametize by two angles (latitude and longitude), giving six variables to optimise over. The result

$$\mathbf{M} = \begin{pmatrix} 0.998800 & -0.066900 & -0.000100 \\ -0.049700 & 0.988600 & -0.000100 \\ -0.004300 & -0.134600 & 1.000000 \end{pmatrix} \tag{8}$$

is quite close to a unit matrix. Figure 5 shows the original sensor response and the sharpened response. The change is quite subtle but nevertheless leads to a significant improvement in computing a shadow invariant image.

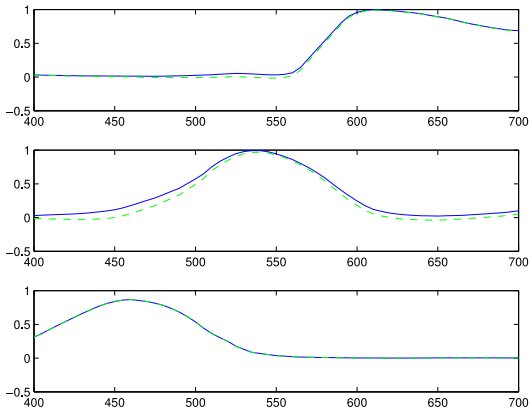
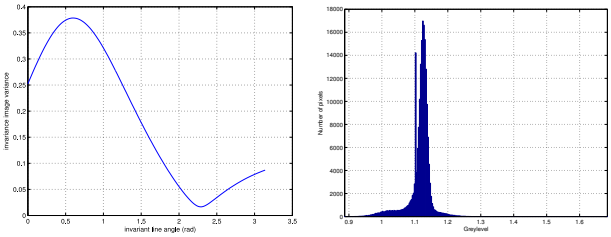


Fig. 5. Original sensor response (blue) and sharpened response (green) for the ICX204 sensor.



(a) (b)

Fig. 6. (a) Example of shadow invariant variance as a function of projection angle. Variance is computed over a set of pixels that belong to the same material. (b) Histogram of shadow invariant pixel values for the image in Figure 1(a).

The result of applying the chromaticity coordinates (6) and the sharpening transform leads to the log-chomaticity plane loci shown in Figure 3(d) which are considerably straighter.

B. Estimating the projection angle

Points on the log-chromaticity plane move parallel to the vector $[c_2, c_2']$ and the orthogonal projection line is therefore $[c_2', -c_2]$. c_2 and c_2' are functions of well known physical constants and the peak spectral response of the green channel (denominator) and the channel $X = R, B$. Theoretically the angle of this line can be shown to be $2.73\,\mathrm{rad}$ but the many assumptions made will introduce errors.

Instead, we train for the angle based on a sample image. The user selects a region (by clicking perimeter points on an image) that comprises the one type of material under varying lighting conditions. For the case of the image shown in Figure 1(a) we highlighted the lit and shadowed road region. We first compute the log-chromaticity image where each pixel is $(\log r'_{uv}, \log b'_{uv})$ and then find the projection angle θ that minimises variance over the region

$$\theta = \arg\ \min_{\theta\in[0,\pi]} \sum_{\langle u,v\rangle\in\mathcal{R}} (g_{uv} - \bar{g})^2 \tag{9}$$

where

$$g_{uv} = (\log r'_{uv},\ \log b'_{uv})^T(\cos\theta,\ \sin\theta) \tag{10}$$

and $\langle u, v\rangle \in \mathcal{R}$ are the pixels in the user selected image region $\mathcal{R}$. A typical plot of variance versus θ is shown

in Figure 6(a) and across many images we see results consistently around $2.3\,\mathrm{rad}$.

C. Practical considerations

There are a number of important practical issues in implementing this algorithm. If any pixel in the scene is saturated, that is it has one or more of $R = 255$, $G = 255$ or $B = 255$ then the true colour of the corresponding world point cannot be known. If any pixel has $R \sim 0$, $G \sim 0$ or $B \sim 0$ then the denominator of (4) or (6) will be small and the result will be either infinity or a large value dominated by pixel noise. Chromaticity values can therefore have any value between 0 and ∞ as well as being indeterminate (NaN). After taking logarithms the result can be in the range $-\infty$ to $+\infty$ as well as being indeterminate (NaN). To prevent some of these extreme values and to deal with saturation we choose only pixels where $\eta < R < 255 - \eta$ and $\eta < G < 255 - \eta$ and $\eta < B < 255 - \eta$ and we have chosen $\eta = 10$.

As shown in Figure 3(d) the result after various ameliorations is still not perfect. That is, a single material maps to an interval in the invariant image. This cannot be avoided unless we use a camera with non-overlapping spectral responses, and some multispectral cameras do have this property.

The dynamic range of the invariant image is quite small and the images appear to be somewhat washed out. Contrast enhancement techniques such as linear stretching or histogram normalisation can be applied. The pixel value distribution, for example Figure 6(b) has long tails due to the numerical issues mentioned above. In this example, although the distribution is not Gaussian, the standard deviation of 0.0252 is a very small fraction of the total dynamic range.

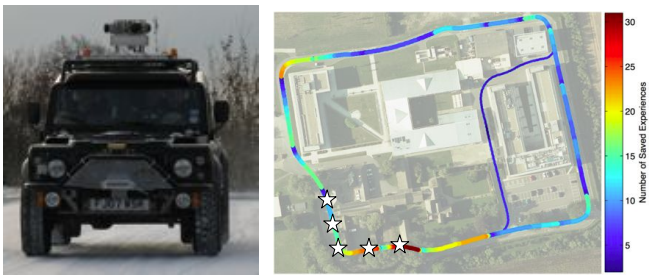
Finally, we note that we have taken the camera specification at face value and have not tested the linearity of the camera or whether the camera channels have any DC offsets.

V. Results

To evaluate the approach we conducted experiments on data collected from our autonomous vehicle, the Wildcat shown in Figure 7(a). The vehicle is equipped with a Point Grey Bumblebee2 stereo camera mounted on the front bumper. The data was collected around our field test site in Begbroke, Oxfordshire, Figure 7(b) over multiple runs under different weather conditions and times of day.

A. Qualitative results

Firstly, we present results on representative images taken from a wide range of outdoor scenes. Figure 10 presents some typical examples of shadow removal in our test images. In frames (a) and (b) foliage shadows are removed, while (c) and (d) remove human and lamp-post shadows. Frame (b) is taken under bright (in an English sort of way) sun which causes shadows on the road as well saturation of the building at the top left. The shadow invariant cannot be computed for these pixels. Note in (d), the yellow bin in the background is mapped to a very dark greyscale value.



(a) Wildcat vehicle (b) Begbroke site

Fig. 7. (a) Wildcat autonomous platform used to collect data for the experiments. The vehicle is equipped with a Point Grey Bumblebee 2 stereo camera mounted on the front bumper of the vehicle. (b) Overhead image of the field site indicating the GPS trajectory taken by the vehicle. The colours indicate the number of experiences created and the white stars are the places we consider.

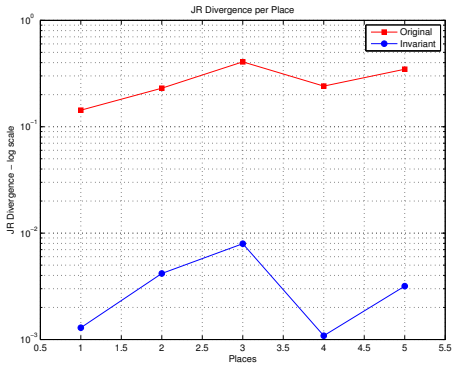
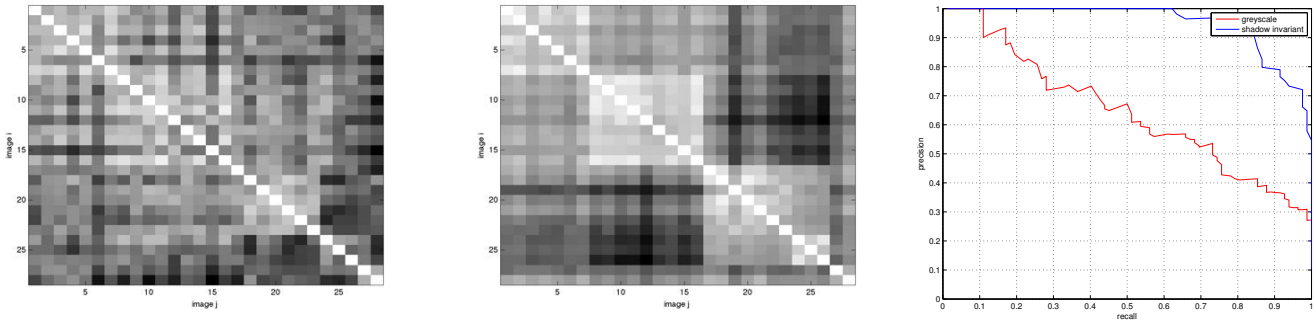


Fig. 9. Comparison of Jensen-Renyi divergence for the set of invariant and original images for 5 places imaged under varying shadow conditions. The divergence values are smaller for the set of invariant images demonstrating that they indeed capture the intrinsic representation of the scene irrespective of shadows effects. Note the log-scale on y-axis.

B. Place recognition

In our recent work on Experience Based Navigation [2], a proposed framework for long term navigation, we allow the visual diversity of a place to dictate the number of representations needed to model that place. We found most regions could be described by a handful of experiences, but sections influenced by transient shadows created more. We consider several locations, marked in Figure 7(b) with white stars, where the system has failed to localise and caused many experiences to be remembered. These sections are characterised by overhanging trees which cast ever changing shadow patterns, producing images with unique appearances. For the 5 selected points around this corner we gather all images that the robot considers to be exemplars of a new experience. The number of images per experience varies from 2 to 9. Figures 11 and 12 show example images (top row each) from two places demonstrating the diversity of visual experiences recorded at a place along the route.

We compute image similarity using the approach similar to [13] where the images are subsampled to 48×64 and compared using the zero-mean normalised cross correlation measure (ZNCC) [14]. Color images are first converted to



(a) Confusion matrix: original images (b) Confusion matrix: shadow invariant images (c) Precision-recall curves

Fig. 8. Image similarity computed for 5 places that were considered as 28 different visual experiences. Confusion matrix (bright is similar) computed on (a) original greyscale images and (b) shadow invariant images. Note that images of the same location in the invariant space are found more similar to each other and more dissimilar to images of other places (appearing as bright blocks) compared to original images (where similarity pattern is chaotic). Precision-recall curves (c) for original greyscale and shadow invariant image similarity for five different places. The curve for the invariant images dominates.

greyscale using the ITU-509 transformation [14], whereas the invariant images are taken as is. The pairwise similarity of all images is computed and displayed as a confusion matrix in Figure 8 for the original and invariant image cases. We can see that using the greyscale image results in a confusion matrix that is quite chaotic showing very little interplace similarity or intraplace dissimilarity, Figure 8(a). The confusion matrix for the shadow invariant case shows much greater structure where images of the same place display higher similarity and greater dissimilarity with images from other places, Figure 8(b).

Figure 8(c) illustrates the precision-recall curves derived from these matrices obtained by varying the threshold at which images are considered to be from the same place. The PR curve for the invariant images dominates the case for original images demonstrating that places can be significantly more reliably recognised using only image similarity if the shadows are first eliminated. In the context of this navigation system [2] this similarity could be used as an additional navigation cue to prevent the number of robot experiences growing without bound due to the almost infinite variety of lighting patterns due to shadows cast on the scene.

C. Measuring intrinsic nature

Next, we quantify the degree to which invariant images capture the intrinsic appearance of a place i.e., yield similar reflectivity results irrespective of lighting. For each place we measure the coherence within the set of invariant images and the original image set respectively.

Formally, we compute the Jensen-Renyi (JR) divergence [15] between the set of images for a single place represented as intensity histograms. The JR divergence evaluates the cumulative dissimilarity between two point sets (histograms in our case). We use a closed-form solution which uses Mixture of Gaussians [15], relying on a kernel density [16] driven non-parametric estimate of the pdf that automatically incorporates smoothness ameliorating bin-quantization effects [17] [18].

Figure 9 plots the result (note the log-scale on the y-axis). The JR divergence values are significantly and consistently lower for the invariant images compared to those for the original images affected by shadow changes. This indicates that the invariant images are indeed closer to generating an intrinsic representation of the scene.

D. Computational cost

The invariant image is computationally cheap to compute — it is not much more expensive than converting a colour image to greyscale. Importantly, it can be computed in constant time. In (6) the cube-root operation is the most expensive but we can first take the logarithms of the R, G and B images and then use addition, multiplication and subtraction to compute the log-chromaticity coordinates. On a 2.6 GHz i7 processor using MATLAB the invariant transform can be computed at 39 ns/pixel, or 40 ms for a megapixel image.

E. Benefits of shadow invariance

Computing the shadow invariant can be considered as a preprocessing step prior to a standard object recognition or feature extraction step. We gain the advantage of processing a scene free of distracting and confusing shadow artefacts. Most object recognition algorithms rely on a region segmentation step based on intensity appearance. If we apply a state-of-the-art graph-based segmentation [19] to the heavily shadowed image of Figure 13(a) we obtain the highly over segmented result shown in Figure 13(b). This clearly shows a complex pattern of segments that do not correspond to the materials present in the scene, rather they reflect the instantaneous pattern of lighting on the scene.

For most robotics tasks the lighting is irrelevant, we are not interested in pixel values but rather the semantics associated with those pixels. For a road navigation task that might be whether the pixel lies on the road or off it. For example, the shadow invariant image is shown in Figure 13(c) and the corresponding segmentation (using the same parameters) is shown in Figure 13(d). We clearly achieve a reduction in the number of segments and a very crisp delineation of the primary navigation feature — the road.

The shadow artefacts are not only irrelevant they are also distracting — corner detectors are drawn to the high contrast

Fig. 14. The approach does not compensate for shadows containing reflected lighting from objects in the scene. The figure illustrates an example where shadows next to coloured walls are not fully removed.

textures induced by shadows rather than the underlying structure. We have applied standard point feature extraction (e.g. SIFT, SURF etc. [20]) to the invariant image with success. Despite the lower SNR of the invariant image all but the smallest scale features reliably associate with material rather than lighting features of the scene.

F. Limitations

One of the limitations of this method is that the model assumes scene lighting by a single Planckian source, (Section IV) and hence cannot fully compensate when shadows are partly-illuminated by light reflected from objects populating a scene. For example, Figure 14 shows a strong shadow next to a building but the shadow is clearly still evident in the invariant image. In this case the shadow region is illuminated by sky light reflected from the coloured wall of the building which makes its spectrum non-Planckian.

VI. Conclusion

In this paper we have described an approach to eliminate shadows from colour images of outdoor scenes that is known in the computer vision community and applied it to a hard robotic problem of outdoor vision-based place recognition. We have described the details of key implementation steps such as minimising camera spectral channel overlap and estimating the direction of the projection line, and discussed approaches to overcome practical problems with low and high pixel values.

VII. Acknowledgements

Peter Corke was supported by Australian Research Council project DP110103006 Lifelong Robotic Navigation using Visual Perception. Winston Churchill was supported by an EPSRC Case Studentship with Oxford Technologies Ltd. Paul Newman was supported by an EPSRC Leadership Fellowship, EPSRC Grant EP/I005021/1. Authors thank Mark Sheehan and Dr. Alastair Harrison for insightful discussion on JR divergence and Dr. Benjamin Davis for maintaining the robotic platform used for this work. We thank Dominic Wang for valuable suggestions on this paper.

References

[1] J. Lalonde, A. Efros, and S. Narasimhan, "Detecting ground shadows in outdoor consumer photographs," *Computer Vision–ECCV 2010*, pp. 322–335, 2010.

[2] W. Churchill and P. Newman, "Practice makes perfect? managing and leveraging visual experiences for lifelong navigation," *IEEE International Conference on Robotics and Automation*, 2012.

[3] J. Zhu, K. Samuel, S. Masood, and M. Tappen, "Learning to recognize shadows in monochromatic natural images," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 223–230.

[4] R. Guo, Q. Dai, and D. Hoiem, "Single-image shadow detection and removal using paired regions," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 2033–2040.

[5] G. Finlayson, M. Drew, and C. Lu, "Intrinsic images by entropy minimization," *Computer Vision-ECCV 2004*, pp. 582–595, 2004.

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

[7] S. Narasimhan, V. Ramesh, and S. Nayar, "A class of photometric invariants: separating material from shape and illumination," in *IEEE International Conference on Computer Vision*, oct. 2003, pp. 1387 –1394 vol.2.

[8] S. Nayar and S. Narasimhan, "Vision in bad weather," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2. Ieee, 1999, pp. 820–827.

[9] S. Narasimhan and S. Nayar, "Chromatic framework for vision in bad weather," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 1. IEEE, 2000, pp. 598–605.

[10] V. Kwatra, M. Han, and S. Dai, "Shadow removal for aerial imagery by information theoretic intrinsic image analysis," in *Computational Photography (ICCP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1–8.

[11] S. Park and S. Lim, "Fast shadow detection for urban autonomous driving applications," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. IEEE, 2009, pp. 1717–1722.

[12] M. Drew and H. Joze, "Sharpening from shadows: Sensor transforms for removing shadows using a single image," in *Color Imaging Conference*, 2009, pp. 267–271.

[13] M. Milford, "Visual route recognition with a handful of bits," in *Proceedings of Robotics: Science and Systems*, Sydney, Australia, July 2012.

[14] P. I. Corke, *Robotics, Vision & Control: Fundamental Algorithms in MATLAB*. Springer, 2011, iSBN 978-3-642-20143-1.

[15] F. Wang, T. Syeda-Mahmood, B. Vemuri, D. Beymer, and A. Rangarajan, "Closed-form Jensen-Rényi divergence for mixture of gaussians and applications to group-wise shape registration," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pp. 648–655, 2009.

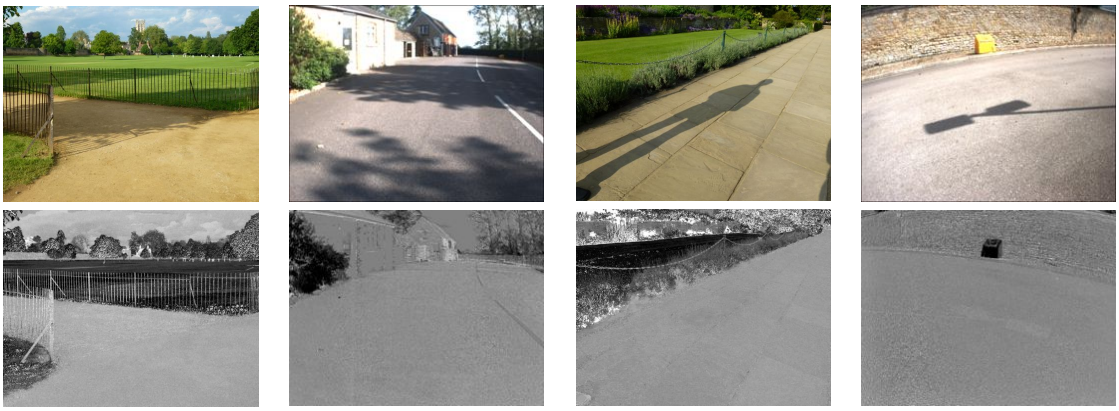
[16] Z. Botev, J. Grotowski, and D. Kroese, "Kernel density estimation via diffusion," *The Annals of Statistics*, vol. 38, no. 5, pp. 2916–2957, 2010.

[17] M. Sheehan, A. Harrison, and P. Newman, "Self-calibration for a 3d laser," *The International Journal of Robotics Research*, 2011.

[18] A. Hamza and H. Krim, "Image registration and segmentation by maximizing the Jensen-Rényi divergence," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Springer, 2003, pp. 147–163.

[19] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[20] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.



(a) (b) (c) (d)

Fig. 10. Gallery of results showing the original (top row) and shadow invariant (bottom row) image for a selection of places. Shadows cast by trees (a, b), people (c) and street lights (d) are removed in the invariant images.

Fig. 11. Original (top row) and invariant (bottom row) images for one of our test places. Note how images with significant and varying shadow effects are transformed back to a shadow-invariant representation.

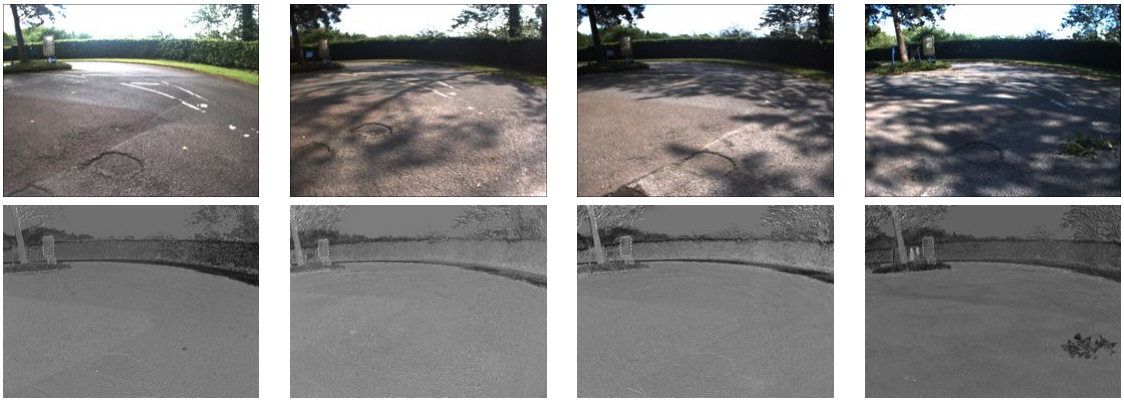
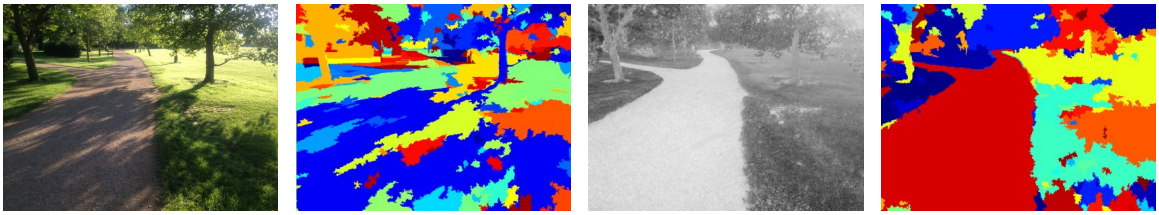


Fig. 12. Original (top row) and invariant (bottom row) images for one of our test places. Regardless of the shadowing effect, it is removed in the invariant image. Also note in the final image, the small piece of foliage fallen on the road that appears in the invariant image as it is a different material.



(a) Original image (b) Colour image segmentation (c) Shadow invariant image (d) Invariant image segmentation

Fig. 13. Comparison of region segmentation [19] based on intensity appearance for original and invariant images. The road path is clearly segmented out unaffected by shadow patterns in the invariant image. The result for the original image is over-segmented with the algorithm confounded by the instantaneous shadow and illumination patterns.