Proceedings of Australasian Conference on Robotics and Automation, 2-4 Dec 2014, The University of Melbourne, Melbourne, Australia

A method to quantify a descriptor's illumination variance

Patrick Ross, Andrew English, David Ball, Peter Corke
ARC Centre of Excellence for Robotic Vision
School of Electrical Engineering and Computer Science
Queensland University of Technology
p6.ross@qut.edu.au

Abstract

This paper presents a new metric, which we call the lighting variance ratio, for quantifying descriptors in terms of their variance to illumination changes. In many applications it is desirable to have descriptors that are robust to changes in illumination, especially in outdoor environments. The lighting variance ratio is useful for comparing descriptors and determining if a descriptor is lighting invariant enough for a given environment. The metric is analysed across a number of datasets, cameras and descriptors. The results show that the upright SIFT descriptor is typically the most lighting invariant descriptor.

1 Introduction

Feature descriptors are necessary for a variety of different robotic vision problems, including SLAM and classification. The ability to robustly match descriptors over long periods of time is important to achieve the goal of persistent operation for these problems. Despite this, many descriptors such as SIFT (Scale Invariant Feature Transform) [Lowe, 2004] and SURF (Speeded-Up Robust Features) [Bay *et al.*, 2006] have been shown to exhibit variation due to illumination changes [Mikulík *et al.*, 2010; Ranganathan *et al.*, 2013; Ross *et al.*, 2013; Valgren and Lilienthal, 2007; 2010], as have more complex colour descriptors [Van de Sande *et al.*, 2008; 2010]. These variations due to illumination reduce the matching accuracy of descriptors over long term operation, and can significantly impact the performance of persistent systems [Kim *et al.*, 2007; McManus *et al.*, 2011]. To this end, a standardised metric for quantifying the illumination variance of descriptors under realistic operating conditions is necessary to determine suitable descriptors, as well as to direct the design of illumination invariant descriptors.

Prior work in the area of illumination invariance has largely focussed on quantifying only matching accuracy, typically for purposes of SLAM [Ranganathan *et al.*, 2013; Valgren and Lilienthal, 2007; Van de Sande *et al.*, 2008; 2010]. These methods are sensitive to the choice of environment and keypoint detectors, although the method presented by [Van de Sande *et al.*, 2008; 2010] is less so given that they utilised densely packed keypoints for one of their experiments. Additionally, any method which compares keypoint matching accuracy over different images requires complete knowledge of correct results, and as such these methods typically require modest human intervention. These methods typically don't separate variance due to keypoint detection from the variance due to the descriptor, and since different keypoint detectors are used for different descriptors, this makes direct comparisons difficult.

Ross et al [Ross *et al.*, 2013] presented a method for quantifying lighting variance in terms of the amount of the descriptor variance accounted for by lighting. These types of variance were assumed to change slowly over time; however in practise it assumed that lighting was constant over a given window of ten minutes. While this assumption is valid throughout the day, when the lighting is quickly changing such as around sunrise and sunset this assumption is invalid. The method required a set of timelapse imagery and choice of a set of keypoints, however after this initialisation required no human input. While this method removed the keypoint variance from the result, the resulting metric was only useful in a relative sense, since it had little physical meaning.

This paper introduces a new metric, the lighting variance ratio. The lighting variance ratio is the fraction of the total variance that is accounted for by illumination changes. As this value decreases the result becomes dominated by other sources of variance, typically intra-class variance. This ratio is more useful than an absolute measure of the variance due to illumination changes, since an absolute measure gives no indication of how it compares to typical descriptor variance. This metric is applicable in a relative sense to determine if a descriptor is more illumination invariant than another, and in an absolute sense to determine if a descriptor is invariant enough.

The effect of algorithm parameters is explored in detail, and it is shown that some parameters can be effectively removed, significantly improving the stability of the result over previous work. Additionally, the previous analysis is extended to a larger variety of descriptors, and is compared over different cameras for the same dataset in order to give a better understanding of the effect of camera properties on the result.

The remainder of the paper is laid out as follows. Section 2 introduces the method for computing the lighting variance ratio. Sections 3, 4 and 5 introduce the descriptors, cameras and datasets respectively. Section 6 details results and Section 7 provides conclusions.

2 Lighting variance ratio

The lighting variance ratio is defined as the fraction of the total variance of a descriptor which is due to illumination changes. This is estimated by utilising a given timelapse dataset. Variance due to illumination changes is expected to be gradually changing over time, and so within a local window of a given size it should be approximately constant. Our method makes no other assumptions about the nature of the variation due to lighting changes other than that it is approximately constant over these small windows. As this window size goes to zero this approximation becomes exact.

Mikulik [Mikulík *et al.*, 2010] demonstrated that for the standard SIFT descriptor changes due to illumination are not linear, and lie on a complex manifold. They showed that the L2 norm between any two given descriptors is therefore only an accurate estimate of their similarity over small differences. These nonlinearities in data will corrupt any mean and variance measures taken from the data.

Isomap [Tenenbaum *et al.*, 2000] is employed first on the data to recover linearised data. Isomap maintains the scaling of the data, so mean and variance measures on the linearised data are still meaningful assuming the mapping is performed correctly.

Isomap is a variant of multidimensional scaling (MDS) where the distance between two points is approximated by a geodesic distance. It makes the same assumptions outlined above, in that the L2 norm is assumed to be a valid distance metric only over local neighbourhoods, and utilises these neighbourhoods to build a connectivity graph for the geodesic distance. It makes no assumptions about the shape of the manifold, and is asymptotically guaranteed to approach the behaviour of the actual embedded manifold as the amount of data increases.

Neighbourhoods for Isomap are constructed using their k-nearest neighbours (kNN) since a fully connected graph is required for this analysis. As the amount of data available goes to infinity this could theoretically be replaced with a maximum distance criterion, however this will have problems in practise due to the quantisation introduced by the camera sensors. The effect of the neighbourhood size is explored in more detail in Section 6.2.

Once a linear mapping of the data has been recovered using Isomap, a number of statistics for the data at each dimensionality are determined. The first of these is the local covariance of the D-dimensional mapping of the data

$$\Sigma^L(D,\Delta) = \mathrm{E}\big[\mathrm{Cov}[X_i(D,\Delta)]\big]$$

where $X_i(D)$ is the ith overlapping window of the D-dimensional mapping of the data, and Δ is the window size in time. This provides an estimate of how much the local data varies around its sample mean. The time range spanned by the ith window is given by

$$T_i = [t_0 + \delta i, t_0 + \delta i + \Delta)$$

where δ is the time difference between the captured imagery, in this case 30 seconds, and t_0 is the time of the first image capture. Choosing δ to be as small as possible in this way produces the maximum number of overlapping windows, and as such produces the most accurate result.

The second statistic determined is the global covariance of the D-dimensional mapping of the data

$$\Sigma^{\mathrm{G}}(D,\Delta) = \mathrm{Cov}\big[\mathrm{E}[X_i(D,\Delta)]\big]$$

It is expected that the illumination variance will be captured mostly by the global covariance, whereas the local covariance will capture noise variables, such as sensor noise, and other local variations. This is based on the assumption that illumination is constant over the duration of the local window. The expectation of the local window is then the effect of the local illumination variables, while the variance is due to other causes. Taking the variance of the local illumination effects then gives an estimate of the variance due to illumination over the dataset.

These two measures are combined to produce an estimate of the D-dimensional lighting variance ratio

$$L(D,\Delta) = \frac{\|\Sigma^{\mathrm{G}}(\mathrm{D},\Delta)\|}{\|\Sigma^{\mathrm{G}}(\mathrm{D},\Delta) + \Sigma^{\mathrm{L}}(\mathrm{D},\Delta)\|}$$

Since it is expected that Σ^L is positive definite, $0 \leq L(D,\Delta) \leq 1$. Smaller values indicate less dependence on the time of day, and indirectly lighting variables. For $L(D,\Delta) > 0.5$, illumination variance is the most significant source of variance in the result. The magnitude is determined using the spectral norm

$$\|\Sigma\| = \max_{|x|\neq 0} \frac{|\Sigma \boldsymbol{x}|}{|\boldsymbol{x}|}$$

where $|\boldsymbol{x}|$ indicates the L2 norm of a vector. This can alternatively be formulated in terms of the maximum eigenvalue, λ_0

$$\|\Sigma\| = \sqrt{\lambda_0(\Sigma^T\Sigma)}$$

where Σ^T indicates the conjugate transpose of Σ.

The value of $L(D,\Delta)$ is monotonically increasing with D, however to an asymptotic limit. Hence

$$L(\Delta) = \lim_{D\to\infty} L(D,\Delta) \approx L(D_{max},\Delta)$$

where D_{max} is the number of dimensions of the original descriptor, since this dimensionality is maintained through Isomap. In practice $L(D,\Delta) \approx L \;\forall\; D > \widehat{D}$, where $\widehat{D}$ is the dimensionality of the underlying data. Smaller values of $L(\Delta)$ are more desirable, since they exhibit lower amounts of lighting variance.

As discussed previously, the lighting variance measure becomes most accurate as the window size Δ goes to zero. Ideally

$$L = \lim_{\Delta\to 0} L(\Delta)$$

This limit is not directly calculable since at a window size of zero the variance and mean cannot be calculated. Instead, the value of $L(0)$ is estimated by extrapolating the behaviour of $L(\Delta)$ for small values of Δ. This is done by fitting a curve of the form

$$L(\Delta) \approx \frac{\Delta + A}{B\Delta^2 + C\Delta + D}$$

where A, B, C and D are parameters of the fit. It can be shown that this form follows the expected behaviour of the function, and results show that this form follows the behaviour of the actual function.

Additionally, it was found during testing that for some of the descriptors fitting this function become numerically unstable, and extrapolated poorly. To combat this, an adaptive algorithm was wrapped around this fitting to penalise large coefficients in the fit while attempting to keep the residual error on the fit within reasonable bounds. This process was found to significantly increase the stability of the fit while demonstrating very little degradation of the residual error on the fit.

Camera	Resolution	Colour	Pixel size	BPP	Dynamic range	SNR (max)	Electron well size
Photonfocus MV1-D1312IE-40-G2	1312×1082	N	8.0 μm	12	120 dB	50 dB	90k
Point Grey GS3-U3-23S6C-C	1920×1200	Y	5.86 μm	10	73 dB	45 dB	32k
iDS UI-5240CP-C-HQ	1280×1024	Y	5.3 μm	8	57 dB	40 dB	12k
Microsoft Lifecam Cinema HD	1280×720	Y	3.0 μm	8	69 dB*	39 dB	13k

Table 1: Summary of the cameras and their capture properties. The dynamic range for the Lifecam was quoted as 69dB at 8x gain, so isn't directly comparable to the others. It is unclear how its dynamic range compares to the others.

3 Feature descriptors

In this work the standard SIFT [Lowe, 2004] and SURF [Bay *et al.*, 2006] descriptors are compared, as well as a number of their variants. These are compared to a baseline of the underlying image data, taken as the RGB values in a 19 by 19 block centred on the keypoint location, and the greyscale data in the same block size.

The same set of fixed keypoint locations were used for all descriptors. The scale was chosen so that each of the descriptors was calculated on the same region of interest (ROI). The rotation of SIFT and SURF were detected from the image, except where upright descriptors were used. The upright variant of SIFT and SURF assumes a fixed rotation of zero.

Since all of the descriptors are calculated on the same support region, it is expected that the block RGB descriptor will provide an estimate of the true variance of the data that the descriptors were calculated from. This gives a constant point of comparison to show any improvements for other descriptors.

Additionally, some of the colour-based variants are compared, as discussed by [Van de Sande *et al.*, 2010]. In this work RGB-SIFT, Opponent-SIFT, and the equivalents in SURF are also compared. RGB-SIFT is a concatenation of the SIFT descriptors calculated on each of the red, green and blue channels independently. This gives it a dimensionality of $128 \times 3 = 384$. Similarly, Opponent-SIFT is the concatenated SIFT descriptors from each of the channels in a CIE-Lab colour converted version of the original image. RGB-SURF and Opponent-SURF are the direct analogues of these descriptors using the SURF descriptor.

SURF descriptors were calculated using a modified version of the OpenSURF MATLAB toolbox[1]. SIFT descriptors were calculated using the VLFeat MATLAB toolbox[2].

Table 2 provides a summary of the descriptors that were investigated in this work. Binary descriptors were not considered in this work due to the difficulty in applying Isomap to them. Future work will aim to include binary descriptors.

4 Cameras

In this work the effect of different cameras on the lighting variance ratio is also investigated. This enables a better understanding of the contribution of sensor noise and dynamic range to the lighting variance ratio.

Table 1 summarises the different cameras used in this work and their important characteristics. Each of the cameras has different dynamic ranges and SNR. This is due to the large differences in the size of the electron wells as well as differences in quantum efficiencies.

The Point Grey and iDS cameras were set to automatically control gain and exposure to control the average image intensity to a given value. The Lifecam utilised its default auto-exposure and gain settings. The exposure and gain of the Photonfocus camera were not adjusted – since the dynamic range was significantly larger than the brightness change from exposure control, there was no noticeable change in the imagery.

Each of the cameras was connected to a data

Descriptor	Colour	Dimensions	Rotation invariant
Block RGB	RGB	1083	No
Block Mono	Mono	361	No
SIFT	Mono	128	Yes
SURF	Mono	128	Yes
U-SIFT	Mono	128	No
U-SURF	Mono	128	No
RGB-SIFT	RGB	384	Yes
RGB-SURF	RGB	384	Yes
Opponent-SIFT	CIE Lab	384	Yes
Opponent-SURF	CIE Lab	384	Yes

Table 2: Descriptors compared in this work. The colour listed is the colour space that the descriptor is calculated on. Rotation invariant descriptors are those that detect their rotation from the image and normalise the descriptor about this point.

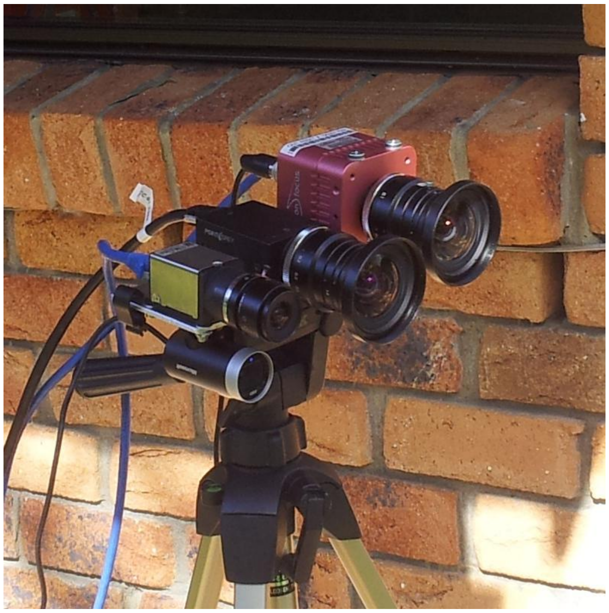

Figure 1: The dataset acquisition platform. The cameras are, top row, left to right: iDS UI-5240-CP, Point Grey GS3-U3-23S6C-C, Photonfocus MV1-D1312IE-40-G2. Bottom: Microsoft Lifecam Cinema HD

[1] Available from http://www.chrisevansdev.com/computer-vision-opensurf.html

[2] Available from http://www.vlfeat.org/

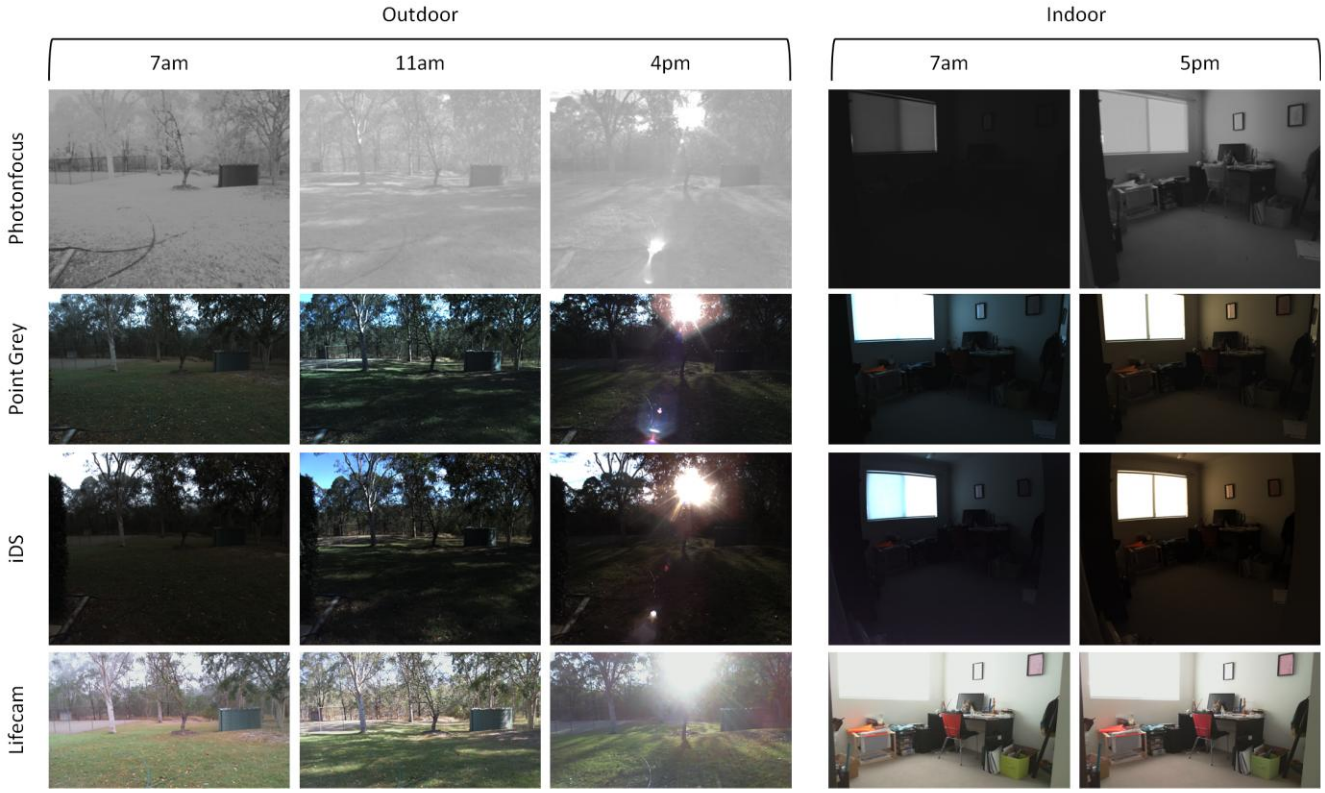

Figure 2: Datasets. Left, outdoor dataset, right, indoor dataset. Images are from periodic points throughout the dataset, highlighting lighting variance on the result. Imagery is, top to bottom, the Photonfocus camera, the Point Grey camera, the iDS camera and the Lifecam.

acquisition rig which ensured that they were pointing the same direction. Each was connected to the same machine, which would save the first image that arrived from the cameras every 30 seconds. Figure 1 shows the camera mounting.

The sensor manufacturer for the Microsoft branded webcam quotes the dynamic range for this camera as 69 dB at 8 times gain, meaning that its dynamic range isn't likely directly comparable to the other cameras. In our experiments it appeared by inspection to be on a similar order of magnitude to that of the iDS camera.

In the case of the first three cameras, the aperture and focus were adjusted such that an identical object at approximately 10m was in focus in each.

The cameras all have varying fields of view. While every effort was made to ensure keypoints were in similar locations in each of the different cameras, it is expected that there will be some source of discrepancy associated with this. Additionally, each pixel on each camera doesn't represent the same area in physical space. As a consequence the support region for the different cameras will be somewhat different, again leading to some difference in results. It is expected that while this may reduce the ability to compare between cameras, some inference will still be possible.

5 Datasets and sample choice

In order to better understand the effect of the actual illumination variance on the lighting variance ratio, two different environments are investigated. These are an outdoor and indoor environment. The outdoor environment exhibits a high degree of lighting variance, whilst the indoor environment is much more controlled in terms of lighting and so has a lower lighting variance overall. This is mitigated by the exposure and gain control of the cameras.

The outdoor dataset covers times from complete darkness prior to dawn through to complete darkness after sunset. The day in question had a large variety of weather conditions, including fog early in the day, sunny at various points, partly cloudy and heavy cloud. There was also a brief shower in the evening.

The indoor dataset covers almost an identical time period as the outdoor dataset. It was taken in a cluttered study with no sources of indoor illumination. The only illumination was through the shuttered window. As can be seen in the imagery, the largest change in the imagery over the portion of the day where there was sufficient illumination is the colour of the lighting.

Figure 2 shows the different camera imagery as well as its evolution over the datasets.

For unknown reasons there were capture issues with some of the imagery from the Photonfocus camera. The problematic imagery was removed from the dataset. It is expected that this will have little to no effect on the result since all variances and means were calculated on a group of descriptors, and the problematic images were spaced throughout the data rather than being grouped in one part.

In the outdoor dataset, a set of 4 keypoints were selected which were all of the same semantic class grass. Similarly on the indoor dataset, a set of 4 keypoints was chosen with the same semantic class carpet.

Dataset	Date	Duration (hh:mm)	Frames
Outdoor	22 July 2014	12:26	1493
Indoor	31 July 2014	12:25	1491

Table 3: A summary of the datasets

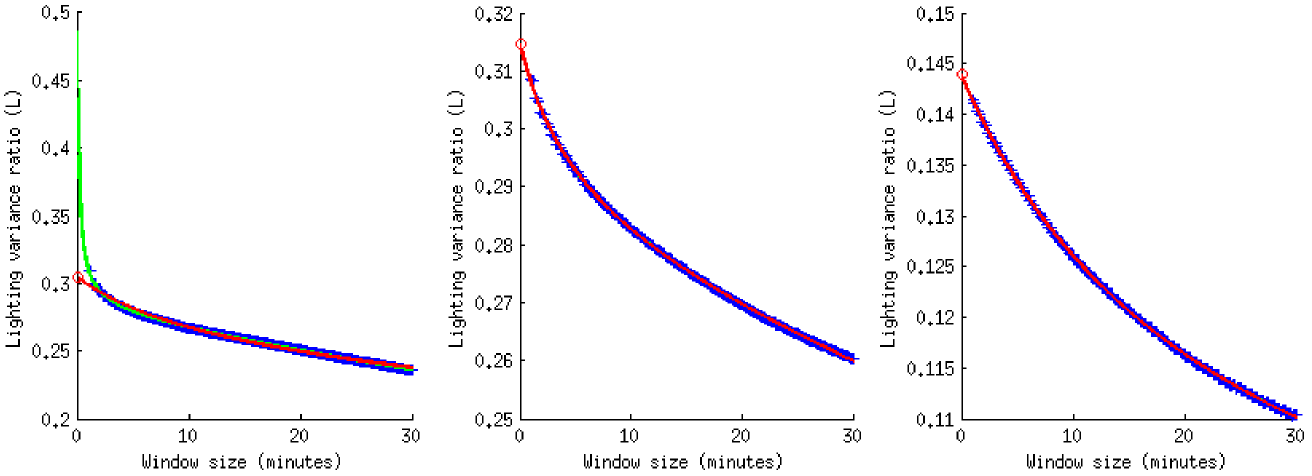

Figure 3: The curve fit for a series of different descriptors and cameras on the outdoor dataset. The blue crosses indicate the calculated values, the red line is the fit to the data after adjusting the limits, the green line is before adjustment. The red circle indicates the value used for the lighting variance estimate. The left-most figure indicates a situation in which the fit was numerically unstable due to the lack of information about the high-curvature components of the fit. The adaptive fit provides a conservative estimate of the lighting variance when the high curvature components cannot be reliably determined.

A consequence of this choice is that the local variance is effectively entirely intra-class variance, assuming the selected keypoints are indicative of the intra-class variance. This makes the interpretation of these results especially prevalent to classification problems in an absolute sense, however the relative ranking of the descriptors is valid for any matching problem. In this case, desirable values for lighting variance would be $L \ll 0.5$, so that intra-class variance is dominating the result. This would make classification based on these descriptors at the very least feasible.

6 Results

6.1 Local time window

The local time window is the size of the windows over which the lighting variables are assumed to be constant. As discussed previously, the lighting variance ratio becomes most accurate in the limit as the window size goes to zero. Figure 3 demonstrates the behaviour of the lighting variance ratio for small window sizes for a number of different descriptors.

As evidenced by the data, for most descriptors the behaviour of the lighting variance ratio as the window size decreases is well described by the fit. This suggests that the fit is valid for extrapolating the behaviour of the lighting variance to the limit.

Also of note is the fact that for some of the results the fitting became numerically unstable. In a few cases this variance was quite significant. This was in all cases due to the fact that there was insufficient data on the behaviour of the lighting variance close to zero, in that the high curvature components could not be accurately estimated. Figure 3 demonstrates this issue for one of the descriptors.

To combat this issue, an adaptive algorithm was used to limit the coefficients of the fitted line to be within reasonable bounds. These bounds were made as tight as possible without significantly reducing the R-squared coefficient of the fit. The optimal bounds are chosen to be those that produce an R-squared coefficient of 99% that of the unconstrained solution. In practice, this gave significantly improved estimation stability for unstable

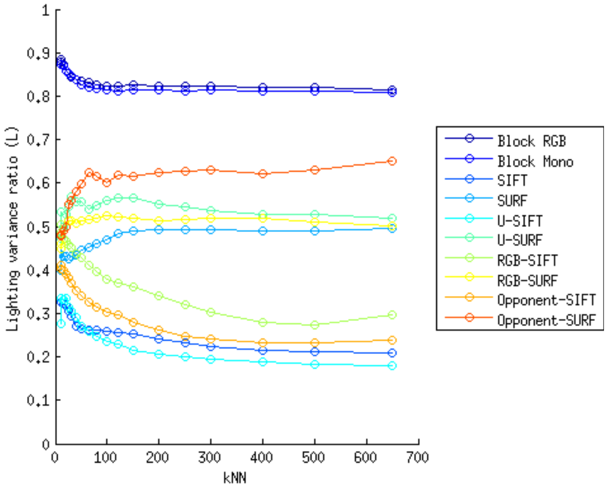

Figure 4: The effect of changing the nearest neighbours on the lighting variance ratio for the iDS camera on the outdoor dataset. In each case the lighting variance stabilises around kNN = 150.

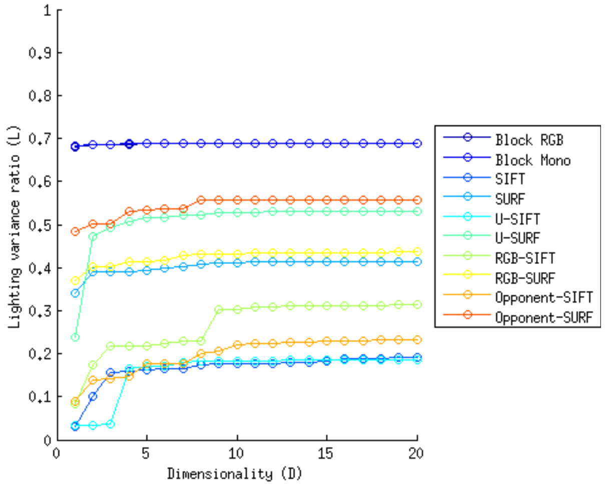

Figure 5: The effect of the dimensionality on the lighting variance ratio for the iDS camera on the outdoor dataset. In each case the lighting variance tends towards an asymptotic limit.

solutions, while having minimal impact on those that were already stable. Figure 3 shows that for typical situations where the fit is well defined there is little to no difference between the fit before and after this adjustment, whereas in the unstable case there is significant variation between the two, leading to a more accurate estimate in this situation.

6.2 Isomap nearest neighbours

The number of nearest neighbours used for constructing neighbourhoods is another important variable in the lighting variance ratio. It is important that the number of nearest neighbours is sufficiently large to generate a fully connected graph, in order for the linear mapping generated by Isomap to be valid, however not so large that it no longer accurately represents the underlying manifold.

Figure 4 shows the result of varying the kNN for a number of different descriptors and cameras in the outdoor dataset. It can be seen that there is a distinct elbow in the lighting variance ratio as the number of nearest neighbours increases above a minimum threshold, which for this data appears to be in the range of 50 – 150 nearest neighbours.

Below this threshold, although the graphs were found to be fully connected, there is clearly some incorrect construction of neighbourhoods, leading to an over-estimation of the lighting variance ratio.

These results suggest that the optimal choice for the number of nearest neighbours is in the range of 100 – 200. It should be noted that this number is still a function of the size of the dataset (number of images) and the number of keypoints. In this case, this corresponds to approximately 1.7 – 3.3% of the number of descriptors for the analysis. It is not yet known whether this result generalises to larger datasets. For further analysis 150 nearest neighbours are used.

6.3 Lighting variance ratio

Figure 5 shows how the lighting variance ratio changes as the dimensionality increases. It can be seen that the lighting variance increases to an asymptotic limit within 10 dimensions for all descriptors. This suggests that the assumption that lighting increases to a limit past the dimensionality of the data is valid. It should be noted that these results are for a constant window size of 10 minutes, since the window size extrapolation step is carried out after the determination of the lighting variance.

Table 4 and Table 5 outline the results for the various descriptors for the outdoor and indoor datasets respectively. It can be seen that the U-SIFT, SIFT and U-SURF descriptors are typically the best performers. Of these, the performance of SIFT and U-SURF is highly variant, while the U-SIFT descriptor always gives nearly optimal results when compared to the other descriptors.

The results for the colour descriptors were far more mixed, making any conclusions difficult. A more in-depth analysis, looking at upright colour descriptors would likely give significantly improved performance, for reasons discussed more below.

The increase in lighting variance from SIFT to its

Descriptor	Lifecam	iDS	Point Grey	Photonfocus
Block Mono	0.9221	0.8144	0.7838	0.9866
SURF	0.3598	0.4884	0.3692	0.9054
U-SURF	0.3412	0.5657	0.1897	0.849
SIFT	**0.1181**	0.2516	0.2936	0.2449
U-SIFT	0.1324	**0.2155**	**0.1482**	**0.2126**
Block RGB	0.9235	0.8248	0.7894	-
RGB-SURF	0.3201	0.5189	0.3596	-
Opponent-SURF	0.3486	0.6146	0.5036	-
RGB-SIFT	0.2898	0.3606	**0.3093**	-
Opponent-SIFT	**0.2866**	**0.2798**	0.3204	-

Table 4: The lighting variance ratio for each of the cameras and descriptors for the outdoor dataset. Note that since the Photonfocus camera is monochrome, the colour descriptors couldn't be utilised.

Descriptor	Lifecam	iDS	Point Grey	Photonfocus
Block Mono	0.9449	0.5651	0.86	0.9949
SURF	0.0564	0.24	0.175	0.0497
U-SURF	**0.0316**	**0.1124**	0.1406	0.0185
SIFT	0.0748	0.1267	0.0815	0.1726
U-SIFT	0.0442	0.1366	**0.0399**	**0.0147**
Block RGB	0.9305	0.5555	0.8602	-
RGB-SURF	0.1213	0.366	**0.1348**	-
Opponent-SURF	**0.1124**	0.3422	0.5406	-
RGB-SIFT	0.2439	0.2515	0.4733	-
Opponent-SIFT	0.4783	**0.163**	0.429	-

Table 5: The lighting variance ratio for each of the cameras and descriptors for the indoor dataset. Note that since the Photonfocus camera is monochrome, the colour descriptors couldn't be utilised.

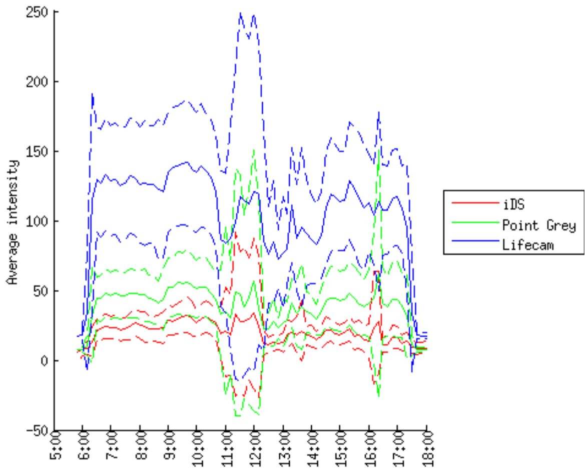

Figure 6: The local average intensity of three of the cameras for the outdoor dataset. Solid lines indicate the mean, dotted is 3 standard deviations from the mean.

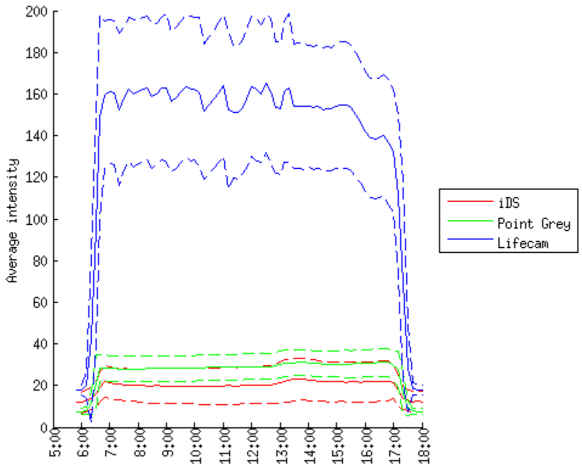

Figure 7: The local average intensity of three of the cameras for the indoor dataset. Solid lines indicate the mean, dotted is 3 standard deviations from the mean.

colour variants (and similarly with SURF) can be attributed to the fact that the orientation of the descriptor parts was not enforced to be in the same direction, and as such introduces a new source of variation above that of the mono variants. This is especially prevalent in the indoor dataset where the rotation estimate may be quite sensitive to noise due to the lower texture appearance.

The reduction in variance from rotation invariant descriptors to their upright counterparts is also observed, and is due to the fact that a source of variance from the data is removed. This would suggest that any additional information that can be used to reduce the variance in the keypoint such as fixed scale or rotation can lead to significant reductions in the variance of the keypoint. This is exacerbated by utilising pre-selected keypoints as opposed to automatically detected keypoints, since they likely do not correspond to salient image points, which leads to poor rotation estimation. This is a limitation of the current strategy.

The results show that the Photonfocus camera has by far the highest lighting variance ratio in both datasets. This is a consequence of the fact that the camera didn't utilise exposure or gain control, and so the DC lighting changes showed much clearer in this data. Combined with the fact that this camera has the best SNR, it is expected that it demonstrates the highest lighting variance ratio.

Of the remaining cameras, the Lifecam was found to have the highest lighting variance ratio. Figure 6 and Figure 7 show that this is due to the fact that this camera exhibited by far the highest range of intensities over the day in both datasets, due to the fact that its capture settings were not controlled in the same manner as the other cameras.

The iDS and Point Grey cameras exhibited similar levels of the lighting variance ratio for the outdoor dataset, while the iDS had a significantly lower value for the indoor dataset. The data for the Point Grey camera shows that it had a brighter image in the region of interest, leading to it having a higher global variance. This suggests that its local variance was also increased relative to that of the iDS camera for the outdoor dataset, which is demonstrated in Figure 6 and Figure 7. This is thought to be due to the increased dynamic range of the Point Grey camera, which had significantly more information is this region of the image, leading to higher amounts of local variance.

Conversely, on the perceptually uniform indoor dataset the Point Grey camera exhibited significantly smaller local variance than the iDS camera, as expected by the SNR of the cameras. This fact combined with the increased brightness of the Point Grey camera contributed to a significantly higher lighting variance ratio for this dataset.

When considering the lighting variance ratios in an absolute sense, as discussed previously $L \ll 0.5$ would ensure that the intra-class variance is much more significant than any lighting variance. Taking an arbitrary condition of $L \leq 0.25$, only the U-SIFT descriptor would be considered lighting invariant enough for all cameras on the outdoor dataset, while each of the SIFT, SURF and their upright variants were lighting invariant enough on the indoor dataset. This suggests that for complex outdoor scenes, the U-SIFT descriptor would need to be the descriptor of choice, however for more structured and uniform indoor scenes there is more choice. One could, for example, justifiably utilise the SURF descriptor for the increased computation and comparison speeds it affords.

Compared to previous work by other authors, Valgren [Valgren and Lilienthal, 2007] concluded that the U-SURF descriptor was the best descriptor for their datasets. This conclusion was based on the favourable processing time compared to SIFT, percentage of correct matches and number of correct matches. They noted, however, that the SIFT descriptor gave the largest number of correct matches despite having a lower percentage correct matches, but due to the different keypoint detector it is difficult to directly compare these result. They also noted that upright descriptors tended to be more lighting invariant, a result that agrees well with the results presented here.

Van de Sande [Van de Sande *et al.*, 2010], conversely, concluded that SIFT and most of its variants are robust to most lighting changes, out-performing SURF and its variants, which agrees well with the results presented here. They found that colour SIFT descriptors are more discriminative over lighting changes than greyscale SIFT descriptors, however classes with smaller lighting variations actually performed worse with colour SIFT descriptors. They concluded that Opponent-SIFT was the best descriptor in the absence of the ability to perform the analysis on data specific to the application, however the plain SIFT descriptor was also a good choice.

The discrepancy between these results and those presented by Van de Sande are likely due to the unconstrained extra rotations in the colour SIFT descriptors. It is expected that fixing the rotation, or pre-detecting it from the greyscale image would give significantly improved results in this regard.

7 Conclusion

This paper has presented a novel metric for evaluating the lighting variance of various descriptors, called the lighting variance ratio. It has been shown that this corresponds well to the fraction of the variance in the descriptors that is associated with lighting changes.

Preliminary results with this metric show that it is useful both in a relative sense for comparing descriptors, and in an absolute sense for determining if a descriptor is invariant enough for a particular environment and camera.

Factors such as automatic exposure and gain control have been shown to significantly reduce the effects of lighting variance when employed effectively, as can utilising additional information to constrain the keypoint; fixing the keypoint location, scale or orientation will reduce the lighting variance ratio of a descriptor.

While certain colour descriptors can reduce the lighting variance ratio compared to their greyscale counterparts, typically they increase the lighting variance ratio due to the increased degrees of freedom in the keypoint arising from the additional rotations.

Variants of SIFT and SURF were most lighting invariant in various situations, however the most consistently lighting invariant descriptor was upright SIFT. For this reason, in the absence of the ability to perform this analysis for a particular application the upright SIFT descriptor is recommended for applications where lighting variance is significant. For other applications where lighting variance may be less significant other descriptors may suffice, and in fact may be more desirable for other properties such as computation and matching time.

There were a number of issues with the dataset capture, as well as the possibility of errors being introduced into the result through the differing field of view of the cameras. Future work will attempt to remove these sources of error from the result, to better compare the different cameras.

The number of keypoints per image is the only remaining parameter without a clear understanding of its effect on the result. In this work its value was chosen arbitrarily. Future work will aim to provide a more in-depth analysis of its effect on the result. Future work will also investigate binary descriptors such as BRIEF, BRISK and ORB.

Acknowledgements

This work was supported in part by the Australian Research Council Linkage Project LP110200375 “Robotics for Zero Tillage Agriculture” awarded to the Queensland University of Technology, SwarmFarm Robotics and the Australian Centre for Field Robotics.

Computational resources and services used in this work were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia.

References

[Bay *et al.*, 2006] Herbert Bay, Tinne Tuytelaars, and Luc Gool. SURF: Speeded Up Robust Features. *Computer Vision – ECCV 2006*, Springer Berlin Heidelberg. 3951: 404-417, 2006.

[Kim *et al.*, 2007] Dongshin Kim, Sang Min Oh, and J. M. Rehg. Traversability classification for UGV navigation: a comparison of patch and superpixel representations. *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, 2007.

[Lowe, 2004] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60(2): 91-110, 2004.

[McManus *et al.*, 2011] C. McManus, P. Furgale, and T. D. Barfoot. Towards appearance-based methods for lidar sensors. *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011.

[Mikulík *et al.*, 2010] Andrej Mikulík, Michal Perdoch, Ondřej Chum, and Jiří Matas. Learning a Fine Vocabulary. *Computer Vision – ECCV 2010*, Springer Berlin Heidelberg. 6313: 1-14, 2010.

[Ranganathan *et al.*, 2013] Ananth Ranganathan, Shohei Matsumoto, and David Ilstrup. Towards illumination invariance for visual localization. *Robotics and Automation, 2013. Proceedings. ICRA '13. IEEE International Conference on*, 2013.

[Ross *et al.*, 2013] Patrick Ross, Andrew English, David Ball, Ben Upcroft, Gordon Wyeth, and Peter Corke. A novel method for analysing lighting variance. *Australian Conference on Robotics and Automation*, Sydney, Australia, 2013.

[Tenenbaum *et al.*, 2000] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* 290(5500): 2319-2323, 2000.

[Valgren and Lilienthal, 2007] Christoffer Valgren, and Achim Lilienthal. Sift, surf and seasons: Long-term outdoor localization using local features. *Proceedings of the European conference on mobile robots (ECMR)*, 2007.

[Valgren and Lilienthal, 2010] Christoffer Valgren, and Achim Lilienthal. SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments. *Robotics and Autonomous Systems* 58(2): 149-156, 2010.

[Van de Sande *et al.*, 2008] Koen E. A. Van de Sande, Theo Gevers, and Cees G.M. Snoek. A comparison of color features for visual concept classification. *Proceedings of the 2008 international conference on Content-based image and video retrieval*. Niagara Falls, Canada, ACM: 141-150, 2008.

[Van de Sande *et al.*, 2010] Koen E. A. Van de Sande, Theo Gevers, and Cees G.M. Snoek. Evaluating Color Descriptors for Object and Scene Recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32(9): 1582-1596, 2010.