# Texture and Distinctness Analysis for Natural Feature Extraction

**Kai-Ming Kiang, Richard Willgoss**
**School of Mechanical and Manufacturing Engineering,**
**University of New South Wales, Sydney, NSW 2052, Australia.**
**kai-ming.kiang@student.unsw.edu.au, r.willgoss@unsw.edu.au**
**Alan Blair**
**School of Computer Science and Engineering,**
**University of New South Wales, Sydney, NSW 2052, Australia.**
**blair@cse.unsw.edu.au**

## Abstract

One of the basic requirements for autonomous navigation in an unexplored and often complex environment is to be able to lock on to natural features. This paper presents a method for extracting features distinctive enough to navigate with. The method consists of three parts. Firstly, it selects a set of interest points from the images which are invariant to most changes in conditions; secondly, it analyses the texture distribution of the local interest regions around interest points selected; thirdly, it picks out distinctive features from the original set of interest points. The method has been implemented within a SLAM framework designed for use in a texture-rich environment such as the Great Barrier Reef. The results have shown that this method has significant advantages over other widely used methods in this specific environment. The speed of implementation is faster and the number of features needed to process is reduced.

## 1 Introduction

Autonomous navigation in an unexplored environment is more challenging than in one that is controlled because of extra effort needed to make sense of sensor inputs. In particular, underwater environments are mostly unexplored and do not have GPS access. Therefore navigation in these environments requires the use of methods such as Simultaneous Localization and Mapping (SLAM) [Csorba, 1997; Williams *et al.*, 2002]. However, most existing SLAM algorithms have relied on point-based artificial landmarks that do not exist in an unexplored environment. SLAM can be unreliable if natural landmarks are used when they lack descriptive representation.

Developments from computer vision research extract features with representations that are invariant to scaling, distortion and perspective [Carneiro and Jepson, 2002; Mikolajczyk and Schmid, 2002; Tuytelaars and Van Gool, 2000]. These developments could potentially be used for robot navigation and are already capturing attention from the robotic communities [Kragic and Christensen, 2005]. In particular, Scale Invariant Feature Transformation (SIFT) [Lowe, 2004] was reported as a method robust in representing features. Its descriptors were claimed to be invariant under changes in scale, rotation, shift and illumination conditions.

A performance test comparing different feature extraction methods was reported by Mikolajczyk and Schmid [2005] that indicated SIFT generally performed the best amongst these methods. Moreover, there was a modification to SIFT using principle component analysis that improved its performance further [Ke and Sukthankar, 2004].

However, as these methods were mostly designed for non real-time object recognition purposes, computational efficiency may not have been the major concern. The methods tended to generate a large number of features that maximized accuracy and stability. For real-time SLAM applications, it is computationally infeasible to compare such large sets of features from a series of images that have been captured.

We presented Texture Analysis (TA) [Kiang *et al.*, 2004] for feature extraction purposes that was designed to improve performance speed. The descriptors for TA represented the frequency distribution of the local interest region of an interest point. This method is potentially an appropriate choice of descriptor for a texture-rich environment such as found at the Great Barrier Reef because there are many textures to work with. However, TA is designed to be a generic solution and can also be used for navigating a mobile robot through a street, guiding an aircraft and negotiating wooded scenery. An improved version of TA is now presented in this paper.

Besides requiring representative features, SLAM also requires a selection method that can minimise the set of features picked for similarity matching. For this reason Distinctness Analysis (DA) was devised as a technique to minimize the number of features selected [Kiang *et al.*, 2005]. DA can extract the distinctively rare features from those initially selected that minimises the number that are needed for processing.

In this paper, besides presenting the improved version of TA, further work on combining TA and DA is presented. Moreover, results are presented that have been

obtained from the implementation of the combined TA/DA method in a SLAM framework for use in an underwater environment.

## 2 Interest Point Selection

The method of feature extraction presented here consists of three stages, namely Interest Point Selection, Texture Analysis (TA) and Distinctness Analysis (DA). This section describes how interest points are selected.

Interest points should be invariant to rotation, shift, scale, illumination and affine transformation such that, when the examining region is to be analysed again under different conditions, the same points would be evident. Two ways of selecting interest points have been reported are Harris Corner [Harris and Stephen, 1988] and extrema in Difference of Gaussian (DOG) [Lindeberg, 1994], which are the most widely used methods. But Mikolajczyk and Schmid [2005] have pointed out that the method of selecting interest points is independent of the choice of method for generating feature descriptors. In this paper, extrema of DOG, the method used in SIFT [Lowe, 2004], has been chosen for interest point selection and will now be briefly described.

A DOG image is computed by convolving the original image $I(x,y)$ with a Gaussian function of standard derivation σ, to obtain a blurred version of the original image $B(x,y,\sigma)$. That is:

$$B(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{1}$$

where * is the convolution operator in x and y plane. The Gaussian function $G$ is applied to the resulting image $B$ sequentially obtaining $B(x,y,k\sigma)$ where $k\sigma$ represents the number of convolutions applied to the original image.

The DOG image is then defined to be the difference of these two images. That is:

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma)) * I(x, y) \\ &= B(x, y, k\sigma) - B(x, y, \sigma) \end{aligned} \tag{2}$$

Note that the variable σ depends on the complexity of the image.

Besides applying Gaussian convolution, which results in a blurring effect, down-sampling is also applied. Images are downsized in ratios of two. A pyramid of DOG images is then constructed. From the pyramid, extrema can be selected directly by comparing each pixel with its neighbouring pixel in spatial and scale domain for a preset radius around the pixel. These DOG extrema become the interest points.

## 3 Texture Analysis

After selecting interest points, the local region of each point was treated as a feature candidate that required further analysis. The local regions were each limited to 32 x 32 pixel rectangular segments. Even though the square segments were chosen to simplify Fourier Transform calculation, a Hanning window was applied on the transformed regions making them approximately circular and centred at an interest point such that invariant properties were preserved.

Different textures show up as different patterns in the Fourier Transform. TA used the Fourier Transform of the local regions of the interest points as a basis of its choice for descriptors. This choice is suitable for representing features from images of a texture-rich environment.

The Discrete Fourier Transform adjusted by a Hanning window was calculated as follow:

$$S_{\mathbf{m}}^{pg}(\omega) = \frac{1}{\|W\|}\left|\sum_{\mathbf{k}\in W} q(\mathbf{k})x[\mathbf{k}]e^{-j\omega^t\mathbf{k}}\right| \tag{3}$$

where W is the original local region and q(k) is the Hanning window function defined as:

$$\begin{aligned} q(k) &= \frac{1+\cos(\frac{\pi k}{\tau})}{2} && \text{for} |k| < \tau, \\ &= 0 && \text{otherwise.} \end{aligned} \tag{4}$$

The resulting transform, which is also an image, represented the distribution of frequency of the interest region. The Hanning window, besides generating a circular region, was also needed to smoothly correct the boundary effect of the Fourier Transform.

This transformed image was then partitioned into 26 useful regions discarding area containing sparse information. Firstly, it was divided into concentric semi-annuli one pixel wide. Then for the second and third smallest annuli, they were further subdivided into 4 and 8 angular sectors respectively. A diagram illustrating this partitioning scheme is shown in Figure 1.

The partitions were more densely distributed in the centre that represented the lower frequencies. The reason for this is that, in natural images, lower frequencies would nearly always contribute more than higher frequencies. A more detailed description of the lower frequencies is therefore important. Moreover, only half of the plane was needed to be analysed because Fourier Transforms are always symmetrical.

In order to have rotational invariance, the angular segmenting process for the inner circles was referenced to the mean gradient direction. The calculation process for this mean gradient direction was the same as in SIFT but without the need to resample the gradient image.

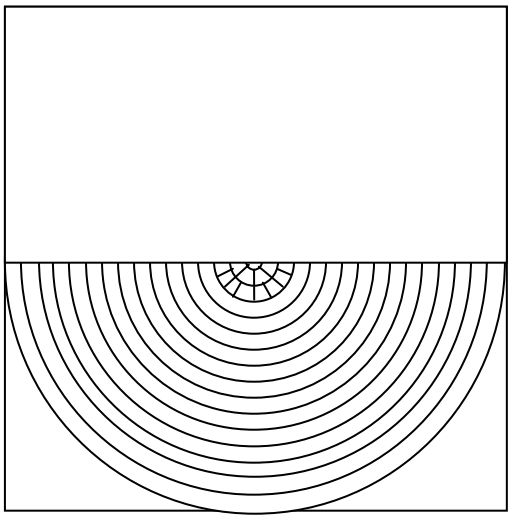

Figure 1: The 26 Partitions of the Frequency Distribution.

Each value in the Fourier Transform was a complex number. Adding magnitudes of each of these complex values within a partition gave the strength of that partition. By applying the calculation to all 26 partitions, a frequency distribution was obtained. The resulting 26 values that represented texture properties were then normalised and could then be used as the descriptors for a particular interest point.

# 4 Distinctness Analysis

## 4.1 The Probability of Occurrence

Numerous interest points were normally generated as features from the first selection process described in Section 2. In the literature, none of the feature extraction methods looked to minimizing this number after the interest points had been transformed into descriptors. This issue is addressed by DA proposed here and is described in this section. The term 'distinctness' has been used in reporting research referring to a special property of a particular type of interest point such as the extrema of DOG or Harris Corners. The special property usually refers to invariants in conditions and stability. It is however not related to the frequency of occurrence of such points within an image or an environment.

In these methods, the number of interest points is not necessarily minimized. Interest points such as the extrema of DOG are often common within an image. The number can be as high as thousands for a 640x480 pixel image. Usually, in object recognition, it is desirable to extract more rather than less interest points to enable robustness in matching. However, in real-time navigation, the computation time is a critical requirement. If all interest points are to be used as landmarks, since the computation time for most cases is proportional to $O(N^2)$, where $N$ is the number of state variables needed to represent the landmarks and the robot pose, the computation time is greatly increased. Therefore the need to minimize the number of interest points while, at the same time, not penalising the performance of recognition is the main objective.

However, at the raw pixel level of an image, it is difficult to find a certain type of point that rarely occurs and is invariantly stable. On the other hand, feature transformation provided a more expressive representation for describing each interest point. Hence, it is preferable to select the distinctive set of interest points at the TA level of abstraction.

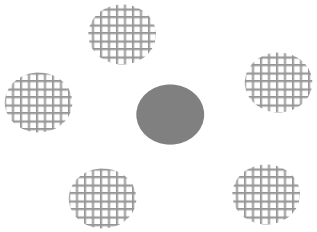

Figure 2: Simple diagram of a distinctive object among other less distinctive objects.

The question then arises as to how a few relevant features out of a potentially large set should be selected. For example, in Figure 2, it would be best to remember the centre object because it is the only one that is unique. If one selects any of the other objects that are similar to each other, it will be hard to distinguish between them later on.

Since the descriptors represent the features, they become elements of feature vectors in the descriptor parametric space. Distinctness can be judged from analysing and comparing these vectors. If we consider all of the descriptors in the feature vectors as independent random variables, the probability of occurrence for each feature can then be calculated by finding a model for its probability density function. In the simplest case, the distribution could be approximated to a multi-dimensional Gaussian. This approximation is to be justified in Subsection 7.2. The probability of occurrence of a feature can then be calculated as follows:

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^m \det(\mathbf{C})}} \cdot \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \tag{5}$$

where $\boldsymbol{\mu}$ is the mean vector:

$$\boldsymbol{\mu} = \frac{1}{|R_j|}\sum_{\mathbf{n}\in R_{j-1}} \mathbf{x}[\mathbf{n}] \tag{6}$$

and $\mathbf{C}$ is the covariance matrix:

$$\mathbf{C} = \frac{1}{|R_j|}\sum_{\mathbf{n}\in R_{j-1}} (\mathbf{x}[\mathbf{n}]-\boldsymbol{\mu})\cdot(\mathbf{x}[\mathbf{n}]-\boldsymbol{\mu})^t \tag{7}$$

DA can then be made on the basis that the lower the probability, the more distinct a feature is judged to be.

## 4.2 Global Distinctness

DA is a process of minimizing the number of features while retaining stability of analysis. Stability refers to the ability to pick out the same feature invariant to any changes in shift, rotation, scale and illumination. However, since features detected in one image need not be the same as in subsequent images, DA must therefore range over many images to embody global distinctness. In doing so, it is then possible to select features that are both distinctive and likely to be found in multiple images in the environment captured at different times and locations.

Denoting the mean and covariance for the global distinctness by $\boldsymbol{\mu}_t$ and $\mathbf{C}_t$ respectively and by $\boldsymbol{\mu}_c$ and $\mathbf{C}_c$ for the current image, $\boldsymbol{\mu}_t$ is obtained and updated using the following formula:

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} \times \lambda + \boldsymbol{\mu}_c \times (1-\lambda) \tag{8}$$

where $\lambda$ is the innovation factor, which determines how much the system relies on history versus new data. $\mathbf{C}_t$ is obtained and updated using the following formula:

$$\mathbf{C}_{t(x,y)} = E(XY)_t - \boldsymbol{\mu}_{t(x)} \cdot \boldsymbol{\mu}_{t(y)}^T \tag{9}$$

where $E(XY)$ is the expectation value of the product of two dimensions X and Y, which can be calculated from:

$$E(XY)_t = E(XY)_{t-1} \times \lambda - E(XY)_c \times (1-\lambda) \quad (10)$$

$E(XY)_{t-1}$ and $E(XY)_c$ can be obtained by rearrangement of the previous formulae using $E(XY)$ as the subject with the appropriate $\boldsymbol{\mu}$ and $\mathbf{C}$.

Equations 8 and 9 are used for iteratively updating. To initialise $\boldsymbol{\mu}_t$ and $\mathbf{C}_t$ , they are assigned to be equal to $\boldsymbol{\mu}_c$ and $\mathbf{C}_c$ for the first input image. $\boldsymbol{\mu}_t$ and $\mathbf{C}_t$ require the system to run over a series of images in order to converge to the true global distinctness. A practical solution is to take a safe walk in the environment of interest before using that data for exploring more of the environment.

## 5 Matching features across images

Having obtained the distinctive set of features extracted from each image, it is then possible to match features across different images. Matching features requires calculating a notional distance between the two feature vectors. In this analysis, distance is defined as the Euclidean distance of the feature vectors. The matching strategy is defined by using a threshold function relating to the closest and second closest match of a particular feature as follows:

$$\|D_A - D_B\| / \|D_A - D_C\| < t \,. \quad (11)$$

where $D_A$ is a feature vector on image one; $D_B$ and $D_C$ are closest and second closest feature vectors from another image respectively. If the above inequality is true, $D_A$ is considered to be the same as $D_B$.

Since every feature is originally an interest point that is either a maximum or a minimum of DOG, this property could also be utilized for matching purposes. It is therefore beneficial to separately storing the maxima and minima and find matches only within the same type.

## 6 Experimental

A submersible vehicle (see Figure 3) which was used for capturing underwater images and acquiring sonar data simultaneously (courtesy of ACFR, University of Sydney, Australia) was chosen to be the source of images used in the present analysis. The configuration of the submersible was set such that the camera was always looking downwards onto the sea floor. This configuration minimised the geometrical distortion that could had been caused by different viewpoints. The vehicle acquired images and sonar data as it travelled underwater. Some of the images are shown in Figure 4. More details of the configuration of this vehicle and implementation can be found in [Williams and Mahon, 2004] where, using the sonar data and the images, the path that it travelled could be retrieved as shown in Figure 5. The original captured underwater images were transferred to an external computer for offline testing and to be available to research groups.

In this paper, TA and DA were written in C++, running code that was embedded in such a way that it could run as the front end to SLAM-controlled guidance. A sequence of approximately 3000 images was used for testing the efficacy of the present technique.

## 7 Results

In the following experiments, TA presented in Section 3 and DA presented in Section 4 are considered as two independent methods. For reference, some of the results presented here were compared with SIFT when applicable. TA is considered as an alternative to SIFT while DA is considered as an add-on to both TA and SIFT.

Figure 3: The underwater vehicle. (Courtesy of ACFR, University of Sydney, Australia)

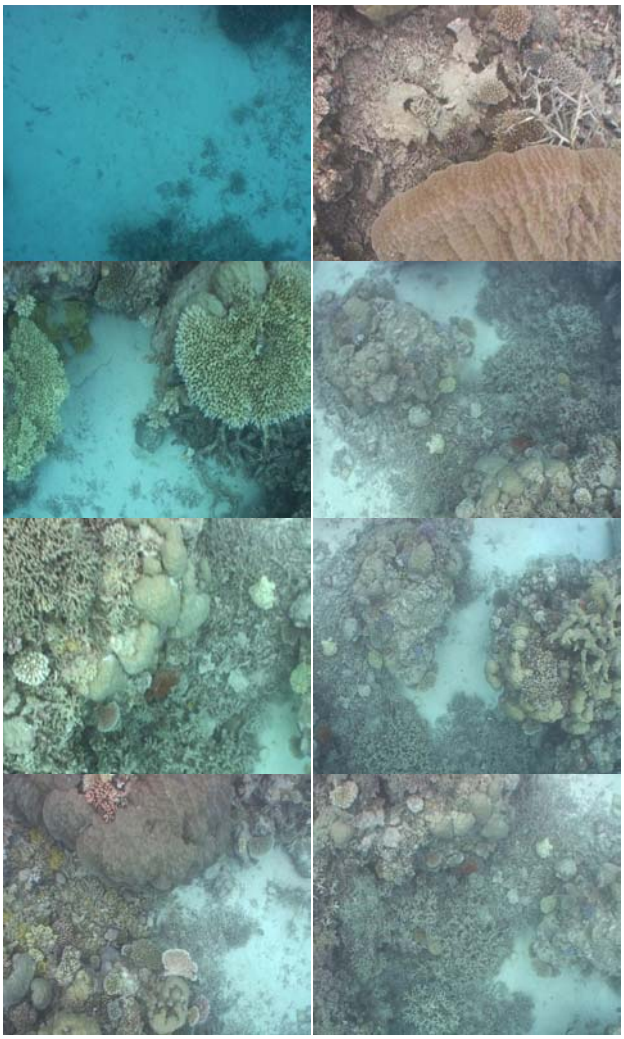

Figure 4: Some Images captured. (Courtesy of ACFR, University of Sydney, Australia)


Figure 5: A path of the underwater vehicle was plotted using the images and sonar data. The image frame numbers along the path is shown with selected images (Original data derived from [Williams and Mahon, 2004] and processed into graph form by the present authors).

### 7.1 Matching Capability

The Matching Capability of TA was tested first. Both SIFT and TA were used to find matches across the sequence of images. In order to have relatively few features on an image such that inspection of matching capability was kept efficient, DA was applied to both SIFT and TA. For example, one of these image pairs, which contained less than 10 feature matches, is shown in Figure 6. These features were matched using TA.

Figure 7 shows the results for a set of image pairs where the matching capabilities TA are tested. The results for SIFT are not shown because the false positive were zero. For TA, the percentage of correct matches was 98.5%. The disadvantage of SIFT was the time required for processing whereas the processing time for TA was approximately 1/3 of that for SIFT. This was due mainly to the lower dimensionality of TA. Provided the number of distinctive features could be limited to around 10 per image, a 98.5% accuracy per match in TA was an acceptable level of matching capability and virtually equivalent to the capability of SIFT.

In Figure 8, the eigenvalues of the 26 dimensional descriptors of TA and the 128 dimensional descriptors of SIFT are plotted and arranged in descending order. Since their dimensions were different in length, these dimensions were adjusted to fit the horizontal axis of the graph. This graph showed that the eigenvalues for TA were spread more evenly across descriptor space than for SIFT. This result suggested that the choice of the descriptors for TA conveyed more information than SIFT per descriptor.

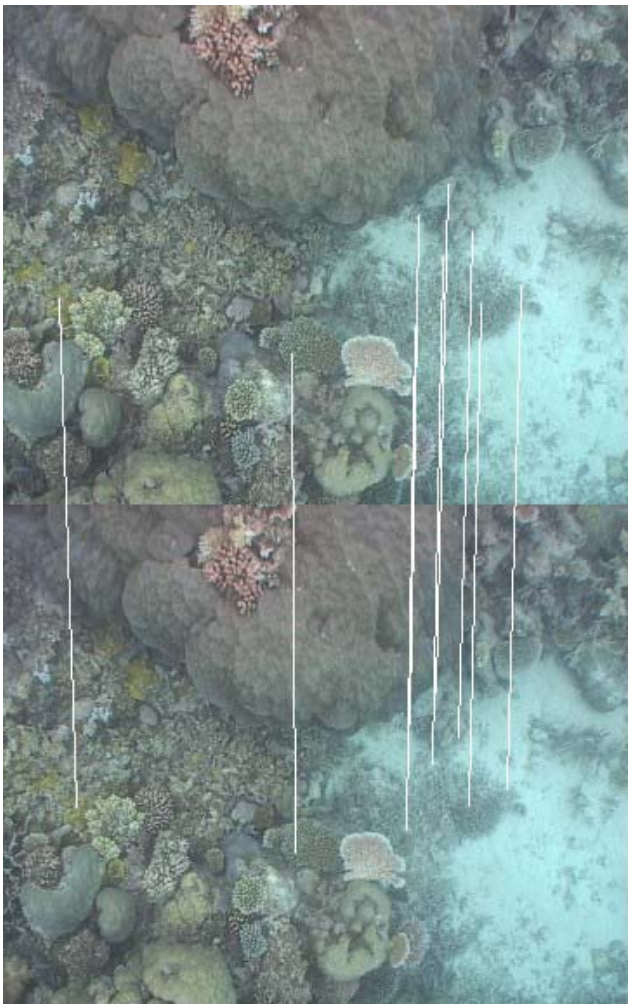

Figure 6: Feature Matching using TA.

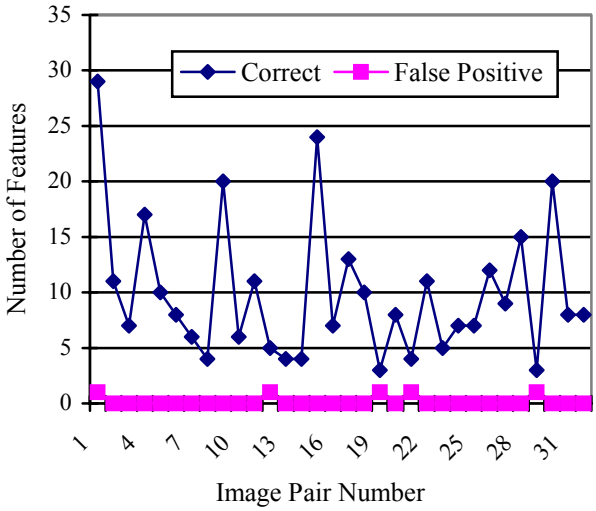


Figure 7: Correct Match and False Positive for Texture Analysis

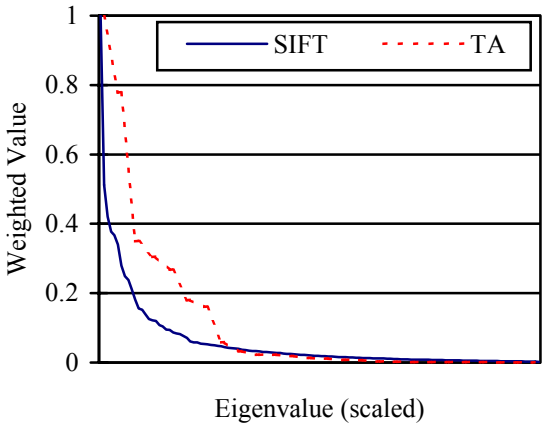


Figure 8: Relative Eigenvalue magnitudes of SIFT and TA features.

### 7.2 Distinctness

The key assumption for testing DA was that the feature descriptors used were based on Gaussian distribution. If the distribution was not Gaussian, DA may not obtain a valid distinctness of an extracted feature. It is important to note that, provided the distribution was unimodal, the distinctness calculated would not significantly deviate from that obtained if the distribution was Gaussian.

In Figure 9, the distribution for a large set (~13000) of texture features generated from different underwater images in the series is plotted. Only the top 3 principle directions of the distribution are shown limiting a principle component analysis to the most significant. It can be seen that the distribution was close to a Gaussian.

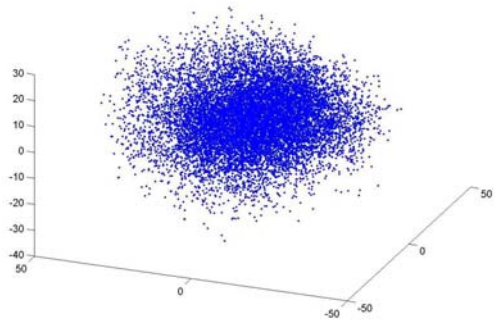

Figure 9: The Distribution of a large list of features plotted in the top 3 principle directions of Texture Analysis.

For comparison, the same features for a SIFT analysis are plotted in Figure 10. As can be seen, SIFT is a bimodal distribution. Such a distribution was caused by the deviation between the two types of features generated by SIFT, namely maxima and minima of DOG. By ignoring the sign during SIFT gradient calculations, the new distribution can be replotted and is shown in Figure 11. In so doing, the distribution of SIFT features, with the sign of the gradient ignored, turned out also to be approximately Gaussian.

Based on the results presented, the distribution for both TA and SIFT could be assumed to be Gaussian.. Therefore DA could potentially be applied to any local descriptor-based feature extraction technique.

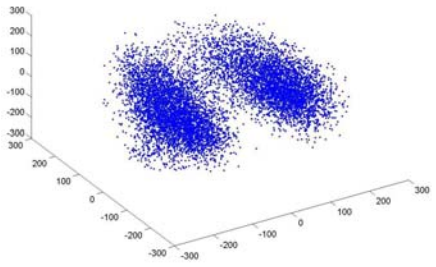

Figure 10: Distribution of SIFT features in the top 3 principle directions.

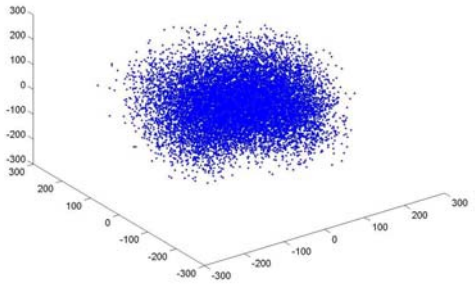

Figure 11: Distribution of SIFT (Gradient sign ignored) features in the top 3 principle directions.

### 7.3 Stability

A final test was conducted to check on the stability of chosen features. It is re-emphasized that by stable, we mean that the same feature should be picked out invariant to any changes in shift, rotation, scale and illumination.

A new series of image pairs were used and to which TA was applied. These image pairs contained overlapping regions such that DA had to range over images in which it was known features were continuous. DA was applied to each image and inspection made within overlapping regions to count the number of distinctive features that appeared within a few pixels in corresponding locations of the images. By comparing this number with the number of features that did not correspond in both of the images, a measure of stability was obtained.

Figure 12 shows the counts of features that were regarded as stable and unstable in overlapping parts of images. Approximately half of the features selected as distinctive in one image appeared in both images. This ratio is largely independent of the number of features detected and is significantly influenced by the stability of the initial selection process using extrema of DOG described in Section 2. The stability of the initial DOG selection process is of itself low and is a major factor in limiting the ratio shown in Figure 12. Based on this fact, the ratio was deemed a relatively high hit rate for tracking distinctive features through image sequences. In a SLAM context, it only requires to match a few stable features correctly across two images to track from image to image and eventually enable loop closure. It was concluded that the results showed significant promise for enabling map building in a SLAM context.

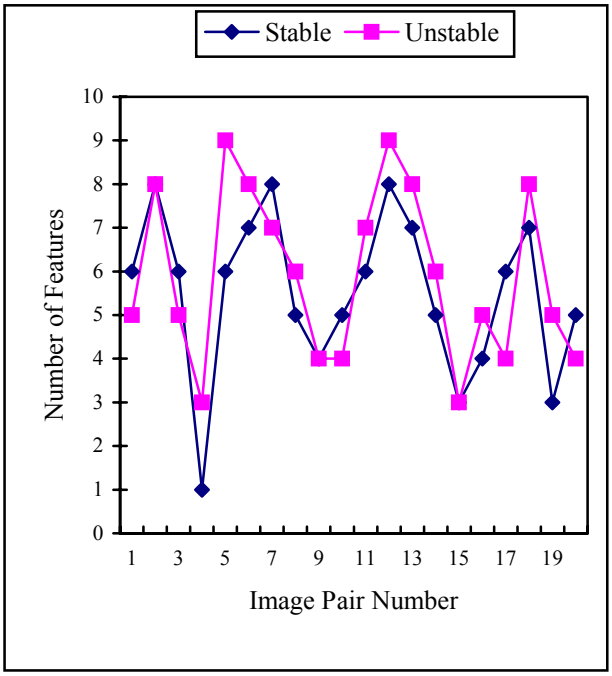

Figure 12: An analysis of finding stable landmarks over 20 pairs of images.

## 8 Conclusion

The work presented in this report showed that Distinctness Analysis and Texture Analysis are suitable choices for use as feature extraction techniques especially in a texture-rich environment such as the Great Barrier Reef. Texture Analysis can be used for extracting features from natural images and represent them with invariant descriptors. Distinctness Analysis can then be used for reducing the set of features generated and maintaining the matching capability with an acceptable level of performance. Based on the results reported here, the opportunity of having a fast and robust feature extraction technique could be considered as feasible.

## Acknowledgements

This work is financially supported by the Australian Cooperative Research Centre for Intelligent Manufacturing Systems & Technologies (CRC IMST) and by the Australian Research Council Centre of Excellence for Autonomous Systems (ARC CAS).

## References

[Carneiro and Jepson, 2002] G. Carneiro and A. D. Jepson, Phase-based local features, *7th European Conference on Computer Vision,* Copenhagen, vol. 1, pp. 282-296, 2002.

[Csorba, 1997] M. Csorba, *Simultaneously Localisation and Mapping*, PhD thesis of Robotics Research Group, Department of Engineering Science, University of Oxford, 1997.

[Harris and Stephen, 1988] C. Harris and M. Stephen, A combined Corner and edge detector, *Alvey Vision Conference*, pp 147-151, 1988.

[Ke and Sukthankar, 2004] Y. Ke and R. Sukthankar, PCA-SIFT: A more Distinctive Representation for Local Image Descriptors, *Computer Vision and Pattern Recognition*, 2004.

[Kiang *et al.*, 2004] K. Kiang, R. A. Willgoss, A. Blair, Distinctive Feature Analysis of Natural Landmarks as a Front end for SLAM applications, *2nd International Conference on Autonomous Robots and Agents*, New Zealand, pp. 206-211, 2004.

[Kiang *et al.*, 2005] K. Kiang, R. A. Willgoss, A. Blair, Distinctness Analysis on Natural Landmark Descriptors, *the 5th International Conference on Field and Service Robotics*, 2005.

[Kragic and Christensen, 2005] D. Kragic and H. I. Christensen, *Advances in Robot Vision, Robotics and Autonomous Systems*, Vol. 52, Issue 1 1-3, 2005.

[Lindeberg, 1994] T. Lindeberg, Scale Space Theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21:2, pp 224-270, 1994.

[Lowe, 2004] D.G. Lowe, Distinctive image features from scale-invariant keypoint, *International Journal of Computer Vision*, 60, 2, pp 91-110, 2004.

[Mikolajczyk and Schmid, 2002] K. Mikolajczyk and C. Schmid, An affine invariant interest point detector, *8th European Conference on Computer Vision*, pp. 128-142, 2002.

[Mikolajczyk and Schmid, 2005] K.Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, *Pattern Analysis & Machine Intelligence*, 2005.

[Thrun *et al.*, 2003] S. Thrun, D. Hähnel, D. Ferguson, M. Montemerlo, R. Triebel, W. Burgard, C. Baker, Z. Omohundro, S. Thayer, and W. Whittaker, A system for volumetric robotic mapping of underground mines, *International Conference on Robotics and Automation*, 2003.

[Tuytelaars and Van Gool, 2000] T. Tuytelaars and L. Van Gool, Wide baseline stereo matching based on local, affinely invariant regions, *11th British Machine Vision Conference*, pp. 412-425, 2000.

[Williams *et al.*, 2002] S.B. Williams, G. Dissanayake, H.F. Durrant-Whyte, Field Deployment of the Simultaneously Localisation and Mapping Algorithm, *15TH IFAC World Congress on Automatic Control*, 2002.

[Williams and Mahon, 2004] S.B. Williams and I. Mahon, Simultaneous Localisation and Mapping on the Great Barrier Reef, *International Conference on Robotics and Automation*, 2004.