Distinctness Analysis on Natural Landmark Descriptors

Kai-Ming Kiang[1], Richard Willgoss[2], and Alan Blair[3]

[1] School of Mechanical and Manufacturing Engineering, University of New South Wales, NSW, 2052, Sydney, Australia kai-ming.kiang@student.unsw.edu.au
[2] School of Mechanical and Manufacturing Engineering, University of New South Wales, NSW, 2052, Sydney, Australia r.willgoss@unsw.edu.au
[3] School of Computer Science and Engineering, University of New South Wales, NSW, 2052, Sydney, Australia blair@cse.unsw.edu.au

Autonomous navigation using natural landmarks in an unexplored environment is a very difficult problem to handle. While there are many techniques capable of matching pre-defined objects correctly, few of them can be used for real-time navigation in an unexplored environment. One important unsolved problem is to efficiently select a minimum set of usable landmarks for localisation purposes. This paper presents a method which minimises the number of landmarks selected based on texture descriptors. This enables localisation based on only a few distinctive landmarks rather than handling hundreds of irrelevant landmarks per image. The distinctness of a landmark is calculated based on the mean and covariance matrix of the feature descriptors of landmarks from an entire history of images. The matrices are calculated in a training process and updated during real-time navigation.

1 Introduction

Autonomous navigation in an unexplored environment is more challenging than in a controlled environment. In particular, underwater environments are mostly unexplored and do not have GPS access. Therefore navigation is generally based on methods such as Simultaneous Localization and Mapping (SLAM) [1] [2]. Most existing SLAM algorithms rely on artificial landmarks which do not exist in an unexplored environment. Recently, methods have been developed for extracting natural landmarks with representations that are invariant to scaling, distortion and perspective. Most of these methods select landmarks based on local properties of points, such as extracting the extrema [3] [4] or corner features [5]. The surrounding properties of these points are then analysed and converted to a vector of feature descriptors. These methods can efficiently select invariant natural landmarks from each image, so that

P. Corke and S. Sukkarieh (Eds.): Field and Service Robotics, STAR 25, pp. 67–78, 2006.

the same landmarks can be picked up under different geometric or lighting conditions from different images.

However, such methods tend to generate hundreds of landmarks per image. For real-time SLAM applications, it is computationally infeasible to compare landmarks from the current image against a database of all landmarks previously seen. SLAM does not only require a method for selecting natural landmarks that are invariant, but also requires selection of a small enough set of distinctive landmarks for computing the similarity between landmarks. A method for selecting distinctive landmarks that is both economical and efficacious is described in this paper.

2 Background

Scale Invariant Feature Transformation (SIFT) [4] is a method which has received much attention recently for its robustness in representing landmarks. It analyses the local gradients of the extrema extracted using Difference of Gaussian (DOG) filters [6] . Its descriptors are claimed to be invariant under changes in scale, rotation, shift and illumination conditions.

We have previously devised another method of representing landmarks based on DOG and frequency distribution analysis [3] that could potentially provide more robust descriptors than those based on gradient properties because the frequency properties used are usually less sensitive to noise.

Corner-based approaches, such as the Harris Matrix [7], claim to be also invariant under affine transformation [5]. Other methods including phase congruency [8], wide baseline stereo matching [9], intensity transformation [10] and steerable filters [11], are also designed to provide invariant descriptors. Some comparison of these techniques has been reported [12].

All of the methods mentioned above can be described in two main steps. Firstly, select interest points based on local properties. Secondly, analyse and represent the local properties of interest points by descriptors. The reason that methods mentioned so far tend to generate hundreds of landmarks per image is because the selection process occurs prior to descriptor transformation and is therefore based only on raw image data. The main motivation of distinctness analysis presented in this paper is to have a further selection process based on the landmarks represented by descriptors i.e. a post-descriptor selection process.

3 Distinctivness Analysis

The question arises as to how a few relevant landmarks out of a potentially large set should be remembered. In Figure 1, it would be best to remember the center object because it is the most distinctive among the set. If one remembers any of the other objects, which are similar to each other, it will be hard to distinguish between them later on. The algorithm should maximize the probability of recognizing and localizing correctly, based on the features of just a few chosen landmarks.

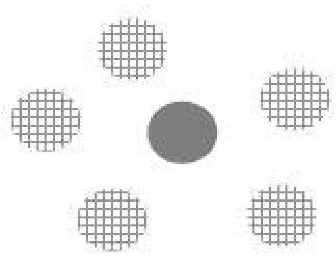

Fig. 1. Simple diagram of a distinctive object among other objects

3.1 Mathematical Distinctness

Mathematically speaking, if each landmark is represented by descriptors using a method noted in Section 2, each landmark becomes a feature vector of descriptors in parametric space. Distinctness can be judged from analyzing these vectors. The general philosophy of distinctness selection is to preserve a set of parameters that appear less frequently whilst deleting those that appear more frequently. If we consider all the feature vectors of landmarks as containing random variables, the probability of appearance for each of them can then be calculated by assuming a Gaussian distribution of the vectors using the formula [13]:

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^m det(\mathbf{C})}} \times exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\} \tag{1}$$

where:

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j \tag{2}$$

and

$$\mathbf{C} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{x}_j-\boldsymbol{\mu}) \cdot (\mathbf{x}_j-\boldsymbol{\mu})^t \tag{3}$$

where n = the number of landmarks.

A distinctness selection can then be made on the basis that the lower the probability, the more distinctive a landmark is judged to be.

3.2 Global Distinctness

Since distinctness selection is a process of minimizing the number of landmarks, the selection criteria must select landmarks consistently from a variety of images. The distinctness of a landmark must have a global meaning for it to be call distinctive i.e. any possible matches should pick out landmarks that are distinctive across images rather than within a particular image.

Referring to Equation 1, the distinctness of a landmark is calculated based on a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\mathbf{C}$. To obtain these two matrices, the sample feature vectors must be selected over a wide range of images of the environment. However, remembering all the sampled feature vectors from each image can accumulate to a huge database. This is avoided because the mean and covariance are updated on every image without the need for recalculation later on.

Let us denote the mean and covariance for the global distinctness by $\boldsymbol{\mu}_\mathbf{t}$ and $\mathbf{C}_\mathbf{t}$ respectively and those for the current image by $\boldsymbol{\mu}_\mathbf{c}$ and $\mathbf{C}_\mathbf{c}$. Then $\boldsymbol{\mu}_\mathbf{c}$ and $\mathbf{C}_\mathbf{c}$ can be calculated from Equations 2 and 3; assuming $\boldsymbol{\mu}_\mathbf{t}$ and $\mathbf{C}_\mathbf{t}$ have been initialised, they can be updated using the formula:

$$\boldsymbol{\mu}_\mathbf{t} = \lambda\, \boldsymbol{\mu}_\mathbf{t-1} + (1 - \lambda)\boldsymbol{\mu}_\mathbf{c} \tag{4}$$

where λ is the innovation factor, which determines how much the system relies on history versus new data.

$\mathbf{C}_\mathbf{t}$ is calculated on the following formula:

$$C_{t(x,y)} = E(XY)_t - \mu_{t(x)}\mu_{t(y)} \tag{5}$$

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

$$E(XY)_t = \lambda\, E(XY)_{t-1} - (1 - \lambda)E(XY)_c \tag{6}$$

$E(XY)_{t-1}$ and $E(XY)_c$ can be obtained by rearranging Equation 5 using $E(XY)$ as the subject with appropriate $\boldsymbol{\mu}$ and $\mathbf{C}$ matrices.

$\boldsymbol{\mu}_\mathbf{t}$ and $\mathbf{C}_\mathbf{t}$ can be updated iteratively using $\boldsymbol{\mu}_\mathbf{c}$ and $\mathbf{C}_\mathbf{c}$. To initialize $\boldsymbol{\mu}_\mathbf{t}$ and $\mathbf{C}_\mathbf{t}$, they are assigned equal to $\boldsymbol{\mu}_\mathbf{c}$ and $\mathbf{C}_\mathbf{c}$ for the first input image. $\boldsymbol{\mu}_\mathbf{t}$ and $\mathbf{C}_\mathbf{t}$ require the system to run over a series of images in order to have confidence in global distinctness. A practical solution is to take a safe walk in the environment of interest e.g. move forward a few steps then move backward a few steps, before using the data for exploration into an unexplored environment.

3.3 Probability of Similarity

Once distinctive landmarks have been extracted, they are compared to form a judgment on how likely any two of them correspond to the same landmark.

This involves calculating the probability of similarity between two selected landmarks from different images.

Each landmark is extracted and converted into a feature descriptor i.e. a p-dimensional vector, which is subject to sources of randomness. Firstly there is random noise from the sensors. Secondly, the descriptor expression is itself a simplified representation of the landmark. Lastly the two images being compared could be viewing the landmark from a different perspective, which causes geometric distortion. Therefore, each landmark can be considered as a single sample of the observing object.

In making inferences from two landmarks in two different images, it is in principle a standard significance test. However, comparison is only made between two single samples. For this reason, the ANOVA test (The Analysis of Variance) cannot be used because the sample size required should be large.

For multidimensional vector comparison, the χ_v^2 (Chi-Squared) distribution test is appropriate. Chi-Squared distribution is a combined distribution of all dimensions which are assumed to be normally distributed. It includes an additional variable v describing the degrees of freedom. Details can be found in [13].

In multidimensional space, the χ_v^2 variable is defined by:

$$\chi_v^2 = N(\overline{\mathbf{x}} - \overline{\mathbf{y}})^t \Sigma^{-1} (\overline{\mathbf{x}} - \overline{\mathbf{y}}) \tag{7}$$

where:

$\overline{\mathbf{x}}$ and $\overline{\mathbf{y}}$ are the mean of the measurements of X and Y respectively;

Σ is the covariance matrix of noise;

N is a function related to the sample size of the two measurements.

Since our sample size is one, then $N = 1$, $\overline{\mathbf{x}} = \mathbf{x}$ and $\overline{\mathbf{y}} = \mathbf{y}$. Equation 7 simplifies to:

$$\chi_v^2 = (\mathbf{x} - \mathbf{y})^t \Sigma^{-1} (\mathbf{x} - \mathbf{y}) \tag{8}$$

If the noise of each dimension is independent of the other, the inverse covariance is a diagonal matrix and hence can be further simplified to:

$$\chi_v^2 = \sum_{i=1}^{p} \frac{(x_i - y_i)^2}{\sigma_i^2} \tag{9}$$

where p is the number of dimensions of $\mathbf{x}$.

Since $\mathbf{x}$ contains p independent dimensions, then the degree of freedom v is p not $(p-1)$ as usually defined for the categorical statistic. Also $\sigma_i = \sqrt{2}\sigma$, where σ is the standard deviation for a single random variable on each dimension.

With χ_v^2 and v obtained, the probability of similarity is defined to be equal to the integrated probability at the χ_v^2 value obtained. The integrated probability of Chi-Square distribution can be found in statistical tables.

4 Experimental

In this section, experiments were conducted on a series of sub-sea images (courtesy of ACFR, University of Sydney, Australia). The configuration was set such that the camera was always looking downwards on the sea floor. This configuration minimised the geometrical distortion caused by different viewpoints.

4.1 Initial Test of the Algorithm

For this experiment, the algorithm was written in Matlab V6.5 running on a PC with a P4 2.4GHz processor and 512Mb of RAM.

To demonstrate how the distinctness analysis algorithm worked, a typical analysis is now explained in detail. In the following example, we have trained the distinctness parameters $\boldsymbol{\mu}_\mathbf{t}$ and $\mathbf{C}_\mathbf{t}$ over 100 images from the series. The texture analysis described in [3] generated invariant landmarks on two particular images shown in Figure 2 which consist of partially overlapping regions.

The distinctness analysis described in Section 3 was then applied to further select a smaller set of landmarks which were considered to be distinctive as shown in Figure 3. The innovation factor λ was chosen to be 0.9 weighting the past significantly more than the present. The threshold for distinctness in Equation 1 was chosen to be 0.2, a value that kept the number of landmarks chosen to relatively few. In Figure 4, the two highest matches of landmarks that scored higher than a threshold probability of 0.8 are shown with linked lines.

The first selection of landmarks based on DOG techniques generated many landmarks scattered all over the two images. More landmarks could usually mean more confidence for matching. However, the computational time for making comparison would also increase. In addition, since non-distinctive objects were not excluded, many of the matches could possibly have been generated by similar objects located at different places.

Figure 3 shows a selection of landmarks that the algorithm chose to be globally distinctive. The number of landmarks was significantly reduced when retaining useful matches between the two images. Since these landmarks should not appear frequently in the environment, the possibility that similar objects appear in different locations is minimised.

The run-time of this algorithm depended on the complexity of the images. On average, the time required to generate landmarks with descriptors took $\sim$6 seconds per image while the selection process of distinctive landmarks required only $\sim$0.05 second per image. Thus the extra time required to select distinctive landmarks was comparatively small. The time required to calculate the probability between any two landmarks was $\sim$0.001 second. On average, the sub images could generate 150 landmarks. Therefore there were 150 x 149 potential comparisons required to calculate between two images. The maximum time required would be $\sim$0.001 x 150 x 150 = 22.5 seconds. But after

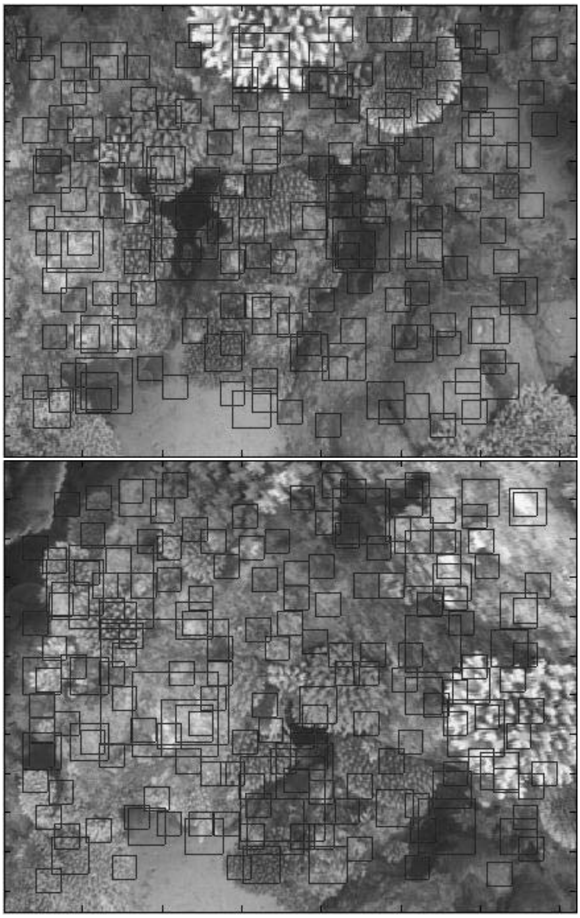

Fig. 2. Two particular images from the Sub sea images. The different sizes of boxes are landmarks generated using texture analysis described in [3].

applying the distinctness selection process, the number of landmarks reduced to ∼10 per image. The time required to make comparison thus reduced to ∼0.1 second. The algorithm is currently being re-implemented in C which should improve its speed significantly.

4.2 Global Distinctness Test

The performance of the algorithm was then tested with different images across the environment. The test should reveal whether the algorithm could select objects that are truly distinctive from a human's perspective. The task is in some ways subjective. A group of images are displayed in Figure 5 together

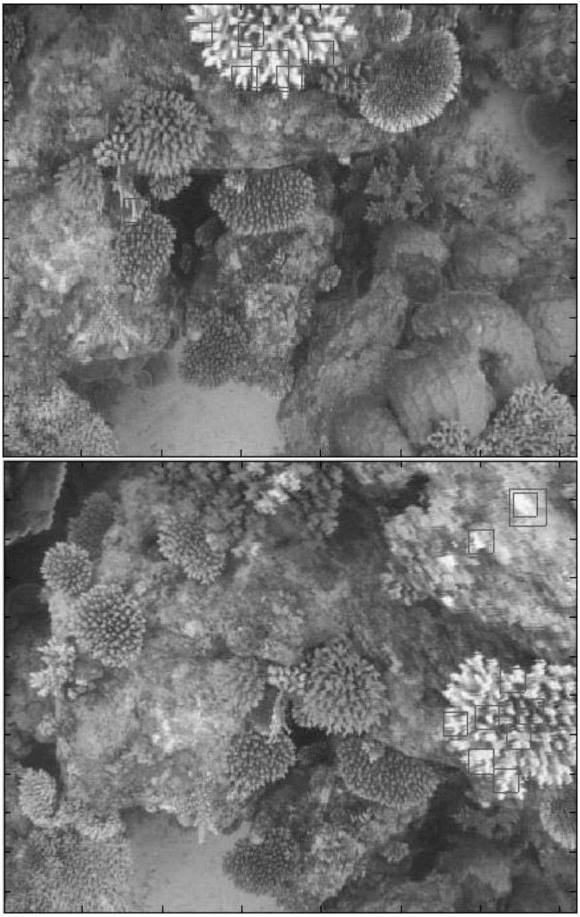

Fig. 3. On the same two images of Figure 2. After applying the Distinctness selection process described in Section 3, the number of landmarks is reduced.

with the landmarks selected by the algorithm. The reader can judge the performance of the algorithm by noting what has been picked out.

As can be seen, the distinctive landmarks are usually the complicated textural corals which tend to be sparsely distributed.

It can be seen that in some of these images, there is a single distinctive object, in which case, the algorithm has concentrated the landmarks in that region. However, in images that contain no obvious distinctive objects, the algorithm has chosen fewer distinctive landmarks scattered over the whole image.

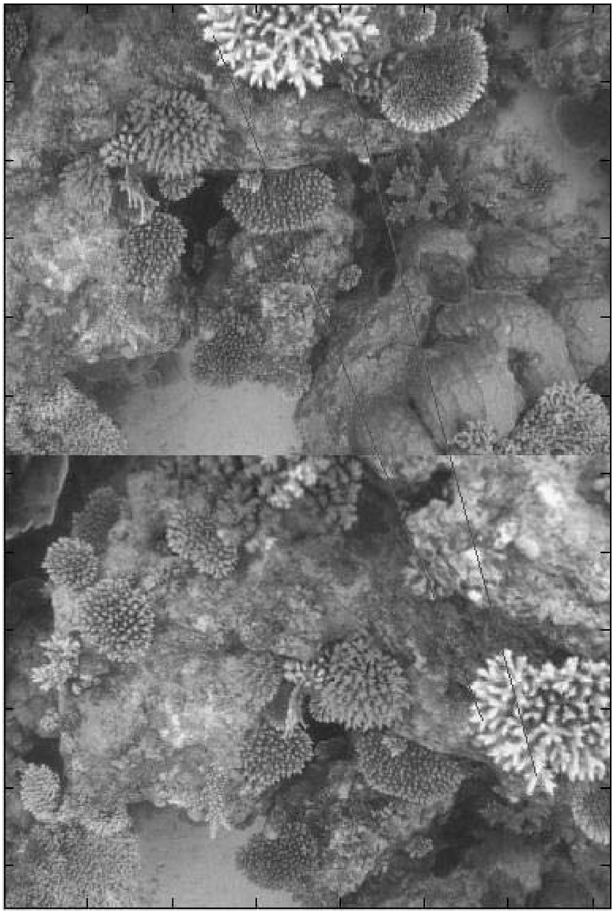

Fig. 4. After comparing each distinctive landmarks, two highest matches that contains probability of over 0.8 are joined by lines for illustration.

4.3 Stability Test

A final test was conducted to check on the stability of chosen landmarks. By stability, we mean that the same landmark should be picked out invariant to any changes in shift, rotation, scale and illumination. A selection of image pairs was made such that these pairs contained relatively large changes in the previously mentioned conditions and contained overlapping regions. After the algorithm was applied to each image to pick out distinctive landmarks, an inspection was made within the overlapping region to count the number of distinctive landmarks that appeared within a few pixels in corresponding locations of the two images. By comparing this number with the number of landmarks that did not correspond in both of the images, a measure of stability was obtained. For example in Figure 3, there were four distinctive

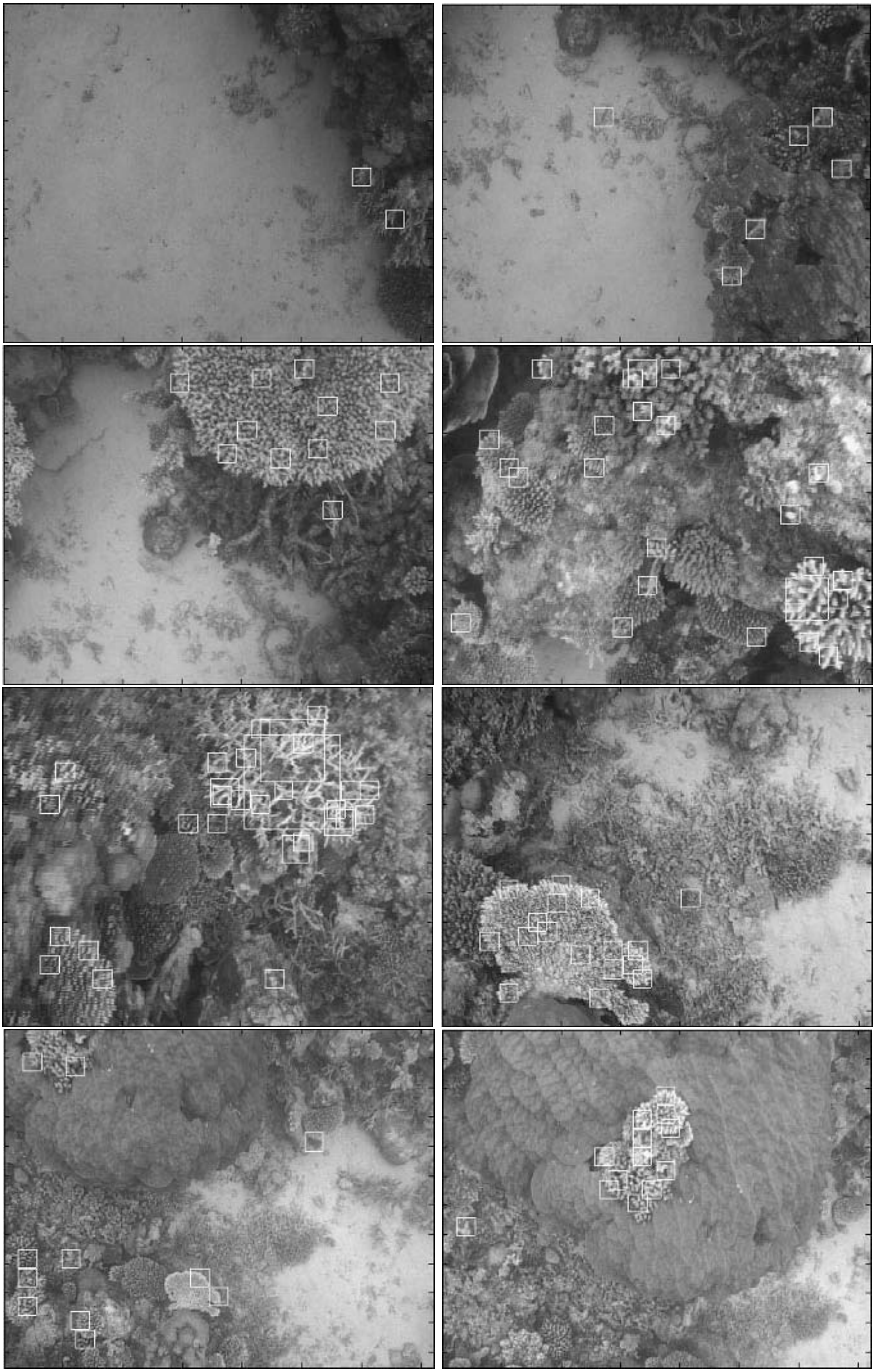

Fig. 5. Sample images from sub-sea series (courtesy of ACFR, University of Sydney, Australia)

landmarks appearing in corresponding locations of both images. On the other hand, there were three which do not correspond in both images.

In Figure 6, 20 pairs of images have been analysed in the way indicated above. On average, 47% of the landmarks selected as distinctive in one image appeared correspondingly in both images. This was deemed a relatively high hit rate for tracking good distinctive landmarks through image sequences and shows promise for enabling map building in a SLAM context.

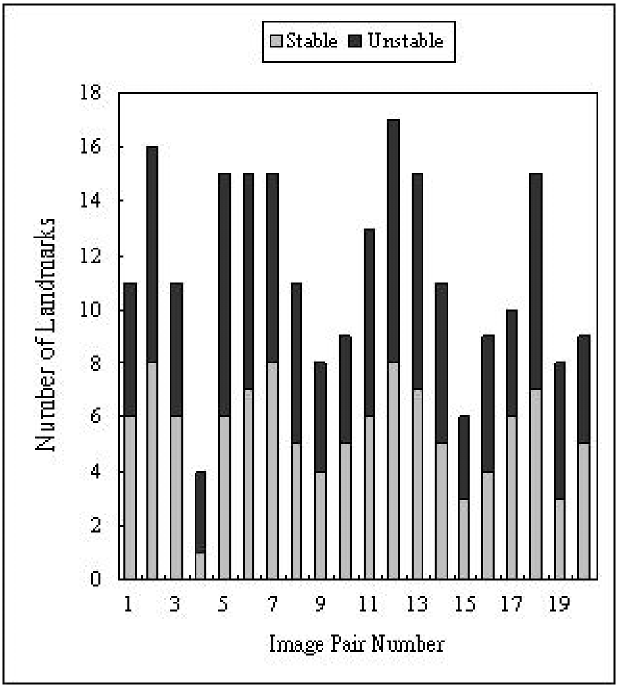

Fig. 6. An analysis of finding stable landmarks over 20 pairs of images.

5 Conclusion and Future Work

The work reported here has shown that it is possible to differentiate image data in such a way that distinctive features can be defined which can be tracked on images as the features progress through a sequence of images in an unexplored environment.

The paper presented an extended algorithm for selecting distinctive landmarks among numerous candidates, that could also be adapted and combined with existing invariant landmark generation techniques such as SIFT or Texture Analysis. In our experiments, the algorithm is demonstrated to discriminate a small enough set of landmarks that would be useful in techniques such as SLAM.

We are currently working to incorporate this landmark selection algorithm with inertia sensor information to form a functioning SLAM system and deploy it in a submersible vehicle.

Acknowledgment

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

1. Csorba M (1997) Simultaneously Localisation and Mapping. PhD thesis of Robotics Research Group, Department of Engineering Science, University of Oxford
2. Williams S B (2001) Efficient Solutions to Autonomous Mapping and Navigation Problems. PhD thesis of ACFR, Department of Mechanical and Mechatronic Engineering, the University of Sydney
3. Kiang K, Willgoss R A, Blair A (2004) "Distinctive Feature Analysis of Natural Landmarks as a Front end for SLAM applications", 2nd International Conference on Autonomous Robots and Agents, New Zealand, 206–211
4. Lowe D G (2004) "Distinctive image features from scale-invariant keypoint", International Journal of Computer Vision, 60, 2:91–110
5. Mikolajczyk K and Schmid C (2002) "An affine invariant interest point detector", 8th European Conference on Computer Vision Czech, 128–142
6. Lindeberg T (1994) "Scale-Space Theory: A Basic Tool for Analysing Structures at Different Scales", J. of Applied Statistics, 21, 2:224–270
7. Harris C, Stephen M (1988) "A combined Corner and edge detector", 4th Alvey Vision Conference Manchester, 147–151
8. Carneiro G, Jepson A D (2002) "Phase-based local features", 7th European Conference on Computer Vision Copenhagen, 1:282–296
9. Tuytelaars T, Van G L (2000) "Wide baseline stereo matching based on local, affinely invariant regions", 11th British Machine Vision Conference, 412–425
10. Schmid C, Mohr R (1997) "Local grayvalue invariants for image retrieval", Pattern Analysis and Machine Intelligence, 19, 5:530–534
11. Freeman W, Adelson E (1991) "The design and use of steerable filters", Pattern Analysis and Machine Intelligence, 13, 9:891–906
12. Mikolajczyk K, Schmid C (2003) "Local grayvalue invariants for image retrieval", Pattern Analysis and Machine Intelligence, 19, 5:530–534
13. Manly B (2005) Multivariate Statistical Methods A primer 3rd edition, Chapman & Hall/CRC