

Distinctive Feature Analysis of Natural Landmarks as a Front end for SLAM applications

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

Abstract

This paper presents a method for extracting distinctive textural features from images taken from natural scenes. The aim is to use natural landmarks for navigation in an unexplored environment. Natural features are all different and complex in shape. To be able to use them for navigation, informative representation of these features and a careful selection process is required. The present method is termed as 'Distinctive Texture Analysis'. It has three parts. Firstly, a method of selecting Interest Points from the filtered images is presented. Secondly, Texture Analysis of the local properties of Interest Points are applied and stored as descriptors. Thirdly, to reduce the number of landmarks selected for storage and comparison purposes, Distinctness Analysis is applied. Current results have shown that the most distinctive features as concurred by viewing the simple images are able to be selected and correctly matched. Results provided for the complex underwater images illustrated the difficulty and limitation. However, when this method is applied with multiple numbers of landmarks such that correlation of landmark positions is considered, certainty for SLAM can increase. Future works can include consideration of such correlation.

Keywords: navigation, image analysis, distinctive features, natural landmarks, SLAM, scale invariant

1 Introduction

Navigating in an unexplored environment such as underwater is very different to a controlled indoor environment. Traditional SLAM (Simultaneous Localization and Mapping) algorithms use point-based features and usually rely on artificial landmarks, which do not exist in an unexplored environment. If it is proposed that natural landmarks can be used instead, any new method must be able to handle the complicated structures of such landmarks that can come in many shapes and textures without prior knowledge of their characteristics.

To compare the chosen extracted landmarks, a representation that is invariant to geometric transformation is needed. Hence, a careful choice of descriptors of texture properties is also necessary. Moreover, a method must be designed to extract only the distinctive objects from background to use as landmarks for comparison to save computational time and storage space.

This paper proposes a method that extracts and stores only distinctive points of an image for comparison. This approach contains three stages, extracting interest points from original images, comparing the local properties of these interest points by texture analysis and lastly selecting the most distinctive points to store and compare.

1.1 Related research

The traditional manner of providing feature data for SLAM has been characterised by point-based feature representation. Many attempts have been made with this approach [1] [2] [3] using Kalman filtering [4]. Every landmark in point-based feature representation is located specifically in the environment for easy recognition. In an open unstructured environment, it is hard to find such easily identifiable landmarks. Therefore, using these methods in an open environment is only possible if artificial landmarks are introduced. This requires putting beacons or reflectors in the environment prior to robot exploration. However, this negates the original aim i.e. navigating in an open and unexplored environment.

Methods that recognise landmarks by images commonly use Principle Component Analysis (PCA) [5]. PCA represents objects weighted by the most common features of exemplars and uses these principle components for matching. This algorithm effectively reduces the problem to approximately 20 to 30 dimensions. The major limitation is that it requires having pre-knowledge of possible landmarks.

More recent methods for reducing dimensionality include Isomap and LLE [6] [7] [8] which reduce from a global and a local perspective correspondingly while preserving neighbourhood relationships. Again,

these methods require pre-knowledge of possible landmarks.

Another method of landmark recognition called Scale Invariant Feature Transformation (SIFT) [9] can be characterised by analysing the local properties of interest points of the images, noting they are usually invariant under certain conditions, namely scale, rotation, shift, illumination and affine transformation. SIFT is an efficient method for representing landmarks in this manner based on local gradients of extrema using Difference Of Gaussian (DOG) filters.

In addition to SIFT, some Harris Matrix [10] based approaches claim to be invariant under affine transformation [11] [12]. Other methods including phase congruency [13] [14], wide baseline stereo matching [15], intensity transformation [16] [17] [18] and steerable filters [19] are also designed to provide invariant descriptors. Some comparison of these techniques have been reported [20] [21] [22].

The main motivation for the methods mentioned above has come from interests in vision analysis for object recognition not necessarily for navigation and guidance via new feature extraction. For SLAM applications, real time computation is required. Therefore it is important to select the minimum number of landmarks which are necessary and distinct for storage and comparison. This paper reports on a method that has the potential to achieve this objective.

In Section 2, a new method for finding distinctive textural features from images is described. Section 3 presents some test results showing how well the technique works on both a simple and complex natural environment. Section 4 gives conclusions and presents possible future work.

2 Distinctive Feature Analysis

As already mentioned, the method of feature extraction that is reported here consists of three parts, extracting interest points, analysis of texture via comparison of segmentation of Discrete Fourier Transforms (DFT) and selection of distinctive segments from the descriptors for comparison via probability analysis.

2.1 Interest points

The first step is to extract areas of an image as landmarks. These areas must ideally be invariant to rotation, shift, scale, illumination and affine transformation such that when the image is taken from the same scene again, the same features would be evident. In [9] it is proposed that the peaks of the DOG image are invariant to scale, shift, rotation and illumination to some degree.

The 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 original image $B(x,y,\sigma)$.

$$B(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

where $*$ is the convolution operator in x and y plane. The Gaussian function:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

is applied to $B(x,y,\sigma)$ 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 blurred images.

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

Note that the variable σ depends on the complexity of image. It is possible to construct a pyramid of DOG images as described in [9]. However, here only one level is used to keep the computational burden down.

In the technique proposed here, bilinear interpolation is applied before the first convolution takes place, significantly reducing the size of the image and computational time. In the images used in this paper, we are working mainly with lower frequency bands that are designed to capture fewer, but larger objects within images.

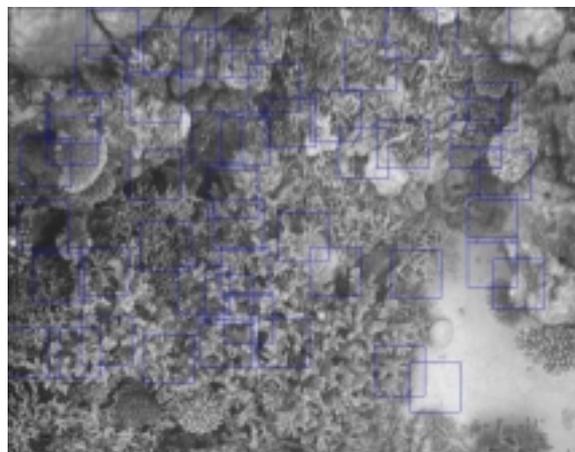


Figure 1: Expanding extrema in a DOG image to areas for use as landmarks.

From the DOG image, extrema can be found directly by comparing each pixel with its neighbour in a preset radius for each image. After locating extrema, a local limited neighbourhood (area) is expanded around the point of interest for analysis. Even though square areas are produced to enable DFT, as will be discussed further in Section 2.2, a Hanning window is applied to all areas making them approximately circular and hence invariant to rotation. These areas, having special properties centred at extrema of DOG images are then used as landmarks.

Since DOG properties are invariant to most geometrical transformations, the same points expanded into areas should be able to repeatedly selected in different images. Chosen areas can be seen in Figure 1.

2.2 Texture analysis

Texture properties are used to describe the areas chosen as detailed in 2.1. Measuring texture is closely related to measuring a frequency distribution which can be obtained from the periodogram [23]. The periodogram is calculated from the squared terms of the individual values from the Discrete Fourier Transform (DFT) as shown below:

$$S_{\mathbf{m}}^{pg}(\omega) = \frac{1}{\|W\|} \left| \sum_{\mathbf{k} \in W} q(\mathbf{k}) x[\mathbf{k}] e^{-j\omega' \mathbf{k}} \right|^2$$

where W is the prototype image window, centred approximately about the origin. $q(k)$ is the Hanning window function defined as:

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

A periodogram represents a frequency distribution as a matrix of values. Each value represents a particular amplitude of frequency. Since for natural images, low frequencies dominate, the distribution in the periodogram reduces the statistical redundancy and provides a lower dimensional representation. The Hanning window is used to smoothly correct the boundary effect of the DFT.

An analysis of the properties of periodogram is conducted as shown in Figure 2, the concentric annuli and angular segments are used to extract the components of the periodogram. After summing the magnitudes within each segment, ratios between segments can be obtained and are meaningful in terms of the spatial frequency distributions obtained in each image and as a function of orientation.

The concentric circles measure frequency distribution invariant to rotation. The angular segments on the other hand, measure angular frequency distribution. The angular segments can also be made invariant to rotation if the segments are compared relative to the highest magnitude segment not the absolute orientation.

When using this method, firstly there is a need to exclude the lower frequencies in the angular segments because too few samples are available to subdivide into angular segments. Secondly, only half of the plane needs to be analysed because the DFT is always symmetrical.

The sums of samples in the concentric annuli and angular segments can be further analysed as descriptors of the image. For concentric annuli, these descriptors can be normalised or retain their absolute magnitude. The number of descriptors used depends on pixel density of the image. For angular segments, descriptors can also be normalised or absolute, but shifted relatively to the maximum value obtained.

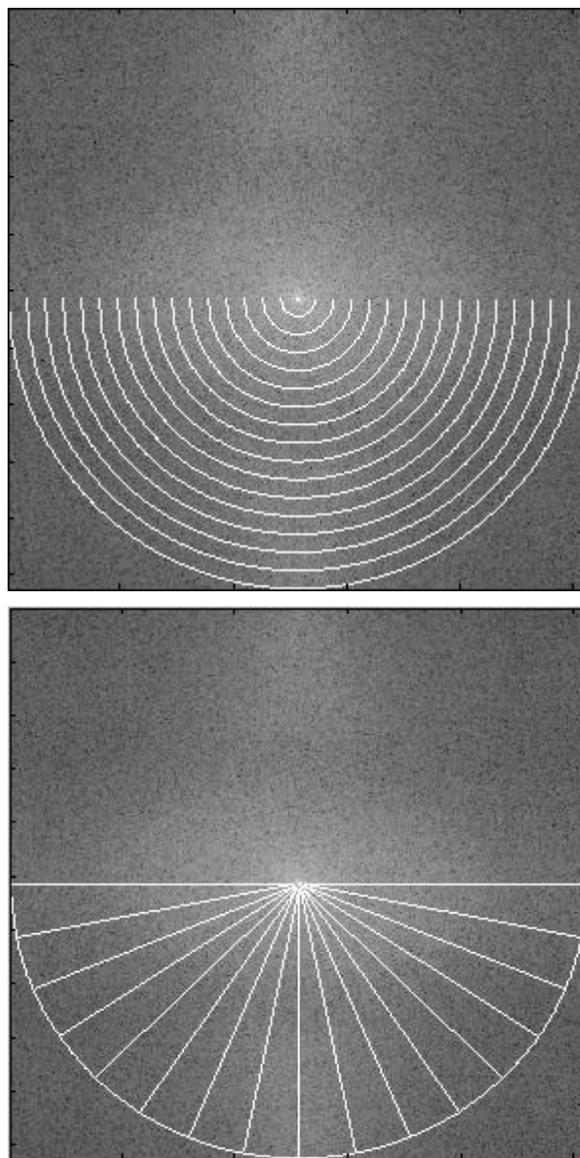


Figure 2: Periodogram features used; top - concentric annuli; lower - angular segments.

The descriptors from the DFT now serve as signatures for further comparison in parameter space by finding Euclidean norms. Hence similarity of features, landmarks and textures is postulated to increase as the Euclidean distance in parametric space decreases.

2.3 Distinctness analysis

Each image can now be represented by descriptors as explained in Section 2.2 derived from the DFT. In

parametric space, further selection of distinctive segments can be made. The general philosophy of distinctive selection is to preserve a set of parameters that appear infrequently whilst deleting those that appear everywhere. Using all the values of detected segments as samples and considering all descriptors as random variables, the probability distribution for each descriptor can then be calculated, assuming a Gaussian profile, from:

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

where:

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

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

A distinctness selection is made on the assumption that a result with the lowest probability represents the most distinctive feature.

3 Results

3.1 Simple images test

The procedure given in Section 2 was applied to two different sets of images. The first set contained simple features of easily recognised objects whilst the second set contained complex underwater features.

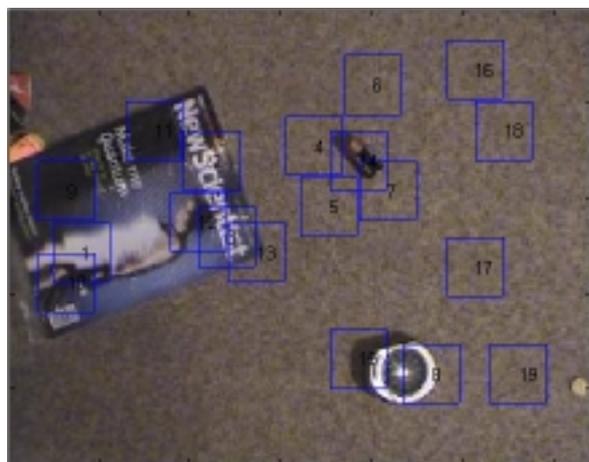


Figure 3: Feature areas selected as extrema in a DOG analysis.

In the first test, the areas selected by DOG extrema analysis as described in Section 2.1 contained features as shown in Figure 3. After applying the distinctness analysis, the number of distinct areas reduced significantly as well as filtering out most of those emanating from the carpet background (see Figure 4: Top).

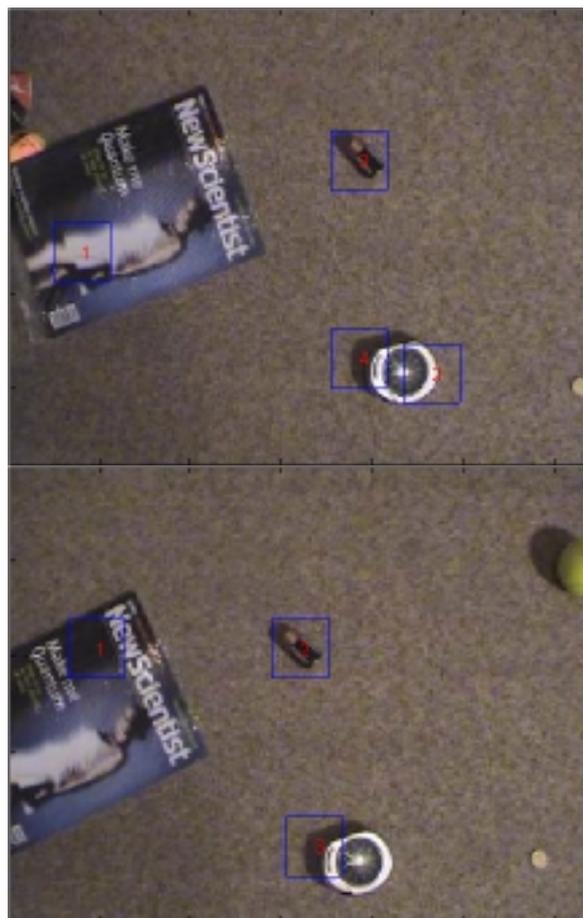


Figure 4: Distinctive feature areas selected after distinctness analysis has been applied to images: Top – selection from image (1) in Figure 3: Bottom – selection from a similar image (2) at the same stage of being processed.

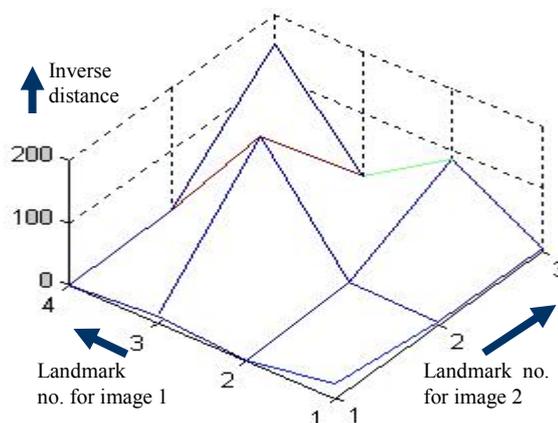


Figure 5: Correlation of features using inverse Euclidian distances for two similar images.

In Figure 4, the analysis has been taken one stage further, selecting for distinctness using Euclidian distances. The two images captured the same scene with a small translation and rotation. The correlation between numbered features in each image is compared using a contour plot in Figure 5.

3.2 Complex underwater images test

In the second test, images from a realistic robotic environment were used where a robot was placed in a coral reef. In this test, one of the feature areas is selected from the first image of the set as a reference for comparison (see landmark 16 - an area within the big coral in the lower right corner in Figure 6).

Interest is centred on whether the procedure can select other feature areas that are similar to the reference whilst rejecting those that are irrelevant. The images shown in Figure 7 labelled 2 to 10 have been processed for comparison with the feature areas from the first image.

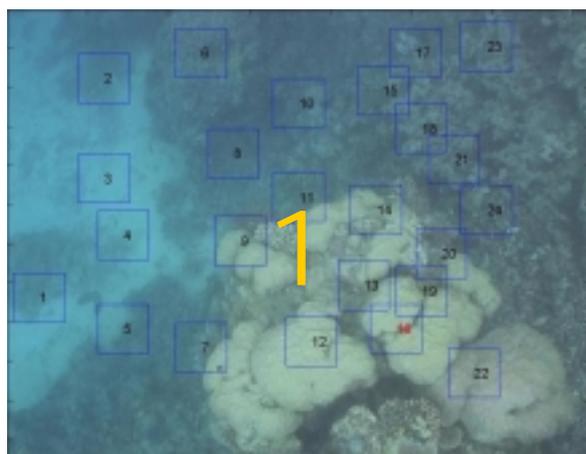


Figure 6: Reference image. (Courtesy of ACFR, University of Sydney, Australia)

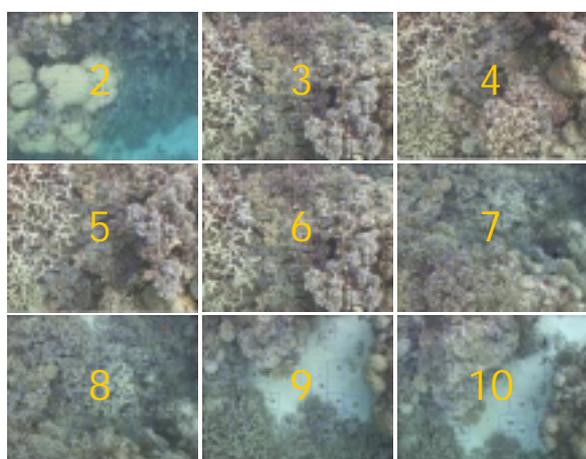


Figure 7: A series of images analysed for comparison with a reference image analysis. (Courtesy of ACFR, University of Sydney, Australia)

After applying the procedure of Section 2, a contour plot of the correlation of inverse Euclidian distances is shown in Figure 8. Similarity between each feature area and the reference area in Figure 6 has been measured. Peaks in the contour plot indicate feature areas that are similar, whereas a flat surface indicates minimal correlation. The results show that image 2 in

Figure 7 has generated many high peaks compared to the other images. Image 2 is the only image that captures the same scene but from a different angle to that of the reference image. Therefore, in this test, the procedure has correctly retained information relevant for identifying features that are persistent as the scene changes and, at the same time, filtering out irrelevant candidate feature areas.

Notice also that in Figure 8 a significant peak can be observed in image 4, landmark 29 from Figure 7. A check with the test images reveals this landmark has very similar texture properties with the reference. Although the texture is similar, the location of image 4 is different to that of image 1. Hence, this method cannot provide a perfect matching capability. Instead it will match landmarks in real-time but with a few false positive candidates.

This method, combined with correlation of different landmarks within the same image, could significantly increase certainty for SLAM in unstructured environments.

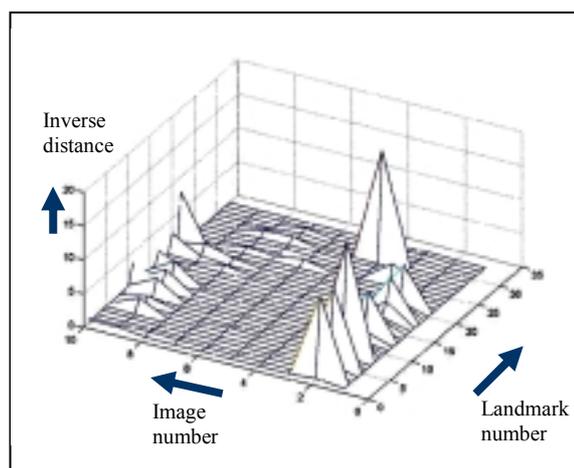


Figure 8: Correlation of features using inverse Euclidian distances between each area of the set of images with the reference.

4 Conclusions and future work

A three-step approach of utilising natural landmarks for navigation is presented in this paper. Results have shown that this method is capable of filtering out irrelevant landmarks while identifying those with similar visual properties.

Future work will investigate how this method can be incorporated within the SLAM framework to achieve autonomous navigation in unstructured environments.

Acknowledgements

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

5 References

- [1] Csorba, M., *Simultaneously Localisation and Mapping*, PhD thesis of Robotics Research Group, Department of Engineering Science, University of Oxford, (1997).
- [2] Williams, S.B., *Efficient Solutions to Autonomous Mapping and Navigation Problems*, PhD thesis of ACFR, Department of Mechanical and Mechatronic Engineering, the University of Sydney, (2001).
- [3] Guivant, J.E., *Efficient Simultaneously Localisation and Mapping in Large Environment*, PhD thesis of ACFR, Department of Mechanical and Mechatronic Engineering, the University of Sydney, (2002).
- [4] Catlin, D.E., *Estimation, Control, and the Discrete Kalman Filter*, vol. 71 of Applied Mathematical Sciences. Springer-Verlag, (1989).
- [5] O'Connell, M.J., *Search Program for Significant Variables*, Comp. Phys. Comm. 8, 49. (1974).
- [6] Tenenbaum, J.B., "Mapping a manifold of perceptual observations", *Advances in Neural Information Processing Systems* 10, Cambridge, MIT Press, pp 682-687 (1998).
- [7] Silva, V. and Tenenbaum J.B., "Global versus local methods in nonlinear dimensionality reduction", *Advances in Neural Information Processing Systems* 15, Cambridge, MIT Press, pp 705-712, (2002).
- [8] Saul, L.K., and Roweis, S.T., "Think globally, fit locally: unsupervised learning of low dimensional manifolds", *Journal of Machine Learning Research* 4, pp 119-155, (2003).
- [9] Lowe, D.G., "Distinctive image features from scale-invariant keypoint", *International Journal of Computer Vision*, pp 91-110, (2004).
- [10] Harris, C. and Stephen, M., "A combined Corner and edge detector", *Alvey Vision Conference*, pp 147-151, (1988).
- [11] Mikolajczyk, K. and Schmid, C., "Indexing based on scale invariant interest points", *ICCV 2001*, pp 525-531, (2001).
- [12] Mikolajczyk, K. and Schmid, C., "An affine invariant interest point detector", *ECCV 2002*, pp 128-142, (2002).
- [13] Carneiro, G. and Jepson, A. D., "Phase-based local features", *ECCV 2002*, vol 1, pp 282-296, (2002).
- [14] Kovese, P., "Phase Congruency Detects Corners and Edges", *DICTA 2003*, pp 309-318, (2003).
- [15] Tuytelaars, T. and Van Gool, L., "Wide baseline stereo matching based on local, affinely invariant regions", *BMVC 2000*, pp 412-425, (2000).
- [16] Schmid, C. and Mohr, R., "Local grayvalue invariants for image retrieval", *PAMI 1997*, vol 19(5), pp 530-534, (1997).
- [17] Van Gool, L., Moons, T. and Ungureanu, D., "Affine/ photometric invariants for planar intensity patterns", *ECCV 1996*, pp 642-651, (1996).
- [18] Florack, L., ter Haar Romeny, B., Koenderink, J. and Viergever, M., "General intensity transformations and second order invariants" *SCIA 1991*, pp 338-345, (1991).
- [19] Freeman, W. and Adelson, E., "The design and use of steerable filters", *PAMI 1991*, vol 13(9), pp 891-906, (1991).
- [20] Mikolajczyk, K. and Schmid, C., "A performance evaluation of local descriptors", *CVPR 2003*, vol 2, pp 257-263, (2003).
- [21] Schmid, C., Mohr, R., and Bauckhage, C., "Evaluation of interest point detectors", *IJCV 2000*, vol 37(2), pp 151-172, (2000).
- [22] Randen, T. and Husoy, J.H., "Filtering for texture classification: A comparative study", *PAMI 1999*, vol 21(4), pp 291-310, (1999).
- [23] Bloomfield, P., *Fourier Analysis of Time Series: An Introduction*, Wiley, (1976).