Texture-based classification of sub-Antarctic vegetation communities on Heard Island

Humphrey Murray a, Arko Lucieer b,*, Raymond Williams c,*

a Datalive Software, 6 George St, Launceston, Tasmania 7250, Australia
b School of Geography and Environmental Studies, University of Tasmania, Private Bag 76, Hobart, Tasmania 7001, Australia
c School of Computing and Information Systems, University of Tasmania, Private Bag 100, Hobart, Tasmania 7001, Australia

**ABSTRACT**

This study was the first to use high-resolution IKONOS imagery to classify vegetation communities on sub-Antarctic Heard Island. We focused on the use of texture measures, in addition to standard multispectral information, to improve the classification of sub-Antarctic vegetation communities. Heard Island’s pristine and rapidly changing environment makes it a relevant and exciting location to study the regional effects of climate change. This study uses IKONOS imagery to provide automated, up-to-date, and non-invasive means to map vegetation as an important indicator for environmental change. Three classification techniques were compared: multispectral classification, texture based classification, and a combination of both. Texture features were calculated using the Grey Level Co-occurrence Matrix (GLCM). We investigated the effect of the texture window size on classification accuracy. The combined approach produced a higher accuracy than using multispectral bands alone. It was also found that the selection of GLCM texture features is critical. The highest accuracy (85%) was produced using all original spectral bands and three uncorrelated texture features. Incorporating texture improved classification accuracy by 6%.

1. Introduction

1.1. Heard Island

Heard Island is a pristine sub-Antarctic island south of the Antarctic Polar Frontal Zone in the Indian Ocean. This Australian territory is a 2800 m high volcanic and glaciated island, and because of its remoteness, human visits to the island are very infrequent. Heard Island is unique in terms of its location, climatic conditions, vegetation communities, geology, volcanic activity, and glacial cover (Bergstrom and Chown, 1999; Bergstrom and Selkirk, 2000; Bergstrom et al., 2002; Scott and Bergstrom, 2006). Up-to-date and accurate spatial information is of crucial importance for sustainable management of the island. Because of the island’s remoteness, satellite imagery provides advanced and cost-effective means to map its land cover and to quantify environmental changes. This information is important for sustainable management of this pristine island, to study the regional effects of climate change, and to assess the effects of human impacts. The glaciers on Heard Island have been receding since 1947 when glacial extent was first estimated from aerial photographs (Thost and Truffer, 2008). This recession was most likely caused by a temperature rise of +0.9 deg C between 1947 and 2004. Glacial retreat has exposed new land that has become available for colonisation of plant species.

During previous expeditions to Heard Island in 1986/1987, 1987/1988, 2000/2001 and 2003/2004 terrestrial plant ecology has been studied and vegetation maps have been produced. These maps were produced manually, based on visual interpretation of aerial photographs and satellite imagery, combined with GPS-based field samples (Bergstrom and Selkirk, 2000; Bergstrom et al., 2002; Scott, 1989; Australian Antarctic Division, 2009). Because of the inaccessibility of Heard Island, field surveys are often expensive and labour intensive, and expeditions can potentially be intrusive. Satellite images have been successfully used in vegetation mapping, monitoring, and ecological applications in the past (Aplin, 2005; Coppin et al., 2004; Jensen, 2000; Xie et al., 2008). Very high spatial resolution imagery (VHR) such as IKONOS (1–4 m spatial resolution, 4 multispectral bands) provides a valuable new source of information for remotely sensed vegetation mapping. Given that the island is rarely visited, satellite image classification could be a suitable technique to produce vegetation maps regularly and accurately. Satellite sensors are also able to capture imagery of large parts of the island or even the entire island, so that complete vegetation maps can be produced. If we can produce accurate vegetation maps from VHR satellite imagery,
we can potentially map and quantify changes in vegetation cover, which for a pristine area like Heard Island provides an important indication of the regional effects of climate change. This study is the first to use satellite imagery for semi-automated vegetation classification on Heard Island.

1.2. Texture-based classification

The multispectral bands of satellite imagery are often transformed into thematic classes using an appropriate classification technique (Lu and Weng, 2007; Tsio and Mather, 2001). Most of these techniques, however, only look at the spectral values of individual pixels and do not take into consideration the spatial context of pixels. With recent VHR imagery, real world objects or regions that were previously represented by only one or two pixels now consist of many pixels. Therefore, techniques that take into account the spatial properties of an image region need to be developed and applied. One approach for including the spatial relationship of pixels is modelling texture. Texture can be defined as the various measures of smoothness, coarseness, and regularity of an image region (Gonzalez and Woods, 1992). Previous studies have shown that combining both multispectral and texture data together can lead to improved classification accuracy (Ruiz et al., 2004; Zhu and Yang, 1998). The popular grey-level co-occurrence matrix (GLCM) texture model (Haralick et al., 1973; Haralick, 1979) has been widely used in remote sensing studies (Clausi, 2002; Franklin et al., 2001; Ouma et al., 2008). Recently, texture-based classification algorithms have been successfully applied to VHR satellite imagery (Aguer et al., 2008; Ouma et al., 2008; Puissant et al., 2005). Tsai et al. (2005) and Tsai and Chou (2006) applied the GLCM to VHR Quickbird imagery to detect invasive plant species. In this study, we apply GLCM texture-based classification to VHR IKONOS imagery of Heard Island.

In addition, the issues of scale and complexity in texture definitions have been raised in previous remote sensing studies. The size of the texture window should ideally match the spatial scale of the object or class under consideration, but this is not always a trivial exercise. The window should be large enough to capture the relevant patterns, but if the window becomes too large edge effects could dominate the results (Puissant et al., 2005). Several studies have looked at the influence of the window size on classification accuracy (Aplin, 2006; Chen et al., 2004; Franklin et al., 1996). The GLCM texture model generates a range of correlated texture features that can be used in a classification (Haralick et al., 1973; Hall-Beyer, 2007). In this study, we systematically examine the effect of the GLCM window size and texture feature selection on classification accuracy.

In addition to spatial scale, thematic scale or complexity is another issue in classifying natural ecosystems (Aplin, 2006). In the last decade or so, a range of studies have explored the use of hierarchy theory in image classification (Akçay and Aksoy, 2008; Blaschke and Strobl, 2001; Benz et al., 2004; Burnett and Blaschke, 2003; Franklin et al., 2001; Ju et al., 2005; Wu, 1999). A classification hierarchy arranges thematic classes into a hierarchical tree with the most generic classes at the top and the more detailed classes further down, inheriting characteristics from their parent classes. In this study we apply a hierarchical classification approach toward classifying Heard Island’s vegetation communities.

1.3. Aim and objectives

In summary, the main aim of this study is to investigate whether incorporating texture improves vegetation classification based on VHR IKONOS imagery of Heard Island. The objectives of the study are (a) to determine an appropriate window size for texture analysis; (b) to determine which texture features should be used for classification; (c) to perform a classification using texture alone; (d) to combine multispectral and texture features in a classification; and (e) to compare this approach to standard multispectral classification.

1.4. Paper Structure

Section 2 of this paper describes the study area and the imagery being used for this research. Section 3 presents the texture-based methods that were used in the study and explains the GLCM which was a fundamental component in all of the techniques described. Section 4 discusses the feature reduction and classification techniques used, as well as the techniques for validation. Section 5 presents and discusses the results obtained while Section 6 presents the conclusions that can be drawn from the methods and results.

2. Study area, imagery and field data

2.1. Study Area

Heard Island is a sub-Antarctic island located in the Indian Ocean at approximately 53°11’S, 73°54’E. Fig. 1 shows how the island is located with respect to Australia and Antarctica. The Territory of Heard Island and McDonald Islands (HIMI) was inscribed on the World Heritage List in 1997 for its outstanding universal natural values. In addition to being recognised internationally for their conservation values, the Heard Island and McDonald Islands and Marine Reserve are significant at the Australian national level for their contribution to the National Representative System of Marine Protected Areas (NRSMPA), their heritage values and their important wetlands (Bergstrom and Selkirk, 2000; Bergstrom et al., 2002; Australian Antarctic Division, 2009; Kiernan and McConnell, 1999; Scott, 1989).

Fig. 2 shows the topography of Heard Island. The island contains a large volcano, Big Ben, which is covered in ice, snow, and glaciers. However, the low lying coastal regions (approximately 20% of the island) are covered by many different vegetation communities. It is the only large sub-Antarctic island free of introduced predators and, because of its remote location, human visits are rare. Any observed changes to the island are likely to have resulted from causes that are not directly related to local human intervention. Therefore by observing changes in the vegetation on Heard Island, a valuable insight into the regional effects of climate change can be obtained.

![Fig. 1. Location of Heard Island (source: Australian Antarctic Data Centre, 2009).](image-url)
Previous studies on Heard Island have focused on vegetation mapping based on field surveys and visual image interpretation (Scott and Bergstrom, 2006). Currently there are vegetation maps of the eastern part of the island. The overall aim of this study is to develop a more objective and automated approach using VHR satellite imagery to enable repeated mapping of the island.

2.2. IKONOS Imagery

The resolution of traditional satellite imagery is coarse compared with the resolution of the latest commercial satellites, such as IKONOS (GeoEye, 2009) and Quickbird (DigitalGlobe, 2009). For example, imagery from the Advanced Very High Resolution Radiometer (AVHRR) has a pixel size of 1km x 1km in size, and was considered to be of a high resolution (NOAASIS, 2005) when the instrument was launched in 1978. However, this imagery is now considered to be very coarse, as recent satellites acquire imagery where a pixel can represent an area as small as 0.6m x 0.6m on the ground (DigitalGlobe, 2009).

The area that was selected for this study is Paddick Valley (53°14’S; 73°68’E), on the southeast coast of Heard Island. The imagery used was acquired by the IKONOS satellite in January 2004.

The original IKONOS imagery contained four multispectral bands with 4 m pixels, along with one panchromatic band with 1 m pixels. Fig. 3 shows a false colour composite of the east side of Heard Island using the 4 m near-infrared, red, and green bands. The location of Paddick Valley is shown by the small rectangle. The 4 m multispectral bands were pansharpened with the 1 m panchromatic band resulting in 1 m multispectral bands. The process of pansharpening was performed by Space Imaging (GeoEye, 2009). Orthorectification was carried out with 163 ground control points (GCPs), a 10 m resolution digital elevation model (DEM) derived from RADARSAT imagery acquired in 2002, and the RPC parameters for the IKONOS sensor. This was done to correct any geometric distortion that could have occurred due to the angle of image acquisition and topographic relief distortion. To fine-tune the rectification process a rubber-sheeting (or triangulation) transformation was applied to the orthorectified image based on the 163 GCPs to ensures that the DGPS coordinates match the points on the image correctly. The final image used in this study is a subset of 1900 by 1600 pixels with four spectral bands at 1 m resolution covering Paddick Valley. Details of the IKONOS sensor are provided in Cook et al. (2001).
2.4. Field samples

Field sampling data were kindly supplied by Prof. Robinson’s team at Wollongong University. These data were collected during a field campaign in 2004 (at the time of IKONOS image acquisition) (Brandner et al., 2003). These data were collected by randomly sampling ten 1 m by 1 m field quadrats within a 30 m radius. The results from these plots were then used to determine a vegetation class for the 30 m by 30 m block. This sample size was chosen to simulate a 30 m SPOT pixel, as SPOT imagery was used in their study. Only the centre of the 30 m sample area was recorded with DGPS. To make these field samples usable for image classification, 30 m buffer zones were created around each block’s centre point, and these reference areas were then classified according to the Brandner et al. (2003) vegetation communities. Further reference areas were created to reflect the water, rock, and wave classes. These reference areas were digitised as regions of interest in ENVI (ITTVIS, 2009) (Fig. 4).

A three level hierarchy of vegetation classes was used to represent different thematic levels of classification (Table 1). The highest level contains only the broadest land cover classes: vegetation, rock, water and waves, whereas the lowest level contains more specific classes representing the six vegetation communities.

The highest level of classification was labelled the 1c-level, as it contains one vegetation class. It also contains the water, waves, and rock classes, as these four classes describe most of the spectral variance in the image. Although this study focused specifically on vegetation classes, it was important to define water, waves, and rock as separate classes rather than a single non-vegetation class, as these classes have distinctly different spectral characteristics. The next level of classification was labelled the 3c-level, as it separates three vegetation classes. These are sparse, medium, and dense vegetation. The most detailed level of classification was labelled the 6c-level, as it separates six different vegetation communities. These communities are shown in Table 1.

Multiple disjoint regions of interest (ROIs) were used to represent each of the land cover classes. For each of the classes a random sample of 200 pixels was selected for use as training data. Another random sample of 200 pixels was selected for use as validation data. No pixel that was used as training data was used as validation data. No pixel that was used as training data was used as validation data.

Table 1 Three-level hierarchy of vegetation classes.

<table>
<thead>
<tr>
<th>1c (Broad)</th>
<th>3c (Medium)</th>
<th>6c (Communities)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rock</td>
<td>Rock</td>
<td>Rock</td>
</tr>
<tr>
<td>Water</td>
<td>Water</td>
<td>Water</td>
</tr>
<tr>
<td>Wave</td>
<td>Wave</td>
<td>Wave</td>
</tr>
<tr>
<td>Vegetation</td>
<td>Sparse</td>
<td>Fellfield (sparse)</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>Tussock (medium)</td>
</tr>
<tr>
<td></td>
<td>Dense</td>
<td>Close Cushionfield (dense)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Open Cushionfield (medium)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mossfield (medium)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Herbfield (dense)</td>
</tr>
</tbody>
</table>

Fig. 4. Subset of the IKONOS image of Paddick Valley showing the locations of the field samples.

The patchy tundra-like vegetation on Heard Island is typical of the sub-Antarctic containing predominantly herbs, grasses and bryophytes. Additionally, the high abundance of the cushion-forming Azorella selago sets it apart from other islands (Scott and Bergstrom, 2006). The vegetation is most developed in areas that are low lying and ice free along the coast. Substantial vegetation is rarely found at altitudes greater than 250 m above sea level. There are only 12 vascular plant species recorded on Heard Island, due to the isolation and limited area available for colonisation, making this the smallest vascular plant flora of any major sub-Antarctic island group. These few vascular plant species are widely distributed across a large range of environments, resulting in highly variable morphologies and levels of performance in different habitats (Bergstrom et al., 2002; Scott and Bergstrom, 2006).

The major environmental variables affecting vegetation on sub-Antarctic islands are wind exposure, water availability, parent soil material, salt spray, and trampling (nutrient enrichment) and trampling by seals and seabirds. Bergstrom et al. (2002) considered that altitude is another important factor. Six basic plant community categories are used in this study to classify the vegetation of Heard Island: closed cushionfield, open cushionfield, fellfield, mossfield, herbfield, tussock. Closed cushionfield is a dense vegetation community that often forms extensive carpets of Azorella selago or cushion plant. Open cushionfield is the most widespread vegetation community found on Heard Island. Regions of open cushionfield are interspersed with various plants. Fellfield is a community that has abundant bare ground, and the total vegetation covers less than 50% of the area. Tussock is a medium dense vegetation community found along coastal areas, and often has an influx of elephant seals on it. Mossfield and herbfield are rich vegetation communities with varying amounts of ground cover. These communities were proposed by Bergstrom et al. (2002) and slightly modified by Brandner (2005) after ordination analysis of field samples. More information on the vegetation communities can be found in Scott and Bergstrom (2006). Despite the low number of vascular plants on the island and the relatively small proportion of ice-free land available for vegetation, these plant assemblages exhibit sufficient variation to warrant classification into sub-communities.

The rapid changes occurring in Heard Island vegetation communities in the absence of human influence present excellent opportunities for measuring and predicting changes. The potential for alien introductions has been identified as one of the biggest conservation issues facing sub-Antarctic islands as frequency of their propagules arrive (Bergstrom and Chown, 1999; Bergstrom et al., 2002). One of the aims of this study is to derive a baseline vegetation community map based on remote sensing techniques, allowing quantification of vegetation dynamics in the future.
validation data, and vice versa. This process is called stratified random sampling.

3. Texture-based classification

Gonzalez and Woods (1992) describe texture as the various measures of smoothness, coarseness, and regularity of a region in an image. For digital imagery, texture quantifies the way each pixel in an image relates to its neighbouring pixels within a small neighbourhood (window) centred on the pixel. The emphasis of image processing techniques for VHR images has shifted from traditional pixel-based techniques to contextual and object-oriented approaches as the size of the object of interest is often larger than one pixel. Many studies have explored the use of texture measures for improving classification or segmentation results by including the spatial domain (Arivazhagan and Ganesan, 2003; Blaschke and Strobl, 2001; de Jong and van der Meer, 2004; Tso and Mather, 2001; Zhu and Yang, 1998).

Spatial measures, such as texture, can play a vital role in the analysis of VHR imagery. With the VHR imagery that is now available, considering pixels in relation to their immediate neighbourhood rather than considering individual pixels in isolation should yield improved results as demonstrated by Aguera et al. (2008); Blaschke and Strobl (2001); Chiu (2004); Ouma et al. (2008) and Puissant et al. (2005). Therefore, texture contained in VHR satellite imagery should contain enough useful information to extract regions of vegetation from an image. This was recently demonstrated by Tsai et al. (2005) and Tsai and Chou (2006) who applied the GLCM to Quickbird imagery to detect invasive plant species. Some of the possible approaches for analysing texture and deriving texture measures are explained below.

3.1. Approaches to texture

The commonly used techniques for discriminating texture can be classified as either structural or statistical (Gonzalez and Woods, 1992; Nalwa, 1993). Structural techniques classify textures by looking for repeating patterns and other structural characteristics in the image. These techniques work best with imagery of artificial environments, but are not effective with imagery of natural environments (Nalwa, 1993). For a foundational summary of texture processing techniques see Haralick (1979).

Statistical techniques classify textures by performing statistical operations (such as calculating the local standard deviation) on the image. The most common way of doing this is to focus on a small window of pixels and by moving this window over the image. Different statistical measures are derived from the pixels within the window, and the results of these measures are then associated with the window’s centre pixel (Parker, 1997). These methods are most suitable for natural image scenes (Nalwa, 1993). Franklin et al. (2000) mention that the final result of the texture analysis relies heavily on the size of this window. The ideal size of the window depends on a number of factors, including the spatial resolution of the image (Ruiz et al., 2004). If the window is too small then it does not contain enough information about the area to perform an accurate analysis. However, if the window is too large, then it can overlap with other types of ground cover, and produce erroneous results. This is referred to as the edge effect (Aplin, 2006; Chen et al., 2004; Franklin et al., 1996; Puissant et al., 2005). It is possible to combine a number of different window sizes to perform multi-scale texture analysis. Coburn and Roberts (2004) found that when this multi-scale approach was applied to imagery of forested areas, classification accuracy improved by up to 8%.

Franklin et al. (2000) mention that it is important to be aware of the border and edge effects when using texture measures. The edge effect can occur when the window is too large, and therefore covers more than one class. The border effect occurs at the border of the image. As a pixel needs to be at the centre of the window to have a texture measure calculated for it, pixels on the border of the image cannot be analysed, resulting in misclassification. These edge and border effects are the main limitations of texture analysis (Ruiz et al., 2004). Ferro and Warner (2002) give an excellent discussion and demonstration of border and edge effects.

3.2. Texture measures

There are many texture measures to choose from so it is necessary to restrict the investigation to those texture measures that are likely to be potentially relevant for this imagery. The Heard Island imagery shows regions of differing textures that gradually change from one type to another. There are few cases where vegetation is separated by an abrupt boundary. Also, the imagery is of a natural area, so exact re-occurring patterns are rare, and therefore structural texture measures are not appropriate for this study. Therefore, our investigation focused on statistical texture measures.

Examples of textures contained within the Heard Island imagery are shown in Fig. 5. Here it can be seen that the texture of closed cushion vegetation (Fig. 5(a)) is quite different to that of tussock vegetation (Fig. 5(b)). Therefore, texture should prove to be valuable for distinguishing vegetation classes.

Fig. 5. Variation of texture between different classes of Heard Island imagery: (a) Closed cushion plant; (b) Tussock.
3.3. The Grey Level Co-occurrence Matrix (GLCM)

The grey level co-occurrence matrix, or GLCM (Haralick et al., 1973; Parker, 1997; Shapiro and Stockman, 2001) is one of the most widely used texture measures, and was first suggested as a mechanism for deriving texture measures by Haralick et al. (1973). Hall-Beyer (2007) presents an excellent tutorial, showing in detail how to perform the calculations. These calculations are computationally expensive, but this has become less noticeable with advances in computing hardware. The implementation provided within the IDL/ENVI software (ITTVIS, 2009) was used for this study.

The GLCM operates by computing a matrix that is based on quantifying the difference between the grey levels of neighboring pixels in an image window. This matrix essentially quantifies the spatial pixel structure within this window. The GLCM is then used to compute a number of texture features that are used to represent the texture (Parker, 1997).

The GLCM is constructed by considering the relationship between two pixels at a time. These pixels are referred to as the reference and neighbour pixels. In this study, the neighbour pixel was at a displacement of (1,1) to the reference pixel. That is, the neighbour pixel was the pixel located one pixel above and to the right of the reference pixel. Given a large enough window size, any offset could be used, but (1,1) is the most commonly used offset (Hall-Beyer, 2007). The matrix is then created with dimensions of \( S \times S \), where \( S \) is the number of possible discrete grey level values. In this study, the grey values of the input pixel values were quantised to 64 discrete levels which led to a 64 by 64 element GLCM.

The GLCM is calculated for an image window of a given size. The size of this window has a large effect on the final texture features. The optimal size of this window depends largely on the image and features being classified. Aplin (2006), Chen et al. (2004), and Franklin et al. (1996) studied the effect of the GLCM window size on classification results. We follow their approach by classifying the image based on a range of GLCM window sizes.

Once the matrix has been calculated summary statistics, characterising the structure of the matrix, are used as texture measures for each pixel. Most of these features are sums of the values within the matrix, but with different weights for each element (Haralick et al., 1973; Haralick, 1979; Hall-Beyer, 2007). For this study, the following eight widely used GLCM texture features were selected: contrast, dissimilarity, homogeneity, angular second moment, entropy, mean, variance and correlation (Haralick et al., 1973). These eight texture measures fall into three highly correlated categories. Contrast, dissimilarity, and homogeneity are contrast based; angular second moment and entropy are \( \text{orderliness} \) based, and the mean, variance, and correlation are \( \text{statistically} \) based. As features in the same category are highly correlated there is no need to use more than one from each category (Hall-Beyer, 2007). Therefore, each feature was individually compared to the other features within the same category, and the most descriptive feature from each category was selected.

4. Feature Reduction, Classification and Validation

4.1. Feature Reduction

In multispectral satellite imagery, some of the spectral bands are likely to be closely correlated with others, thus the image as a whole contains much redundant information. This redundant information can significantly reduce classification accuracy as well as increase the computational resources required to carry out the classification process. This phenomenon is known as the curse of dimensionality (Scott, 1992). It is therefore desirable to reduce the number of bands being analysed, provided this does not significantly reduce the information content in the image.

Principal Component Analysis (PCA) (Jenson and Waltz, 1979) is a well-known technique used for dimensionality reduction. This is accomplished by transforming the feature space so that the first band contains the most variance, the second band the second-most variance, and so on with the last band containing the least amount of variance. This results in the first few bands containing most of the useful information, and the later bands containing mostly noise. Therefore, these later bands can be removed without significantly affecting the information content of the image. In this study, PCA was performed on all eight GLCM texture features, and the first three principal components (PCs) were used in classification.

4.2. Classification

Texture-based classification techniques extend the traditional pixel-based classifiers by adding texture bands to the multispectral bands. Spectral and spatial information are then combined in the classifier. Two classifiers were used in this study: a minimum distance classifier and a maximum likelihood classifier. The minimum distance to mean classifier (shortened to minimum distance) is a basic classifier that calculates the multivariate Euclidean distances in feature space between an unlabelled pixel and the means or centroids of each class. The unlabelled pixel is classified according to the class with the shortest Euclidean distance. The shortcoming of this classifier is that it does not take into account the spectral variance in each class. In this study, we only use the minimum distance classifier to compare individual bands, however, all multi-band classifications were performed using a maximum likelihood classifier.

The maximum likelihood classifier is a widely used classification algorithm based on Bayes' theorem and founded in probability theory. This classifier is parametric in the sense that it models the statistical distribution of classes in multivariate feature space by class means and variance-covariance matrices, effectively approximating the class shape with a hyper-ellipsoid. The main (sometimes limiting) assumption of this classifier is that it assumes a multivariate normal distribution of spectral values of pixels within a class (Lu and Weng, 2007; Richards and Jia, 2006; Tso and Mather, 2001). For both classifiers, implementations provided within the IDL/ENVI software (ITTVIS, 2009) were used.

4.3. Accuracy assessment

The error matrix is a technique which was used for assessing the accuracy of classification. One axis represents the ground truth pixels, and the other axis represents the pixels assigned to classes by the classifier. This gives a matrix showing how many pixels of each class were actually classified into each class. An overall accuracy is calculated from this matrix by taking the sum of the elements on the diagonal and dividing this by the total number of pixels. Another statistic that is generated from the error matrix is the Kappa coefficient. This statistic takes into account all the values in the matrix, and produces a value that indicates how much of an improvement there is compared to randomly allocating pixels to classes (Congalton and Green, 2008).

5. Experimental Procedure

This section describes the experimental procedure used in the study. These experiments and tests were carried out using IDL and ENVI (ITTVIS, 2009). The ENVI platform provides an excellent environment for viewing, manipulating, and classifying remote sensing data. Some additional functionality was added especially...
for this study, and was written in IDL. All of these experiments were performed using the 1c-, 3c-, and 6c-levels within the three-level hierarchy of classes. The procedure can be broken down into three main steps:

1. classifying imagery using only the multispectral bands;
2. classifying imagery using only texture features; and
3. combining both the texture and multispectral classification approaches.

5.1. Multispectral image classification

Pixel-based multispectral classification is the traditional way of classifying satellite imagery. This was done using the four pansharpened multispectral bands. This step was performed to provide a benchmark against which to compare the texture and combination approaches. The performance of a standard classification technique needs to be known in order to measure whether incorporating texture information improves classification accuracy.

Theoretically, the best result should be obtained when classifying all four multispectral bands. However, if too many bands are used in classification, especially with the maximum likelihood classifier, classification accuracy can deteriorate due to the curse of dimensionality. Hence, if the number of multispectral bands used in classification can be reduced without significantly reducing classification accuracy, there will be fewer bands in total when these multispectral bands are combined with texture.

The goal of the multispectral classification step is two-fold; firstly, to determine the best classification accuracy that can be achieved with the Paddick Valley imagery for each group of classes, and secondly, to represent most of the information contained within the four multispectral bands in two bands, such that the two bands perform almost as well as all four bands. Reduction to two bands was chosen as the goal because one band is unlikely to be sufficiently descriptive on its own, and reduction to three bands does not reduce the data volume enough to be of any significant benefit. Fig. 6 gives an overview of the multispectral classification steps.

First, all four multispectral bands were classified using the maximum likelihood classifier. Second, each spectral band was classified individually with the minimum distance classifier. The minimum distance classifier does not produce results as accurate as those obtained using the maximum likelihood classifier, but still gives a good indication of how the bands compare with each other. Each band was then evaluated for accuracy using the overall accuracy calculated from the error matrix. The two bands that produced the best result were then classified using the maximum likelihood classifier. Third, a PCA was applied to the four multispectral bands. To ensure that the bands were equally scaled before applying PCA, the correlation matrix was used to generate statistics rather than the covariance matrix. The first two bands generated by PCA were then classified using the maximum likelihood classifier.

5.2. Texture Classification

Texture analysis was performed to determine the optimal window size and best texture features to use. This step is necessary since the parameters for texture can be determined, thus producing an improved combination of multispectral bands and texture features. Therefore, the goal of this step was three-fold; firstly, to determine the optimal window size for the Paddick Valley imagery, secondly, to determine what texture features or subset of texture features to use, and thirdly, to obtain the best results possible using only texture measures for vegetation classification.

The texture features were applied to the image of Paddick Valley. Fig. 7 gives an overview of the methods and processes used for the texture classification step. First, all texture features were used together in classification. Then, a PCA was applied to all texture features, and the first three PC bands were used in classification. Finally, each texture feature was classified individually and one feature from each texture category was used in classification.

Texture features were calculated using the Grey Level Co-occurrence Matrix (Section 3.3). All eight features (contrast, dissimilarity, homogeneity, angular second moment, entropy,
mean, variance, and correlation) were calculated using window sizes ranging from 3 to 21 pixels (only odd numbers). This calculation produced ten different texture images with eight feature bands each. The texture was only calculated for band 4 (NIR) of the IKONOS image as this band contains most of the vegetation information.

The first test involved classifying all features for each window size with the maximum likelihood classifier. This test utilised all information from all features, and theoretically should produce the best result. For the second test a PCA was applied to all eight texture features. The first three bands produced by the PCA were then classified using the maximum likelihood classifier.

Hall-Beyer (2007) suggests that only one feature from each texture category (contrast, order, statistical) should be used. Each feature was therefore classified individually using the minimum distance classifier; then the overall accuracy provided by the error matrix was used to determine the best feature from each category. The three best features combined were then classified using the maximum likelihood classifier.

To discover the optimal window size, the previous classification steps were performed over a range of window sizes (3–21 pixels). By using the overall accuracy provided by the error matrix, we observed how the accuracy varied as the window size increased. This test was also repeated for each of the different hierarchical levels of vegetation classes. The optimal window size was taken to be the smallest size which still provided adequate classification accuracy to minimize the edge effects.

5.3. Spectral and Texture Combined

The combined spectral and texture technique works by taking one or more bands from the multispectral step (either original bands, or those produced through PCA), and one or more bands from the texture step (either original bands, or those produced through PCA). The technique produces an image with a number of spectral bands and a number of texture bands, which was then classified using the maximum likelihood classifier.

The following four combinations of multispectral and texture bands were used in classification:

1. All multispectral bands and texture bands;
2. All multispectral bands and the first three of the resulting texture bands after PCA;
3. The best two multispectral bands and the first three of the texture bands after PCA;
4. All multispectral bands and the best texture bands from each category.

The combined classification technique with the highest overall accuracy was then compared with the multispectral-only and texture-only techniques.

6. Results and Discussion

This section presents and discusses the results that were obtained from the methods described in Section 5. These results are presented in three parts: multispectral classification, texture classification and combined multispectral and texture classification. As described in Section 2.4 there are three levels of thematic classes used in this study. These are referred to as the 1c-, 3c-, and 6c-levels, as shown in Table 1.

6.1. Multispectral Classification: Paddick Valley

Multispectral classification was performed to provide a benchmark, to determine the optimal spectral bands for the combination approach, and to assess the loss in accuracy with a reduction in the number of spectral bands. A summary of the results for multispectral classification is presented in Table 2.

For each classification, the near-infrared band (band 4) outperforms all other bands, indicating that band 4 is the most discriminating band for classifying vegetation. Also, as the number of vegetation classes increases, the amount that the overall accuracy for band 4 is greater than the other three bands increases. For example, when considering the 1c classification, band 4 is only 3.00% more accurate than band 1. However, band 4 is 14.25% more accurate with the 3c classification, and 14.83% more accurate with the 6c classification. Therefore, it can be seen that band 4 contains much more discriminating information about vegetation classes than any other band.

Band 1 is the second most accurate spectral band. However its accuracy changes in proportion to bands 2 and 3. Therefore band 1 is equally good at discriminating between all of the classes, rather than just vegetation. Once the performance of each band was assessed, the best two bands (bands 4 and 1) were classified using the maximum likelihood classifier.

Table 3 presents the overall accuracy and kappa coefficients obtained from the different multispectral techniques for the 1c, 3c, and 6c classifications. These different approaches are classified using all four spectral bands, the first two bands produced by PCA, and the best two spectral bands. All of these approaches used the maximum likelihood classifier.

It can be seen that using all of the multispectral bands produces the best result for all classifications. The approaches using only two bands do not contain as much information, so therefore do not perform as well. For classification within the 1c and 3c classifications, there is less than 1% difference in overall accuracy between using all spectral bands, and the approaches using two bands. Using only two spectral bands may assist in minimising the problem of dimensionality when these bands are combined with texture in the combination approach.

### Table 2

<table>
<thead>
<tr>
<th>Classes</th>
<th>Spectral Band</th>
<th>Overall Accuracy</th>
<th>Kappa Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1c</td>
<td>All 4 Bands</td>
<td>99.87</td>
<td>0.9983</td>
</tr>
<tr>
<td></td>
<td>Best 2 Bands</td>
<td>99.25</td>
<td>0.9900</td>
</tr>
<tr>
<td></td>
<td>PCA Bands 1 and 2</td>
<td>99.63</td>
<td>0.9950</td>
</tr>
<tr>
<td>3c</td>
<td>All 4 Bands</td>
<td>87.42</td>
<td>0.8490</td>
</tr>
<tr>
<td></td>
<td>Best 2 Bands</td>
<td>87.25</td>
<td>0.8470</td>
</tr>
<tr>
<td></td>
<td>PCA Bands 1 and 2</td>
<td>87.17</td>
<td>0.8460</td>
</tr>
<tr>
<td>6c</td>
<td>All 4 Bands</td>
<td>78.78</td>
<td>0.7612</td>
</tr>
<tr>
<td></td>
<td>Best 2 Bands</td>
<td>75.28</td>
<td>0.7219</td>
</tr>
<tr>
<td></td>
<td>PCA Bands 1 and 2</td>
<td>72.56</td>
<td>0.6913</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Classes</th>
<th>Technique</th>
<th>Overall Accuracy</th>
<th>Kappa Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1c</td>
<td>All 4 Bands</td>
<td>99.87</td>
<td>0.9983</td>
</tr>
<tr>
<td></td>
<td>Best 2 Bands</td>
<td>99.25</td>
<td>0.9900</td>
</tr>
<tr>
<td></td>
<td>PCA Bands 1 and 2</td>
<td>99.63</td>
<td>0.9950</td>
</tr>
<tr>
<td>3c</td>
<td>All 4 Bands</td>
<td>87.42</td>
<td>0.8490</td>
</tr>
<tr>
<td></td>
<td>Best 2 Bands</td>
<td>87.25</td>
<td>0.8470</td>
</tr>
<tr>
<td></td>
<td>PCA Bands 1 and 2</td>
<td>87.17</td>
<td>0.8460</td>
</tr>
<tr>
<td>6c</td>
<td>All 4 Bands</td>
<td>78.78</td>
<td>0.7612</td>
</tr>
<tr>
<td></td>
<td>Best 2 Bands</td>
<td>75.28</td>
<td>0.7219</td>
</tr>
<tr>
<td></td>
<td>PCA Bands 1 and 2</td>
<td>72.56</td>
<td>0.6913</td>
</tr>
</tbody>
</table>

However, for the 6c classification, there is a significant difference in overall accuracy. Classifying with all four bands still produces the highest overall accuracy, but the approaches using only two bands perform significantly worse. If only bands 1 and 4 are used, the overall accuracy drops by 3.50%, and if the first two PCA bands are used, the overall accuracy drops by 6.22%. This drop in accuracy is significant, therefore feature reduction is not recommended for the 6c classification.

6.2. Texture Classification: Paddick Valley

Texture classification involves using just texture features to perform classification. Texture classification was performed on the Paddick Valley IKONOS image to compare the multispectral and combination approaches, to determine the optimal texture features to use for the combination approach, to determine the optimal texture window size, and to assess whether there is a significant loss in accuracy if the number of bands used to represent the texture features is reduced.

Fig. 8 shows how the overall accuracy of the 6c classification changes as the window size increases from 3 to 21. All eight features from each window size were classified with the maximum likelihood classifier. As can be seen from these figures, classification accuracy increases as the window size increases. Therefore, from a statistical perspective it would seem sensible to choose a large window size, but this is not necessarily the case.

Upon visual inspection of the image, it can be seen that many regions of vegetation are 30 pixels or less in diameter, so a small window size is needed to reduce edge effects. The optimal window size is a compromise between having a good overall accuracy and retaining a small enough window size so that edge effects become insignificant. Therefore, the window size of 7 was chosen for this study, which is where the slope of the accuracy graph decreases. Puissant et al. (2005) also used a window size of 7 when they classified urban regions in VHR satellite imagery.

Each feature was then classified individually using the minimum distance classifier. This was done to be able to compare the different features to each other. Also, as discussed in Section 5.2, it was desired to use only one texture feature from each category of texture features. Therefore, by classifying each feature individually, the best feature from each category could be identified.

Table 4 shows the overall accuracy and the kappa coefficient of each feature for the 1c, 3c, 6c classifications respectively using a window size of 7. For the contrast group, the best feature was dissimilarity; from the statistical group, the best feature was entropy; and from the orderliness group, the best feature was entropy.

The best features from each category were classified using the maximum likelihood classifier. Fig. 9 compares this technique to using all eight features and using the first three resulting bands from PCA. It would be expected that using all features would produce the best accuracy, but this does not seem to be the case. Since features within the same category of textures are highly correlated, there is no need to include more than one feature from each category (Hall-Beyer, 2007). Fig. 9 shows that classification of the first three PCA bands performs better than using all eight texture features. However, using only the best band from each texture category results in the highest classification accuracy for each level in the classification hierarchy.

In summary, a window size of 7 was chosen, and it was found that the best texture features to use are mean, dissimilarity, and entropy. It is these parameters that are expected to perform best when combined with the multispectral bands in the combination classification.

6.3. Combination Classification

Several combinations of multispectral and texture were tested:

1. All multispectral bands and texture features;
2. All multispectral bands and the first three of the texture PCA bands;
3. The best two multispectral bands and the first three of the texture PCA bands;
4. All multispectral bands and the best texture feature from each category.

Fig. 10 shows the comparison of the overall accuracy of these different combinations. For each of the 1c, 3c, and 6c classifications, combination 4 has the highest accuracy, and combination 1 has the lowest accuracy for all but the 3c classification. It was demonstrated in Section 6.1 that there is a significant loss of 6c classification accuracy when only two spectral bands are used. Therefore, the poor performance of combination 3 for the 6c classification is to be expected.
Combination 1 uses a total of twelve bands: all four spectral bands and eight texture features. Apart from combination 3, this combination consistently has the lowest overall accuracy, due to the problem of dimensionality. Combination 2, which utilises three bands produced from PCA performs better than combination 1.

As expected from the findings of Sections 6.1 and 6.2, the best overall accuracy is achieved when all the spectral bands are combined with the best texture feature from each category. This is because this technique contains only seven bands, but still retains a large amount of discriminating information. Fig. 11 shows how this combination approach compares with the multispectral and texture approaches.

Fig. 11 shows that there is no significant increase in overall accuracy for the combination approach when considering the 1c and 3c classifications. However, as the level of detail in classes increases, so does the improvement of the classification accuracy for the combination approach compared to the multispectral approach. For the 6c classification the improvement in overall accuracy is significant at 5.94%. This improvement confirms the hypothesis that texture significantly improves the classification.
accuracy of Heard Island vegetation. The classification results show that the vegetation communities on Heard Island are insufficiently characterised by the four spectral bands from VHR satellite imagery. Similar improvements in overall accuracy have been achieved in other studies. For example, Chiu et al. (2004) reported a 5% increase in accuracy for mapping wetlands when combining texture with multispectral bands. Puissant et al. (2005) combined texture and multispectral bands, increasing classification accuracy by 4.4% compared to multispectral classification alone. Franklin et al. (2000) reported an increase of 12% in classification accuracy when classifying forest species composition from VHR airborne multispectral images.

Finally, Fig. 12 shows the user and producer accuracies of the individual vegetation communities for the multispectral approach and combination approach. These accuracy statistics show an improvement of the classification accuracy for all vegetation communities with the combination of spectral and texture information. The error matrix shows that the highest confusion occurs between closed cushionfield, open cushionfield, tussock, and herbfield. The greatest confusion exists between herbfield and closed cushionfield, and between closed cushionfield and open cushionfield. The former can be explained by the fact that both herbfield and closed cushionfield represent very dense communities. The latter can be explained by the fact that the fellfield, open and closed cushionfield classes represent an ecotone from very sparse cushion plants with a large proportion of rocks to very dense cushionfield. There is a gradual transition between these classes, both thematically as well as spatially, rather than a crisp delineation. In future research, we are planning to use fuzzy classifiers to quantify and map these ecotones (Zhang and Foody, 2001). The increase in the user and producer accuracies for the combination approach illustrates that image texture is an important descriptor of Heard Island's vegetation communities improving differentiation between communities compared to spectral information alone.

7. Conclusion

This study investigated whether using texture features can improve the accuracy of vegetation classification within VHR satellite imagery. More specifically, we focused on two important aspects of the grey level co-occurrence matrix (GLCM): selection of texture features derived from the GLCM, and the optimal size of the texture window. In addition, we systematically explored several combinations of multispectral and texture-based classifications for a three-level hierarchy of land cover and vegetation classes (one (1c), three (3c), and six (6c) vegetation classes in addition to three basic non-vegetation land cover classes).

Firstly, a multispectral classification was performed on a subset of a pansharpened IKONOS image of sub-Antarctic Heard Island. We found that using all four multispectral bands produced the highest overall classification accuracy. However, when classifying only 1 or 3 vegetation classes, using only two of the four IKONOS bands (bands 1 and 4) does not significantly reduce accuracy. Band 4 (infrared) contains more information about vegetation than any of the other bands.

Secondly, GLCM texture analysis was performed on the image using an optimal window size of 7 pixels. Smaller window sizes did not provide an accurate description of the texture, and larger window sizes produced unacceptable edge effects. Out of the eight basic GLCM texture features we found that using the features of mean, dissimilarity, and entropy provided the highest overall classification accuracy. By selecting only these three key texture features, the negative effect of a high dimensional feature space on classification can be reduced.

Thirdly, these three feature textures were combined with all four spectral bands to produce a combination classification approach. When classifying at the 1c- or 3c-levels, the combination approach did not significantly increase the accuracy. However, when classifying at the 6c-level the combination approach increased accuracy by almost 6% to an overall accuracy of 84.7%. In summary, this study supports the hypothesis that combining texture with multispectral classification improves vegetation classification of VHR satellite imagery. This is the first study to produce a baseline vegetation community map of sub-Antarctic Heard Island using IKONOS imagery and texture-based classification techniques.

Acknowledgements

The authors wish to thank the Australian Antarctic Data Centre (AAD) at the Australian Antarctic Division (AAD) for providing the IKONOS image used for this study and for the maps of Heard Island and the McDonald Islands. Dr Jenny Scott has kindly provided DGPS points for image orthorectification and valuable field samples and photographs. Assoc. Prof. Sharon Robinson is acknowledged for providing field samples for Paddick Valley. Finally, we would like to thank the anonymous referees who have provided valuable feedback on an earlier version of this paper.

References


