

An Ensemble of Classifiers Approach to Coral Distribution Mapping

A M. MUSLIM^{a,*}, D. David.^a, Z. BACHOK^a, W.A. SABRI^a

^aInstitute of Oceanography (INOS), Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia

Abstract - Accurate, up-to-date and accessible information on the state of coral reef ecosystem is necessary for informed and effective management of these important marine resources. However, environments containing these habitats are challenging to map due to their remoteness, extent and costs of monitoring. In this research, the capabilities of satellite remote sensing techniques combined with *in situ* data were assessed to generate coral habitat map of Lang Tengah Island. Several classification techniques were utilized in identifying coral distribution. This study aims at utilizing a new approach to increase accuracy, mainly using an ensemble of classifier. The core principle of the ensemble methodology is to weigh several individual classifiers, and combine them in order to reach a classification that is better than the one obtained by only one classifier. This ensembles offer promise in boosting the overall classification accuracy. Initially 5 classifiers were used to classify the study area mainly, Parallelepiped, Minimum distance, Maximum likelihood, Fisher and K-Nearest Neighbor. Results from the classification shows that each method produced different accuracy based on the bottom type. The K-Nearest Neighbor method produced the most accurate results for dense coral class (71.43 %) while for sand class Maximum likelihood was the most accurate (70.37%). Using these results an ensemble of classification approach was applied and results show that accuracy was the highest with an overall accuracy of 73.02 % in comparison to Parallelepiped (52.38), Minimum distance (50.79), Maximum likelihood (60.37), Fisher (31.75) and K-Nearest Neighbor (50.79)

Keyword: Ensemble of Classifier, Hard Classification, Coral Reef Mapping.

1. INTRODUCTION

Coral reefs are considered to be one of the most magnificent marine ecosystems on earth displaying an extremely rich biodiversity (Jill *et.al.*, 2002). They also represent valuable socio-economic resources (Jill *et.al.*, 2002). Many threats put stress on coral reefs (Jill *et.al.*, 2002 and Eric *et.al.*, 2002). The combination of local stresses from overfishing, physical destruction (Lokeet *et.al.*, 2002), coastal pollution and increasing sedimentation (Clive, 2008) together with the growing threat from climate change may result in permanent degradation of the coral reef ecosystem. This complex marine ecosystem represents an economically important ecosystem in Malaysia as the foundation of a significant percentage of the country's tourist industry. However, a lack of comprehensive management programs is leading to degradation of this important economic resource. This situation becomes worse by inadequate information on the status and location of the reefs, further hindering management efforts. The conventional methods such as topographic map, manta tow and LIT are commonly used for mapping and monitoring of reef

habitat. However, this method may not be practical for monitoring large or remote coral reef ecosystem (Eric *et.al.*, 2002) as it is expensive and time consuming (Mumby *et.al.*, 2003). So there is a need to advocate improved methods for mapping coral reefs and providing rapid assessment of the changes and stress responses of this important and threatened ecosystem.

Remote sensing may provide the information about the configuration and composition of coral reefs (Mumby *et.al.*, 2003) and for monitoring and management. Remotely sensed data seems to be potential source of useful information as these maps require frequent time updating, time consuming and expensive by conventional methods.

In this study, the capabilities of satellite remote sensing techniques are combined with *in situ* data collection and assessed for generating coral reef distribution map of Lang Tengah. A multi spectral Quickbird imagery with a spatial resolution of 2.5 m was acquired of the study area and used as the basis of the research.

This study emphasize on the application of ensemble classifiers to improve coral classification. An ensemble approach constructs a set of predictive models (called a classifier). An ensemble gives a prediction for a new data instance by combining the predictions of its classifier for that instance. For regression tasks, the predictions can be combined by averaging the outputs of the classifier. In order for an ensemble to be more accurate than any of its individual members, the individual classifier needs to be accurate and diverse. An accurate classifier is one that performs better than random guessing on new examples. A set of classifier is diverse if the classifier makes different errors on new examples. The diversity in an ensemble can be set (changing the weight of examples or changing the weight of attributes) or by manipulating the learning algorithm used to obtain the models.

Several classifiers were used to classify the Quickbird imagery with different degrees of accuracy in classifying benthic types. The classification techniques used were Parallelepiped, Minimum distance, Maximum likelihood, Fisher and K-Nearest Neighbour. Using the ensemble approach these results were later combined to produce a more accurate benthic habitat map.

2. STUDY AREA

Lang Tengah is one of the nine islands in Redang Archipelago in Malaysia. It lies between 5°47'45" north and 102°53'45" east, off the Terengganu coastline on the north-eastern coast of Malaysia. This island has been gazette as a Marine Park. The brilliant underwater world of this island is one of the best in Malaysia where their crystal clear sea-water, pristine beaches and untouched tropical jungle become the main attraction to the tourist to come.

* Corresponding author: Tel: +609 668 3191; Fax: +609 669 2166. Email address: aidy@umt.edu.my

** The authors wish to acknowledge MOSTI for funding this research.

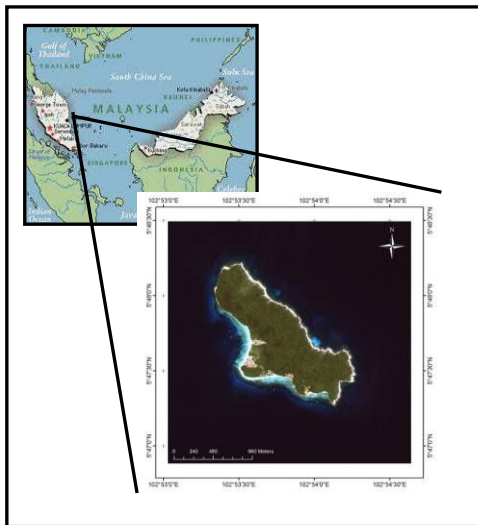


Figure 1. Location of Lang Tengah

3. METHODOLOGY

3.1 Data Sets and Ground Data

This research utilised a multi-spectral Quickbird imagery of the study area. The imagery was captured on 25th August 2005. For ground verification an accuracy assessment, field data was collected around Lang Tengah Island on 1st of December 2009 and 10th of February 2010. This field survey is essential to identify benthic habitats types such as coral, rubble and sand that are present in the study area.

As there is a difference of about 5 years between satellite data acquisition and field observation, a general survey were conducted to determine the status of coral in 2005 with diving operators at the island. The field observation was updated with this information. An example would be if a certain location is observed as dead coral in 2010 but was coral in 2005 (based on dive operator information) it's reclassified as coral. As this paper focuses on the classification techniques being developed classification errors between 2005 and 2010 would be systematically introduced in the results thus maintaining the integrity of comparison between the performances of the classifiers.

The equipment consists of a differential GPS (dGPS) to accurately determine the positions of different types of benthic habitat surrounding the island. The bottom type was determined by a conventional method such as diving, snorkeling and manta tow.

3.2 Image Pre-Processing

Image of Lang Tengah was geometrically corrected where a total of 10 Ground Control Points (GCPs) were used, which resulted in a Root Mean Square (RMS) error less than 0.5 pixels.

In order to obtain a depth independent spectral measurement of the benthic type, it is necessary to compensate for the water column effect (Mumby *et al.*,2000) from the influence of the water column which was accounted for by the empirical

approaches "Depth-Invariant Bottom-Indices" (Lyzena 1978, 1981). Standard atmospheric corrections were applied to correct for path radiance.

3.3 Image Processing

3.3.1 Image Classification Process

In an ensemble of classifier approach several supervised classification output need to be derived from the satellite imagery. Supervised classification is known to be accurate for mapping classes, but depends on the cognition and our skills to manage it. The recognition of conventional classes or meaningful classes in a scene from prior knowledge, such as personal experience with what is present in Lang Tengah, or more generally, the region it's located in, by experience with thematic maps, or by on-site visits will make it easier to choose the training site.

Before the training site was formed, the thematic information from water bodies was extracted where all upland and terrestrial features were masked out from the image. This land mask was applied to restrict the spectral range of brightness values to aquatic features and allow detailed feature discrimination. When this method is properly applied, the diversity of spectral features within a given scene decreases as the level of spectral discrimination increases.

As mention earlier, training pixels were defined based on visual image interpretation, with the guidance of field description. Pixels of well-known ground areas were selected as training sites. The benthic habitat class characteristic was assigned following the classification scheme. The habitat classes are describing in Table 1.

Table1. Descriptive for field survey data.

Habitat Class	Characteristics
Dense coral	More than 50% is coral- covered substrate. Including hard coral, benthic algae and sponges.
Sparse coral	Less than 50% is coral covered substrate.
Dead coral	Algae colonizing <i>dead coral</i> / skeletons, dead coral skeletons, rubble.
Sand	Carbonate sand/ rubble with occurrence of sparse green algae.

Classification algorithms were run to cluster the data representing each class. Classification then proceeds by statistical processing in which every pixel is compared with the various signatures and assigned to the class whose signature comes closest. Different supervised techniques were applied to the image of Lang Tengah. Figure 2 presented the thematic images that generated using different supervised techniques mainly Parallelepiped, Minimum distance, Maximum likelihood, Fisher and K-Nearest Neighbour. Using these 4 classification output classes with the highest accuracy within the classification techniques were combined to produce single classification imagery. The images were overlaid producing ensemble classification imagery. Classification accuracy for each technique based on benthic habitat class is shown in Table 2.

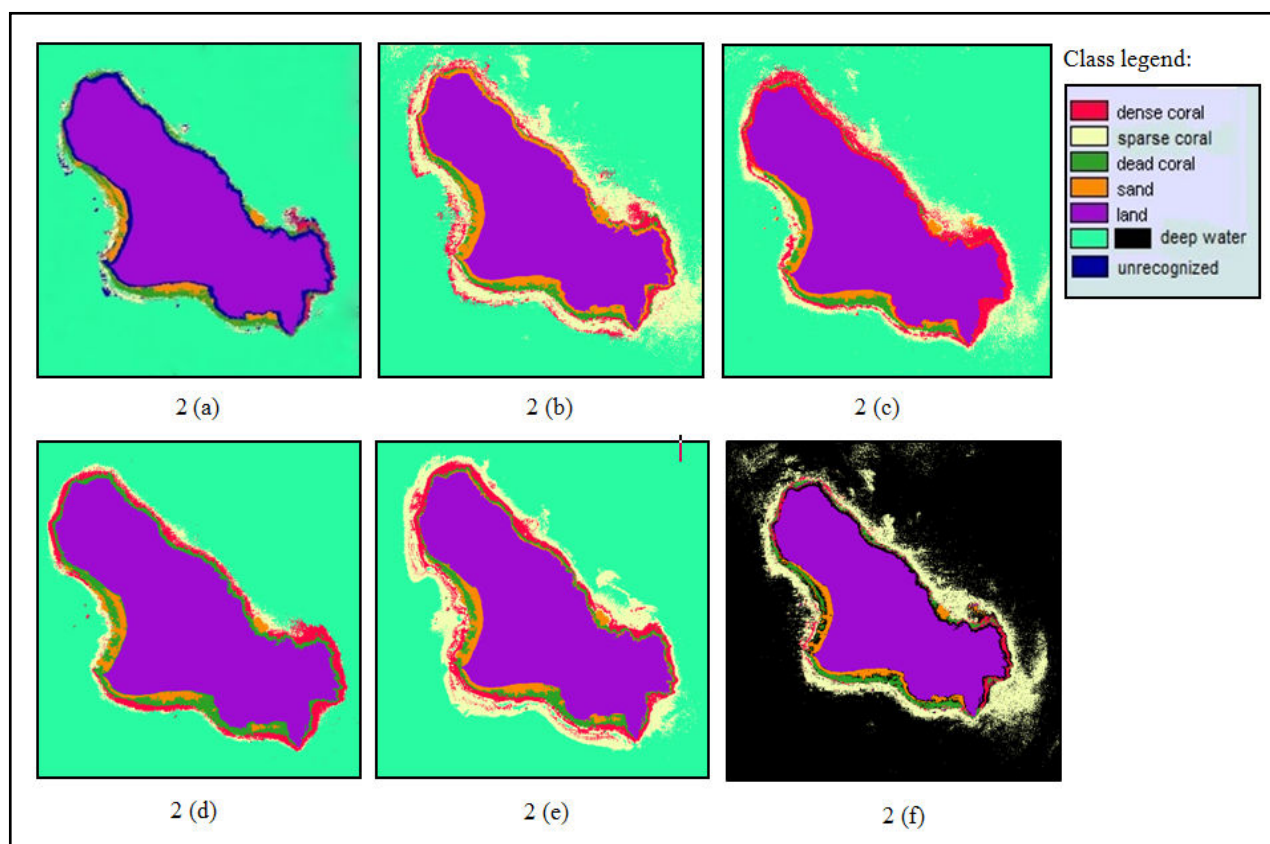


Figure2. Thematic images that generated using different supervised techniques and ensemble of classifier ; (a) Parallelepiped (b) Minimum Distance (c) Maximum Likelihood (d) Fisher and (e) K-Nearest Neighbour (KNN) (f) Ensemble of Classifier.

Table 2. Accuracy for every technique

Classifiers	Accuracy for every class (%)				Total % Accuracy
	dense coral	dispersed coral	dead coral	sand	
Parallelepiped	43	50	71	48	53
Minimum distance	43	88	36	52	55
Maximum likelihood	57	63	43	70	58
Fisher	50	75	29	11	41
K-Nearest Neighbour	71	75	36	41	56
Ensemble of Classifier	71	88	71	70	75

4. RESULT AND DISCUSSIONS

4.1 Accuracy Assessment

In this study, a comparative evaluation of the classified images with *insitu* data revealed the overall accuracy of the techniques used. Accuracy assessments were performed for each of the classified images where the number of reference points or ground truth sites was used to test the overall classification accuracy of the benthic habitat classification. Table 2 shows the percentage of every class produced by different techniques with their total accuracy. The output maps from different supervised techniques had an overall moderate accuracy with range from 41% to 56%. As reported by Mumby *et al.* in 1998, although such results are only moderate by statistical standards, these are not unrealistic values, especially in remote sensing of marine environments as this could be caused by a number of factors such as the delay in the time of ground truthing and the time when the satellite images was taken and the water column correction algorithm may not remove the effects of water attenuation completely.

The image also shows a response with significant confusion in discriminating most of the benthic classes. This may be caused by the spectral that overlaps between optically similar objects. In this study, the algal over growth is often similar to zooxanthellae densities and pigmentation occurring in coral features while the calcium carbonate skeleton of dead coral always misinterpreted as sand.

In this study, the ensemble classification considers only the supervised technique that constructed on highest accuracy of their benthic classes which then were combined to increase the accuracy and to form better and more robust thematic image. In this technique some areas are classified as unrecognised, which are produce when combining the classification classes. This is due to conflicts between classes and accuracy between the classification output. Further research would be conducted to resolve this.

5.0 CONCLUSION

Different classification techniques vary in accuracy when classifying habitat classes. Their performance relies on how a certain class is distributed and its surrounding areas. For example maximum likelihood classifier is the most accurate in classifying sand but not at detecting coral. By combining the most accurate classification output into a single ensemble of classification, the accuracy of the produced map is increased. But further research needs to be done to determine the most accurate method of combining the classes and dealing with conflicts between accuracy and classes.

REFERENCES

- Andre'foue't, S. and Riegl, B. (2004) Remote sensing: a key tool for interdisciplinary assessment of coral reef processes. *Coral Reefs*, vol 23, p.p. 1-4, 2004.
- Clive, W. (2008). Status of Coral Reefs of the World: 2008. Global Coral Reef Monitoring Network.
- Eric, J.H., Marlin, J.A. and Serge, A. (2002). Spectral Reflectance of Coral Reef Bottom-Types Worldwide and Implications for Coral Reef Remote Sensing. *Remote Sensing of Environment*, vol 85, p.p. 159-173.
- John, A. R. and Xiuping, J. (2006). *Remote Sensing Digital Image Analysis*. Springer. Germany, p.p. 193-207.
- Jill, M., Sunil, N., Donald, C.R., Richard, L.P., John, Schalles, Kevin, H. and Jennifer, K. (2002). Classifying and Mapping General Coral Reef Structure Using IKONOS Data. *Photogrammetric Engineering & Remote Sensing*. 1297-1303.
- Loke, M.C., Vo, S.T., Philreefs, Thamasak, Y., Annadel, C., Suharsono, and Ith, K. (2002). Status of Southeast Asia Coral Reefs.
- Miller, S. L. and Crosby, M. P. (1998). The Extent and Condition of US Coral Reefs. In *NOAA's State of the Coast Report*. http://state-of-coast.noaa.gov/bulletins/html/crf_08/crf.html. Accessed on 26 August 2009.
- Mumby, P.J., William, S., Alan, E.S., John, T.H., Ellsworth, F.L.D., Eric, J.H, Rick, P.S. and Laura, T. D. (2003). Remote Sensing of Coral Reefs and Their Physical Environment. *Marine Pollution Bulletin*, vol 48, p.p. 219-222, 2004.
- Rokach, L. (2009). *Ensemble-based Classifier*. Springer vol 33, p.p. 1-39.
- Sharon, K. (2003, December 1). Guardians of the Natural Heritage. *The Sun*, p.p. 19.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of Institute of Oceanography and its staff, Mr.Mohamad Nasir Bin Mohamad, Mr.Azri Bin Mohamad. We would also like to thank Department of Marine Park Malaysia for permission to conduct our research. Last but not least, the authors would like to thank Mr. Mohd Kaharudin, Mr.J azlan, Mr. Najmudin, Mr.Khairul Ihsan and Miss Nurliyana for assisting during field verification.