SEGMENTATION OF MULTISPECTRAL IMAGES AND PREDICTION OF CHI-A CONCENTRATION FOR EFFECTIVE OCEAN COLOUR REMOTE SENSING

Jinchang Ren[1], Xuexing Zeng[2], and David McKee[3]
jinchang.ren@strath.ac.uk, x.zeng@cugb.edu.cn, David.McKee@strath.ac.uk

[1] Department of Electronic and Electrical Engineering, University of Strathclyde, Scotland, U.K.
[2] Dept. of Information Engieering, China University of Geosciences, Beijing, China.
[3] Department of Physics, University of Strathclyde, Scotland, U.K.

ABSTRACT

With the development of new sensors and data processing techniques, ocean colour remote sensing has undergone rapid development in more accurately measurement of coastal shelf classification and concentration of chlorophyll. In this paper, multispectral images are employed to achieve these targets, using techniques including region-growing based segmentation for pixel classification and support vector regression for ChI-a prediction. Interesting results are reported to show the great potential in using state-of-the-art data analysis techniques for effective ocean colour remote sensing.

***Index Terms*—** ocean colour remote sensing; coastal classification; chlorophyll concentration measurement; image segmentation; multispectral/hyperspectral imaging.

1. INTRODUCTION

Multispectral/hyperspectral imaging recently has been applied in a wide range of application areas such as remote sensing [1], forensics [2], pharmaceuticals [3] and food analysis [4]. Hyperspectral imaging collects information from across the electro-magnetic spectrum, and thus produces dense sampling in the spectral domain, and can provide much richer information and better discrimination ability than visible light images.

As a particular application in remote sensing, ocean colour remote sensing has transformed our ability to monitor dynamic relationships between physical and biogeochemical processes that underpin the role of natural waters in the global carbon cycle and the redistribution of suspended and dissolved materials across the globe. With more than a decade of continuous daily global coverage observations, there is a growing archive of ocean colour data that has the potential to act as an essential Global Climate Variable that may inform studies of regional, global and rapid climate change.

With the rapid development of sensor and data processing techniques, ocean colour remote sensing has undergone rapid changes in the degree of sophistication of our understanding of signal measurement technology and data interpretation issues. Sensor technology has developed from early satellite sensors with four operational wavebands (e.g. CZCS), through the current generation of multispectral sensors (e.g. SeaWiFS, MODIS and MERIS) to the recent deployment of the first operational hyperspectral sensors (e.g. HICO) [5]. These technological advances have supported the development of improved atmospheric correction processes, new and upgraded product algorithms (e.g. chlorophyll, inherent optical properties, diffuse attenuation coefficients…) and new ways of partitioning the global ocean into bio-optical (biogeochemical) provinces (e.g. IOCCG Report 9).

This paper will focus on two important tasks of ocean colour remote sensing, i.e. segmentation based coastal shelf classification and ChI-a measurement. In ocean colour remote sensing, data prediction and data classification are emphasized [5], and typical approaches include support vector machine (SVM) [6] and k-means clustering [7]. In general, SVM requires that the appropriate training datasets are selected, and thus it may not suitable for the case that includes too many classifications or the case that very small region belongs to some classifications. In addition, K-means is very sensitive to the initial parameters and noise. As a result, in our work, we will focus on the method of region growing [8] for hyperspectral datasets segmentation. For ChI-a measurement, support vector regression is applied. Relevant techniques are discussed in detail in the next section.

2. THE APPROACH

With multispectral image data used as input, the first task is to identify various coastal shelf regions. According to the remote sensing reflectance data, segmentation based classification is employed, in which images are segmented into regions and followed by pixel based clustering for data classification. For multiple spectral image segmentation, seedless region growing is applied to maintain the spatial coherency when similar pixel vectors are grouped together.

Let I represent a M bands multispectral image as input, and I_m denotes one band image, where $m \in [1, M]$. Let S_{ij} represent a seed pixel for region growing whose spatial co-ordinate is (i, j). Starting from the top-left pixel S_{11}, we sequentially scan the image to identify any pixel which has not been segmented into any groups. Then, this pixel will be picked up and used as a new seed for region growing. Pixels which are spatially adjacent to the seed pixel and satisfy certain conditions will be grouped into the clusters of the new seed. For any pixel which has been grouped into the cluster, its spatial adjacent pixels will be iteratively examined. This process will be repeated until all the pixels have been checked one by one in an iterative way.

Similarity constraint is the major criterion used in determining whether a pixel needs be grouped into a cluster or not. For the pixels in a cluster, their mean vector μ and co-variance matrix Φ can be obtained. A new pixel can be grouped into the cluster if its pixel vector $\mathbf{p}$ satisfy

$$(\mathbf{p}-\mu)^T \Phi^{-1} (\mathbf{p}-\mu) < \theta \quad (1)$$

where θ is a given threshold.

Please note as the number of pixels in a cluster may increase step by step, the corresponding mean vector μ and co-variance matrix Φ need be updated accordingly. At the first stage when there is only one pixel in the cluster, we set μ as the 10% of the seed pixel, and Φ the unit matrix (or the identify matrix).

Although pixel based region growing helps to segment the input image into regions, different regions at a distance may have similar average vector and need be further clustered as they actually represent the same kind of coastal shelves. To achieve this, for each pair of segmented regions, the similarity of their average spectral vector is calculated. If the similarity is over a given threshold, the two regions are clustered together. Eventually, each region is assigned to a new index as the results of region clustering.

For the second task of ChI-a concentration prediction, support vector machine (SVM) is employed for training and prediction. With available ground truth data, support vector regression is applied to map from input spectral data to the ChI-a concentration, using a radial basis function (RBF) kernel. The learnt model is then applied for pixel based prediction, i.e. to predict ChI-a concentration for each pixel vector extracted from the multispectral image. For pixels clustered into a group, the average ChI-a value is also obtained for comparison. In fact, RBF kernelled SVM has been proven successful in a number of other applications [9-12], and detailed description of SVM and other machine learning approaches can be found in [1, 4, 9-10].

3. EXPERIMENTS AND RESULTS

The multispectral ocean dataset around U.K. that collected on May, 2007 was used for coastal shelf classification. This dataset include 9 bands with the following wavelengths: 412, 433, 488,

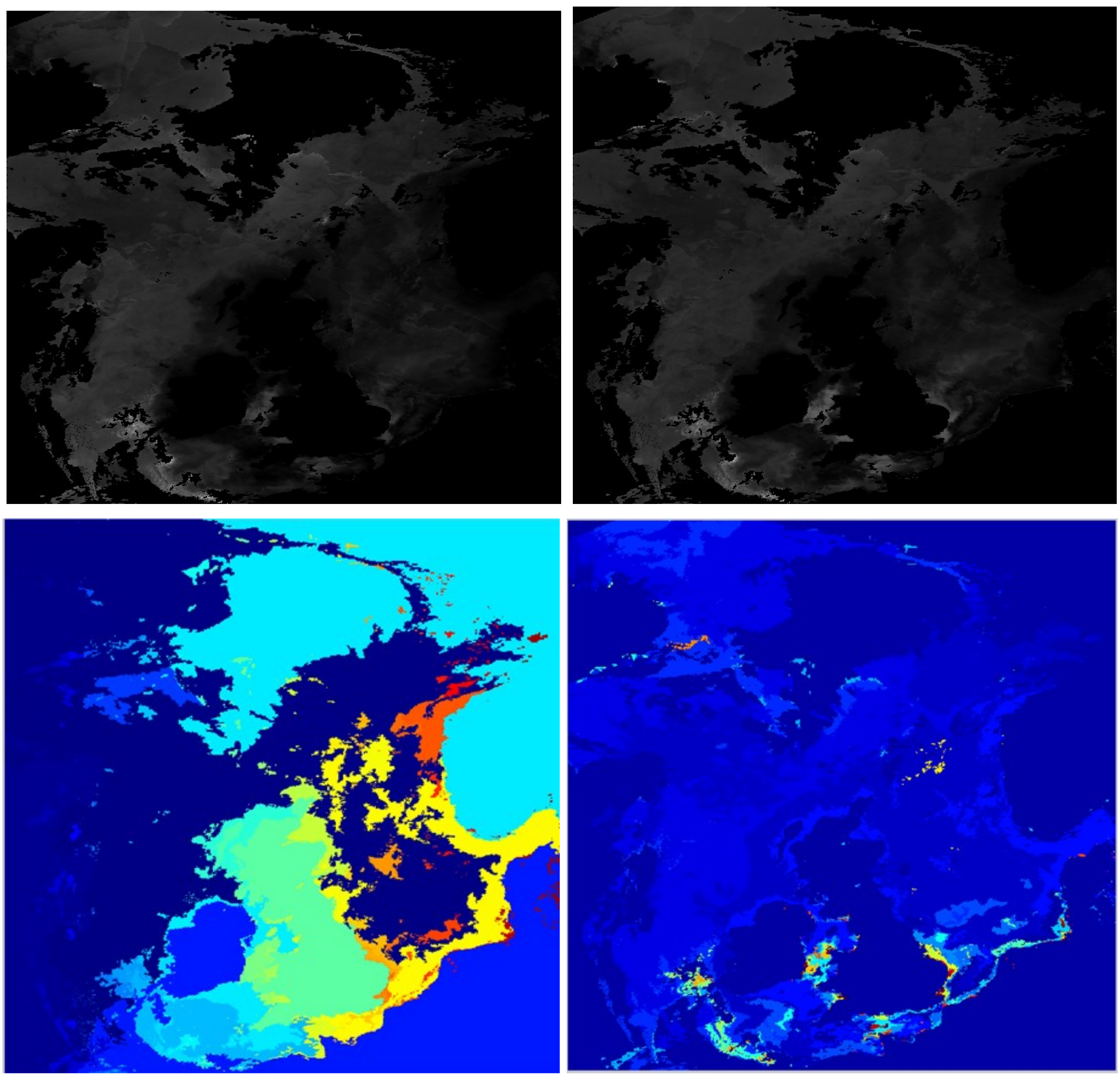

Fig.1: The first two band images (top), the segmented results (bottom-left, over 2200 regions) and the final result after clustering of similar regions (only 31 clusters).

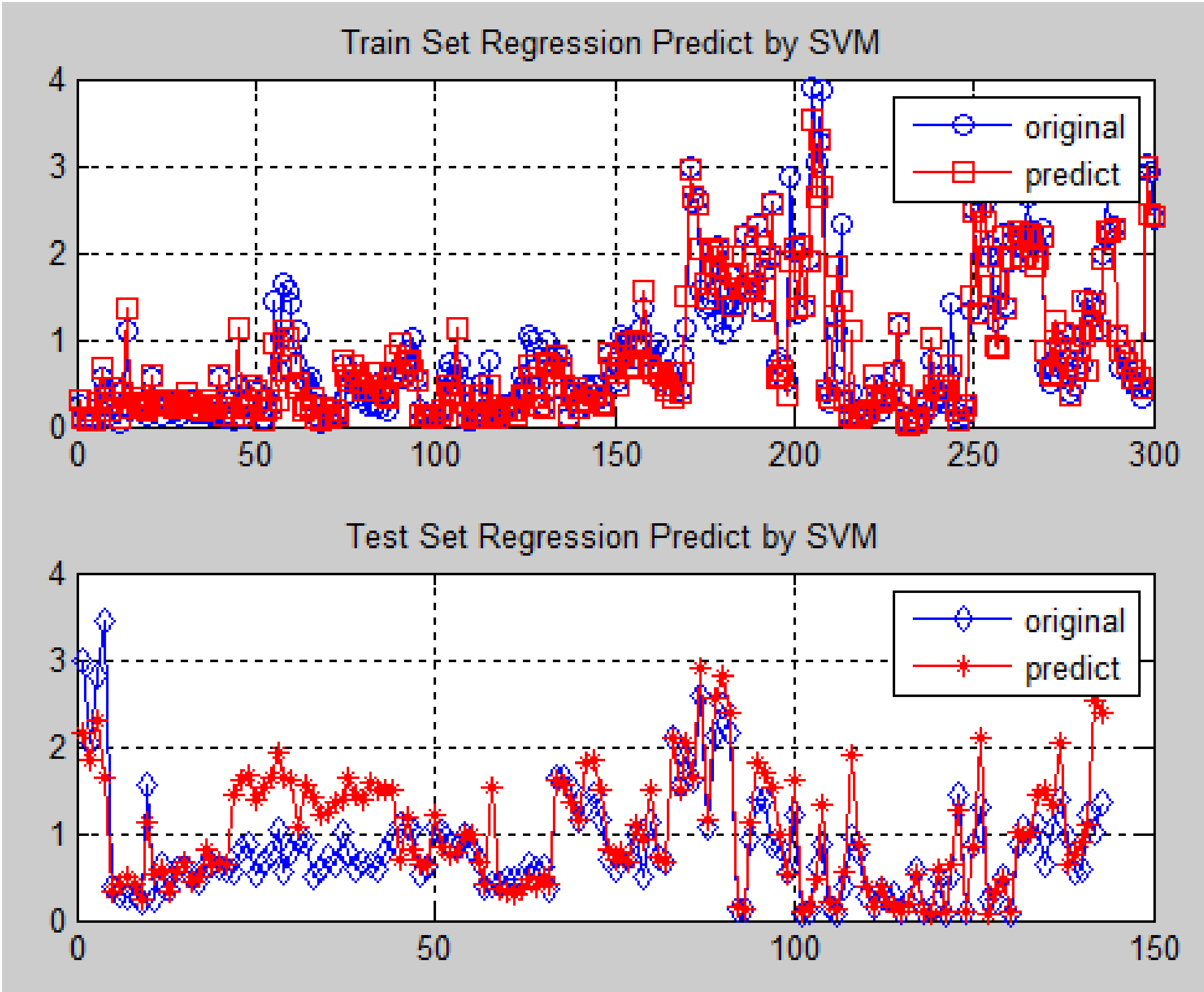

Fig. 2: Results of ChI-a concentration prediction from multispectral data.

531, 547, 667, 678, 748 and 869 nm respectively. The spatial dimension of the image is 1000 by 1000 pixels.

The first two band images are shown in Fig. 1, along with the segmented results from pixel based region growing and region clustering. As can be seen, region-growing based segmentation has successfully segmented the multi-spectral image. However, the number of regions is over 2200 which is too large to be interpreted by human experts. After region clustering, most lands and open water regions are grouped together. Afterwards, we have only 31 clusters remained, which can be found closely adjacent to the coastal lines and are more easily interpretable for ocean physicists in checking their physical meaning.

To predict the ChI-a concentration, another dataset from NOMAD (NASA bio-Optical Marine Algorithm Data set, http://seabass.gsfc.nasa.gov/seabasscgi/nomad.cgi) was used. After removal of missing data, 443 valid samples are obtained. In total 300 samples were used for training the SVM and the remaining 143 for testing. The squared correlation coefficients of training and testing achieved by us are 0.91 and 0.63, respectively. In addition, the mean squared error of training and prediction is 0.004 and 9.14, respectively.

It is worth noting that based on the extracted coastal shelf lines, we can define our region of interest and then apply ChI-a concentration prediction on these regions accordingly. This can not only improve the efficiency in data processing but also enhance the efficacy of the data prediction as ocean physicists can then combine the information together in more effective interpreting and predicting the observed data.

4. CONCLUSION

In this paper, two tasks for multispectral ocean colour remote sensing are covered. For the first task, seedless region growing followed by region clustering is used for classification of coastal shelf regions. For the second task, SVM is used for predict of ChI-a concentration from multispectral data. Promising results from real datasets are produced from the proposed approaches.

5. REFERENCES

[1] J. Ren, et al, “Effective feature extraction and data reduction in remote sensing using hyperspectral imaging,” IEEE Signal Process. Mag., vol. 31, no. 4, pp. 149–154, Jul. 2014.

[2] K. Gill, et al, “Quality-assured fingerprint image enhancement and extraction using hyperspectral imaging,” in ICDP 2011, London, UK, 2011.

[3] L. Zhang, et al, “Multivariate data analysis is for Raman imaging of a model pharmaceutical tablet,” Analytica Chimica Acta, Vol. 545, pp. 262-278, 2005.

[4] T. Kelman, et al, “Effective classification of Chinese tea samples in hyperspectral imaging,” Artif. Intell. Res., vol. 2, no. 4, pp. 87–96, Oct. 2013.

[5] D. McKee, et al, “Optical water type discrimination and tuning remote sensing band-ratio algorithms: Application to retrieval of chlorophyll and K_d (490) in the Irish and Celtic Seas”, Estuarine, Coastal and Shelf Science, Vol. 73, pp. 827-834, 2007.

[6] F. Melgani and L. Bruzzone, “Classification of hyperspectral remote sensing images with support vector machines”, IEEE Trans. On Geoscience and Remote Sensing, 42(8): 1778-1790, 2004.

[7] M. S. Alam, et al, “Object detection in hyperspectral images by using K-means clustering algorithm with a preprocessing,” Proc. of the SPIE, vol. 6574, pp. 1-9, 2007.

[8] R. Adams and L. Bischof, “Seeded region growing”, IEEE Trans. on Pattern Analysis and Machine Intelligence, 16(6): 641-647, 1994.

[9] J. Zabalza, et al, “Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing,” ISPRS J. Photogrammetry Remote Sens., 93(7): 112–122, Jul. 2014.

[10] J. Zabalza, et al, “Novel 2D singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging,” IEEE Trans. Geoscience and Remote Sensing, 2015.

[11] J. Alkhateeb et al, “Multiclass classification of unconstrained handwritten Arabic words using machine learning approaches,” Open Signal processing, 2, 2009

[12] J. Zabalza, et al, “Singular spectrum analysis for effective feature extraction in hyperspectral imaging,” IEEE Geoscience and Remote Sensing Letters, 11(11): 1886-1890, 2014.

[13] T. Ijitona, et al, “SAR sea ice image segmentation using watershed with intensity-based region merging,” IEEE Int. Conf. on Computer and Information Technology (CIT), pp. 168–172, 2014.

[14] J. Ren, et al, “Fusion of intensity and inter-component chromatic difference for effective and robust colour edge detection,” IET Image Processing, 4(4), 2010

[15] C. Zhao, et al, “Improved sparse representation using adaptive spatial support for effective target detection in hyperspectral imagery,” Int. J. Remote Sensing, 34(24):8669-8684, 2013.