UNIVERSITI PUTRA MALAYSIA

IMPROVING CLASSIFICATION OF REMOTELY SENSED DATA USING BEST BAND SELECTION INDEX AND CLUSTER LABELLING ALGORITHMS

TEOH CHIN CHUANG.

FK 2005 49

IMPROVING CLASSIFICATION OF REMOTELY SENSED DATA USING BEST BAND SELECTION INDEX AND CLUSTER LABELLING ALGORITHMS

By

TEOH CHIN CHUANG

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

May 2005

Specially Dedicated To:

My Beloved Wife,

Ong Keat Khim

My Beloved Parents,

Teoh Gin Soon and Hoo Siew Eng

Abstract of the thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirements for the degree of Doctor of Philosophy

IMPROVING CLASSIFICATION OF REMOTELY SENSED DATA USING BEST BAND SELECTION INDEX AND CLUSTER LABELLING ALGORITHMS

By

TEOH CHIN CHUANG

May 2005

Chairman: Professor Shattri Mansor, PhD

Faculty: Engineering

Methods for improving supervised and unsupervised classification of remotely sensed data were developed in this study. Supervised classification of remotely sensed data requires systematic collection of training samples for classes of interest. Image visual interpretation is important in training samples collection because it incorporates association information of surrounding pixels, such as texture and context, hence making the training samples collection process more easy and accurate. Once training samples for each class are collected, the training statistics for each class and band are extracted to select those bands, which are most effective in discriminating each class of information from all others for classification. In remote sensing application, deciding the best band combination for image visualization and classification is relatively difficult and time consuming. In addition, the best band selected for image classification is not necessarily the best for classification.

A Best Band Selection Index (BBSI) algorithm was developed which is capable of selecting the best band combination for image visualization and supervised classification. This BBSI is calculated by two components, one based on class mean (or cluster mean) difference and the other based on correlation coefficients. Using Landsat Thematic Mapper (TM) and Modis/Aster Airborne Simulator (MASTER) images as the test datasets, the BBSI algorithm was compared to the Optimum Index Factor (OIF) algorithm in selection of the best three-band combination for image visualization. The comparison results between BBSI and OIF indicated that, both algorithms correctly predicted the best three-band combination that provided useful information for image visualization in the Landsat TM dataset. However, both algorithms tested on MASTER dataset produced different results. The image quality of band combination selected by BBSI was smoother and better than OIF.

The BBSI was also compared to the Jeffreys-Matusita distance (JM-distance) algorithm in selection of the best four-band combination for supervised classification of Landsat TM and MASTER datasets. The comparison results between BBSI and JM-distance showed that, both algorithms accurately selected the best four-band combination that yielded the highest overall accuracy classification map with value of 91% in the Landsat TM dataset. Meanwhile, the comparison results in the MASTER dataset showed that, the overall accuracy classification map for band combination selected by BBSI with value of 89.7% was slightly higher than band combination selected by JM-distance with value of 89.2%.

Unsupervised classification of remotely sensed data consists of cluster generation and cluster labelling steps. A method was developed to improve the cluster generation and clusters labelling processes in unsupervised classification of the Landsat TM and MASTER datasets. In cluster generating process, the developed BBSI algorithm was used to select the best band combination for generating cluster by using Iterative self-Organizing Data Analysis (ISODATA) technique. The cluster generation results showed that, the BBSI accurately selected the best four-band combination generating very low mixed classes of clusters.

In cluster labelling process, a cluster labelling algorithm based on calculation of minimum-distance (MD) between cluster mean and class mean was developed to label the clusters. This algorithm was compared to co-spectral plot method for labelling clusters the clusters generated in Landsat TM dataset. The comparison results show that, the clusters labelled by the cluster labelling algorithm were the same as using co-spectral plot. The cluster labelling algorithm was also compared to maximum-likelihood supervised classifier in the production of classification map for MASTER dataset. The comparison showed that, the accuracy of the unsupervised classification map with value of 88.4% that was generated by using the cluster labelling algorithm was slightly more than the maximum-likelihood supervised classification map with value of 87.5%. The advantage of the cluster labelling algorithm compared to co-spectral plot and maximum-likelihood classifier was the algorithm provided a rapid production of high accuracy classification map.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PEMBAIKAN PENGKELASAN DATA REMOTE SENSING MENGGUNAKAN ALGORITMA INDEKS PEMILIHAN JALUR TERBAIK DAN ALGORITMA PELABELAN KELOMPOK

Oleh

TEOH CHIN CHUANG

Mei 2005

Pengerusi: Profesor Shattri Mansor, PhD

Fakulti: Kejuruteraan

Kaedah pembaikan pengkelasan yang terselia dan pengkelasan yang tidak terselia bagi data remote sensing telah dibangunkan di dalam kajian ini. Pengkelasan yang terselia bagi imej remote sensing memerlukan pengumpulan sampel-sampel latihan untuk kelas-kelas yang diminati secara teratur. Pentafsiran imej secara penglihatan adalah penting di dalam pengumpulan sampel-sampel latihan kerana cara ini mengambilkira maklumat berkaitan seperti tekstur dan konteks untuk piksel-piksel di sekeliling. Ini akan membuat proses pengumpulan sampel-sampel lebih mudah dan jitu. Selepas mengumpulkan sampel-sampel latihan, maklumat statistik sampel-sampel latihan bagi setiap kelas dan jalur diperolehi dan digunakan untuk memilih jalur-jalur yang paling berkesan. Jalur-jalur dipilih ini dapat membezakan setiap kelas dengan berkesan di dalam pengkelasan imej. Dalam aplikasi remote sensing, pemilihan kombinasi jalur-jalur yang baik untuk penglihtan dan pengkelasan imej adalah sukar dan membazirkan masa. Tambahan pula, jalur-jalur yang terpilih untuk penglihatan imej tidak semestinya sesuai untuk pengkelasan

imej. Satu algoritma Indeks Pemilihan Jalur Terbaik [Best Band Selection Index (BBSI)] telah dibangunkan, di mana ia berupaya memilih kombinasi jalur-jalur yang paling baik untuk penglihatan dan pengkelasan imej. BBSI ini dikira oleh dua komponen, satu berdasarkan perbezaan purata kelas (atau purata kelompok) dan satu lagi ialah berdasarkan pekali sekaitan. Dengan menggunakan imej-imej 'Landsat Thematic Mapper' (TM) dan 'Modis/Aster Airborne Simulator' (MASTER) sebagai dataset ujian, BBSI telah dibandingkan dengan algoritma Faktor Indeks Optimum [Optimum Index Factor (OIF)] di dalam pemilihan kombinasi tiga-jalur yang paling baik untuk penglihatan imej. Keputusan-keputusan daripada perbandingan di antara BBSI and OIF menunjukkan bahawa, kedua-dua algoritma dapat meramalkan kombinasi tiga-jalur dengan tepat, di mana kombinasi jalur-jalur tersebut dapat membekalkan informasi yang berguna untuk pentafsiran imej Landsat TM dengan cara penglihatan. Walau bagaimanapun, keputusan yang berlainan dihasilkan apabila kedua-dua algoritma tersebut diuji pada dataset MASTER. Kualiti imej yang menggunakan kombinasi tiga-jalur yang dipilih oleh BBSI adalah lebih licin dan baik daripada OIF.

BBSI juga dibandingkan dengan algoritma Jarak Jeffreys-Matusita (Jarak-JM) di dalam pemilihan kombinasi empat-jalur yang paling baik untuk pengkelasan data Landsat TM dan MASTER secara terselia. Keputusan perbandingkan di antara BBSI and Jarak-JM menunjukkan bahawa, kedua-dua algoritma dapat memilih kombinasi empat-jalur dengan tepat, di mana kombinasi jalur-jalur dipilih itu dapat menghasilkan peta pengkelasan yang paling baik dengan kejituan keseluruhan 91% untuk dataset Landsat TM. Manakala, keputusan perbandingan untuk dataset MASTER menunjukkan bahawa,

kejituan keseluruhan yang dipilih oleh BBSI adalah 89.7% dan kejituan ini adalah lebih tinggi sedikit daripada Jarak-JM yang mempunyai nilai 89.2%.

Pengkelasan yang tidak terselia bagi data remote sensing terdiri daripada langkah-langkah penjanaan kelompok dan pelabelan kelompok. Satu kaedah telah dibangunkan untuk pembaikan proses-proses penjanaan kelompok dan pelabelan kelompok di dalam pengkelasan tidak terselia untuk dataset Landsat TM dan MASTER. Di dalam proses penjanaan kelompok, algoritma BBSI digunakan untuk memilih kombinasi jalur-jalur terbaik untuk menjanakan kelompok-kelompok dengan menggunakan teknik Analisis Data Aturan-Diri Berulang [Iterative Self-Organizing Data Analysis (ISODATA)]. Keputusan-keputusan penjanaan kelompok menunjukkan bahawa, BBSI dapat memilih kombinasi empat-jalur dengan tepat, di mana jalur-jalur dipilih dapat menghasilkan kelompok-kelompok yang kurang bercampur kelas.

Di dalam proses pelabelan kelompok, satu algoritma pelabelan kelompok berdasarkan pengiraan jarak-minimum di antara purata kelompok dan purata kelas telah dibangunkan untuk melabelkan kelompok-kelompok. Algoritma ini telah dibandingkan dengan kaedah plot co-spektrum untuk melabelkan kelompok-kelompok yang dihasilkan di dalam dataset Landsat TM. Keputusan-keputusan daripada perbandingan menunjukkan bahawa, kelompok-kelompok yang dilabelkan oleh algoritma pelabelan kelompok adalah sama dengan menggunakan plot co-spektrum. Algoritma pelabelan kelompok juga dibandingkan dengan pengkelas kebolehjadian maksimum di dalam penghasilan peta pengkelasan untuk dataset MASTER. Keputusan-keputusan daripada perbandingan

menunjukkan bahawa, kejituan keseluruhan peta pengkelasan yang dihasilkan dengan menggunakan algoritma pelabelan kelompok memberi nilai 88.4% yang adalah sedikit lebih tinggi daripada peta pengkelasan yang dihasilkan oleh pengkelas kebolehjadian maksimum yang memberi nilai 87.5%. Kelebihan algoritma pelabelan kelompok dibandingkan dengan plot co-spektrum dan pengkelas kebolehjadian maksimum ialah algoritma tersebut dapat menghasilkan peta pengkelasan yang mempunyai kejituan tinggi dengan cepat.

ACKNOWLEDGEMENTS

My heart-felt gratitude goes to my supervisor, Professor Dr. Shattri Mansor for his guidance, technical assistance, encouragement and support that has made this thesis possible.

I would like to extend my appreciation to Associate Professor Dr. Abdul Rashid Mohamed Shariff, Associate Professor Dr. Noordin Ahmad and Dr. Radzali Mispan for their kind cooperation and assistance through out my study and also to Malaysia Centre for Remote Sensing (MACRES) for providing the remotely sensed dataset.

Finally, I wish to convey my special thanks to my wife and family members, who had contributed and supported me toward the successful completion of this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES