EXPERT SYSTEM BASED ON OBJECT-ORIENTED APPROACH FOR LAND COVER MAPPING

Zhang Lei, *, Zhou Yueming, Wu Bingfang

Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101, China.
- zhang@irsa.ac.cn

KEY WORDS: Expert classification, Object-oriented classifier, land cover, decision tree, segmentation

ABSTRACT:

The hybrid classifier is presented in combination of expert system and object oriented approach, for which increased information is added for classification and improve the accuracy. Instead of the original image bands, derived data are prepared for classification, which contains physical meaning and clear separation of recognition and assessment of object classification, the variables of NDVI, seasonal change vegetation index, vegetation brightness index, hard surface brightness index, moisture stress index are derived from raw image data. The 17 classes of land cover are possible for classification and make high overall accuracy of 86%, the classes of evergreen needle leaf, evergreen shrub, sands/beach, lake/reservoir get high accuracy in classification, the classes of grass, grass dominated urban vegetation, bare land/rocks have poor classification, the influence to classification is the factors of DEM, slope, shadow. Expert system based on object orient approach has many advantages of multi-spectral cluster separation and object analysis for land cover classification, especially for more than 10 land cover classes. Multiple variables are better solution in criteria setup than single variable and reduce the hierarchical levels.

1. INTRODUCTION

With increasing concerns on global environment, climate change and bio-diversity, land cover/use researches get great progresses in the recent decades, instead of the traditional approaches of classification for land cover, the new methodologies are developed, such as fuzzy logic (Foody 1996), decision tree (Breiman et al., 1984; Hansen et al. 1996), neural network (Carpenter et al. 1997), support vector machine (Chang, 2001), expert system (William, 2001), object oriented classification (Geneletti, 2003) and hybrid model (van der Meer 1995). Each classification strategy has its advantage and limitation, the univariate decision tree is not based on any assumptions of normality within training statistics and is well suited to situations where a single cover type is resented by more than one cluster in the spectral space, the decision tree yields a set of rules which are easy to interpret for an expert analyst in correcting splits associated with faulty or contradictory training data (Hansen, 2000). neural network is also independent from the data prior probability distribution and capable of self-study, self-adaptive and improve the classification accuracy 10-20% high over traditional classifiers, suitable for complicate classification (Richard, 2001); object oriented approach is taken on the vector domain information extraction of spatial geometry, spatial relations, suitable for high resolution image application, Expert systems allow the expert establish the logical decision tree based on feature space and class physical features for the universal application in other places, and integration of remotely sensed data with other sources of geo-referenced information such as land use data, spatial texture, and digital elevation models (DEMs) to obtain greater classification accuracy (William, 2001).

Nevertheless, the classification of land cover on meso-scale is still at low accuracy for large area due to the spectral mixture among classes, spatial heterogeneity and image contamination, the strategy of multiple classifier combination is considered as an approach for accuracy improvement, this paper is discussed on the combination classifiers of object oriented classifier and expert system for land cover classification. Object oriented classifier is applied for derivation of the contextual and spatial geometric information of classes from the images, especially for application on high resolution and meso-scale resolution classification, expert system is allowed to other ancillary data sources adding to classification or post-classification, such as DEM, slope which are more relative to spatial distribution of land cover, with its strategy of hierarchical structures of decision tree, it is not based on any as sumptions of normality within training statistics and is well suited to situations where a single cover type is resented by more than one cluster in the spectral space, decision trees yield a set of rules which are easy to interpret and suitable for deriving a physical understanding of the classification process.

The study area is selected in the mountain area in Three Gorge area, Chongqin, China. It occupies approximately 100 km2 in mid-latitude warm temperate zone with dominated natural vegetation of evergreen shrub and deciduous broad leaf forest. With high density of population, the agricultural land is major landscape of land cover. SPOT5 XS and Pan data are acquired between 2004-2006 for cloud free images, two scenes of different seasons for each area is selected for vegetation detection, one time scene for leaf on date is selected in July or August, which vegetation is grow up most, another scene for leaf off data is selected in January or February, which most deciduous vegetation is dormant, for the confusion of cropland

---

with natural vegetation, the phenology of crops are considered in acquisition selection, one of scens must be in harvest and another season in growing up. The top map of 1:50, 000 is come from Bureau of State Surveying and Mapping for derivation of DEM and slope data.

2. METHEDOLOGY

2.1 Data pre-processing

The images are corrected for the effects of atmosphere, and converted to calibrated reflectance using commercially available software that incorporates the MODTRAN3 radiative transfer code (ATCOR2 for the ERDAS Imagine software; GEOSYSTEMS GmbH, 1997). A middle-atitude summer/winter, rural aerosol concentration model therefore was used as input to the radiative transfer code, estimated visibility is input from ground climate station at the same day. SPOT XS and Pan data are fused to 2.5 m grid with the approach of High Pass Filter (HPF), which keep least spectral change during the fusion processing, All coverages were geo-referenced to the Beijing 54, Gauss Kruger Zone 108 or 111, coordinate system.

The radiometric influence of topography on recorded sensor signals has distorted the vegetation features and physical structure. This can lead to image classification error as well as error in forest classification and parameter estimates, The topographic calibration on reflectance values is necessary in the mountain area (Equation 1,2).

$$L(\lambda,e) = L_n \cdot (\lambda) \cdot \cos i \qquad \text{(Eq. 1)}$$

L ： solar radiance

λ ： bands

e ： ground slope

L_n ： corrected solar radiance

i ： incident angle

$$\cos i = \cos(90 - \theta_s) \cdot \cos\theta_n + \sin(90 - \theta_s) \cdot \sin\theta_s \cdot \cos(\phi_s - \phi_n)$$
$$\text{(Eq. 2)}$$

θ_s ： solar elevation angle

θ_n ： ground slope

ϕ_s ： solar azimuth angle

ϕ_n ： ground azimuth angle

Instead of the original image bands, derived data are prepared for classification, which contains physical meaning and clear separation of recognition and assessment of object classification, in this study, NDVI, seasonal change vegetation index (SCVI) (Equation 3), vegetation brightness index (VBI: Nir+Swir), hard surface brightness index (HBI: Green+Red), moisture stress index (MSI: Nir/Swir) are produced. The ancillary

$$SCVI = \frac{NDVI_{LON} - NDVI_{LOFF}}{NDVI_{LON} + NDVI_{LOFF}} \quad \ldots\ldots\ldots\text{(Eq. 3)}$$

$NDVI_{LON}$: maximum NDVI year round, normal in Month 7-8

$NDVI_{LOFF}$: minimum NDVI year round, normal in Month 1-2

data are regarded as environment factors affected indirectly on the land cover, include the DEM, Slope, Aspect, DEM and slope are applied in this classification due to their high relative to the vegetation spatial distribution. The geometric parameters is obtained from object oriented approach to derive the object parameters, such as ratio of length/width, ratio of perimeter/area, roundness, spatial relations, object hierarchical relationship of multi-scale.

2.2 Classification system

FAO land cover classification system (LCCS) is accepted in this study, LCCS is comprehensive, standardized a priori classification system, independent of the scale or means used to map. It enables a comparison of land cover classes regardless of data source, economic sector or country (Antonio, 2004). There are 17 classes of land cover are found in study area (table 1).

Table 1 FAO land cover classes in Three Gorge area

1. evergreen broad leaf
2. evergreen needle leaf
3. evergreen shrub
4. deciduous shrub
5. grass
6. tree dominated urban vegetation
7. tree and shrub dominated urban vegetation
8. grass dominated urban vegetation
9. herbaceous plantation (upland)
10. aquatic graminoids (paddy)
11. urban area
12. industry
13. sands/beach
14. bare land/rocks
15. river
16. pond
17. lake/reservoir

2.3 Hybrid classification

The hybrid classification is based on the combination of object oriented approach and expert system classification. The approach of expert system has the strategies of non-parameter separation and multi-variable decision rules, Use of the expert system has an advantage to classify multiple clusters of feature space for one class, such as upland and sands/beach, which is not allowed to operate in maximum likelihood classification (MLC) approach ，expert system has higher accuracy of classification than MLC in complicate spectral classes; The object oriented appoach is focused on the extraction of spatial geometry information, some artificial classes have similar spectral features, but have different geometric and relation properties, which are helpful to recognize and divide, such as reservoir/pond, industry and urban vegetation.

The hybrid classifier model was constructed using the Definiens 5.0 image processing software. The classification is performed in object oriented method in advance and then the expert system followed (Fig. 1), that compose of 4 major steps of segmentation, object creation, expert knowledge

Figure1 follow chart for hybrid classification of land cover

rule setup and classification. The segmentation is based on the unsupervised classification to form multiple scale cluster layers, for this study, 10 and 20 scales are selected to shape different size cluster, most parts of classed are segmented in 10 scale, but some classes with unique spectrum or the phenomenon of salt and pepper are segmented in 20 scale, such as water body and artificial classes, the purpose of scaling segmentation is to form best shape for the representation of object properties. vector objects are then transformed from the segmented clusters of

raster types to extract both object geometric and spectral information, each object spectral information includes average value, standard deviation, maximum value, minimum value of inner pixels among each band, each object geometry includes the features of area, size, spatial relation to other objects. These integrated data are input to the database of expert knowledge. According to the object classified rules, the top down decision trees are set up for classification. The selection of physical meaning classification rules and ancillary data are considered to ensure the reliable result and stable classified procedure, 13 decision trees with one or multiple criteria are made for classification of 17 classes (Figure 2). The design of expert system is based on the image spectrum, object properties and ancillary data features, the maximum differences of features are considered on upper hierarchical steps to reduce the omission errors. According to the analysis of samples on the image, upper step is to classify three groups of water-body, vegetation and non-vegetation, then classify classes in each groups.
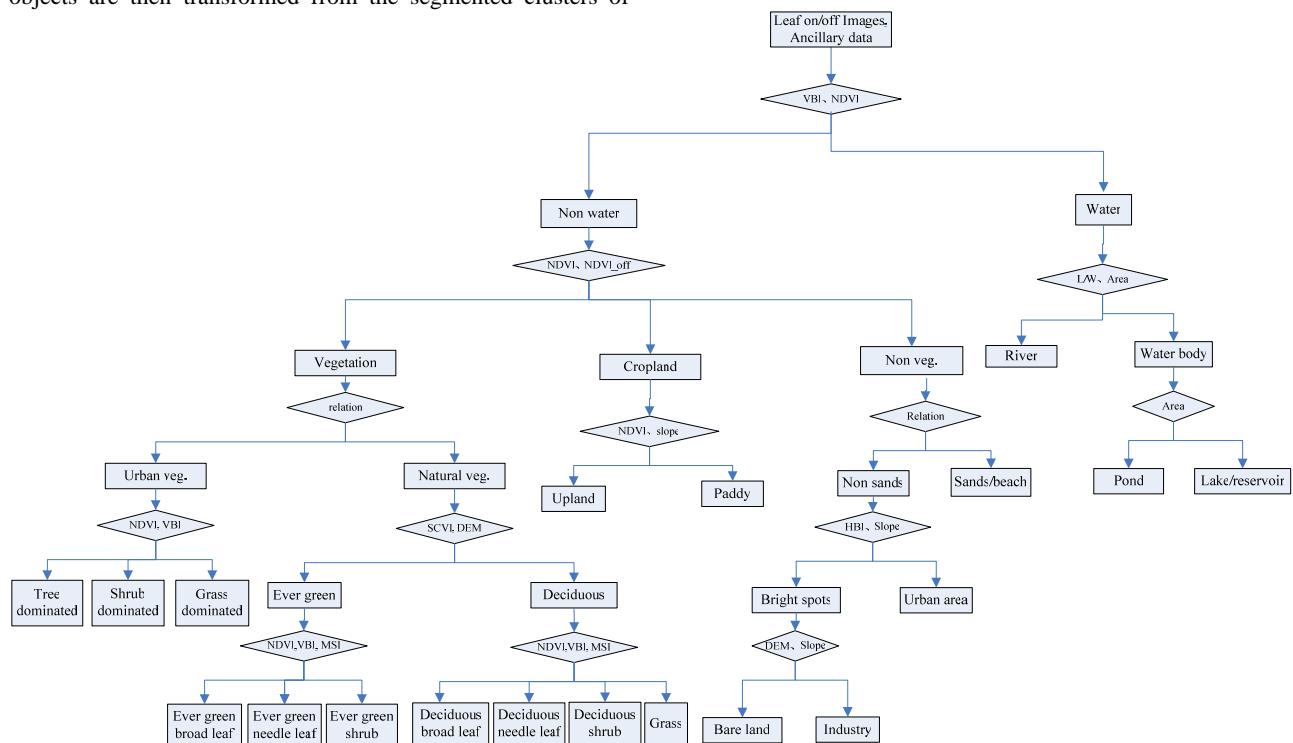
Figure 2 the classification rules and steps of expert knowledge system

3. RESULTS AND ANALYSIS

The accuracy assessment of classification was measured using randomly selected points for which land cover was determined using an orthophoto mosaic geo-referenced to the image. The product accuracy, user accuracy, overall accuracy are analyzed from the sample error matrix assessment (Table 2). Overall classification accuracies for the 17 classes reach 86%, for product accuracy, the classes with little omission numbers are evergreen needle leaf, evergreen shrub, tree dominated urban vegetation, upland, industry, sands/beach, lake/reservoir, for which the accuracy are more than 88%. For user accuracy, the classes with least commission numbers are the classes of evergreen broad leaf, evergreen needle leaf, deciduous shrub,

residence, sands/beach, pond lake/reservoir, for which the accuracy are also more than 88%. The classes with both low omission and commission are evergreen needle leaf, evergreen shrub, sands/beach, lake/reservoir which have least errors in the classification. Big errors are occurred in the classes of grass, grass dominated urban vegetation, bare land/rocks, for which accuracy has about 45%.

The water-body is easy to separate with the two variables of winter VBI and NDVI when chlorophyll content is little in the water, but the hill shadows which have same spectral feature are larger than in the summer, and result in some errors. Then river, reservoir, pond are divided based on the object orient approach which are impossible classified with MLC. The groups of

natural vegetation, cropland and non-vegetation are at the same level for classification with two variables of winter NDVI and summer NDVI, the rules are two high NDVI values for the class of vegetation, one high NDVI for cropland and two low NDVI for non-vegetation, non-vegetation obtains high accuracy classification, but cropland is mixed with natural vegetation, because some crops during harvested season cover with natural grass instead of bare land, other reason is local elevation difference resulting in the crop phenological change, and spectral heterogeneity produces some errors in classification. SCVI is better variable for separation of deciduous and evergreen vegetation, VBI is more sensitive to vegetation type than NDVI and MSI in the classification of tree, shrub, and grass. Urban vegetation classification is based on the rule of spatial relation of closeness to the residence, the variable of relative border ratio is used, some outside natural vegetation is classified to urban vegetation, and some inner urban vegetation object is omitted in classification.

4. CONCLUSIONS AND DISCUSSIONS

Expert system based on object orient approach has many advantages of multi-spectral cluster separation and object analysis for land cover classification, especially for more than 10 land cover classes. Multiple variables are better solution in criteria setup than single variable and reduce the hierarchical levels. The 17 classes of land cover are possible for classification and make high overall accuracy of 86%, the classes of evergreen needle leaf, evergreen shrub, sands/beach, lake/reservoir get high accuracy in classification, the classes of grass, grass dominated urban vegetation, bare land/rocks have poor classification, the influence to classification is the factors of DEM, slope, shadow. Some classes which are difficult to classify in MLC are available in hybrid classifier, such as urban vegetation from vegetation, reservoir from river.

However, the rule setup of expert system is complicate, time-consumed, and requires some experience of land cover features and spatial distribution.

5. REFERENCES

Antonio Di Gregorio, Land Cover Classification System, Food and Agriculture Organization of the United Nations, Rome,2004

Breiman, L., Friedman, J. H., Olshend, R. A., and Stone, C. J.1984, Classification and Regression Trees, Wadsworth, Monterey, CA.

Carpenter, G. A. Gjaja,M. N., Gopal, S., and Woodcock, C. E.,1997, ART neural networks for remote sensing: vegetation classi. cation from Landsat TM and terrain data. IEEE Transactions on Geoscience and Remote Sensing, 35, 308–325.

Chang C.C., Lin C.J.2001, Training v-support vector classifiers: theory and algorithms, Neural Computation Vol. 13(9), pp. 2119-2147.

Foody, G. M.,1996, Approaches for the production and evaluation of fuzzy land cover classi. cations from remotely-sensed data. International Journal of Remote Sensing, 17, 1317–1340.

Geneletti, D, GORTE B. G. H.,2003, A method for object-oriented land cover classification combining Landsat TM data and aerial photographs, International Journal of Remote Sensing, Vol.24, No.6,1273-1286.

Hansen, M., Dubayah, R., and DeFries, R.,1996, Classi. cation trees: an alternative to traditional land cover classi. ers. International Journal of Remote Sensing, 17, 1075–1081.

Hansen, M., Dubayah, R. S. DEFRIES†, J. R. G. TOWNSHEND and R. SOHLBERG.,2000, Global land cover classification at 1 km spatial resolution using a classification tree approach. International Journal of Remote Sensing, 21, No6-7, 1331-1364.

Richard O. Duda, PeterE. Etc.2001. Patern Classification. ISBN:0-471-05669-3.

Van der Meer, F.,1995, Spectral unmixing of Landsat Thematic Mapper data. International Journal of Remote Sensing, 16, 3189–3194.

William L. Stefanov, Michael. Ramsey, Philip R. Christensen,2001, Monitoring urban land cover change: An expert system approach to land cover classification of semiarid to arid urban centers, Remote Sensing of Environment 77, 173–185

type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	U accuracy
1	797186				11595		31597											95%
2	18004	3192410	22124			3613												99%
3	127727	14897	3611747	27596		7690			12889		15597							95%
4				70122														100%
5		4701	15094		63129				28210									57%
6		6513	93345	23170		93257	8488											41%
7	139094	34487		2213	5108		149267			16877								43%
8	64208	41298	4204		6995			44304	2913		4497							26%
9	462743	75690	341893	28788	39411		10597	58112	4021139		48098							79%
10			17011							50405								75%
11	89669	40985							15021	5800	2136491			9307	2107	86898		90%
12											467048	756733				14809		61%
13	6179	46065											1198759					96%
14														9587	8912			52%
15	4206	1501									9610				63698			81%
16				7985							14702	18785				293666		88%
17																	13401	100%
total	1709016	3458546	4105419	159874	126239	104559	199950	102417	4080172	73082	2696043	775518	1198759	18893	74717	395372	13401	
P accuracy	47%	92%	88%	44%	50%	89%	75%	43%	99%	69%	79%	98%	100%	51%	85%	74%	100%	
Overall accuracy	86%																	

Table 2 Error matrix for hybrid classification of land cover in Three Gorge area