ESTIMATION OF IMPERVIOUS SURFACE BASED ON INTEGRATED ANALYSIS OF CLASSIFICATION AND REGRESSION BY USING SVM.

Cheng Xi[1]* *Luo Jiancheng*[1] *Shen Zhanfeng*[1] *Zhu Changming*[1] *Zhang Xin*[1] *Xia Liegang*[1]

1:Institute of remote sensing applications, Chinese academy of sciences, Beijing, China

ABSTRACT

Impervious surface percentage(ISP) is the key parameter for urban regional environment research. This paper proposes the method of ISP estimation by using support vector machine(SVM) on TM image: (1) extract the ISA pixels which occupies any portion of the constructed impervious class based on SVM classification for spatial inputs of ISP estimation (2) estimate ISP of ISA pixels by using SVM regression model, build sample-ISP regression model based on various spectral features inputs and apply ISP-model for regional imperviousness mapping. On the TM image of Tianjin urban area, select high resolution image(Quickbird) classification result of college, industrial and residential districts as training sample(7500 items) and testing sample(2000 items), the mean square error(RMSE) of SVM model is 15.4%; adding "greenness" of tasseled cap transform as SVM feature, the RMSE decrease to 12%. The results of the study indicate that SVM model is suitable for large area ISP mapping without insufficient sample because of the non-linear characteristic and good performance of small-sample generalization. Additionally, to build a typical sample library for large-area ISP mapping will be our future research directions.

Index Terms— impervious surface area, impervious surface percentage , support vector machine,estimation

1. INTRODUCTION

Impervious surface percentage (ISP), has close relationship of regional hydrological environment and boundary layer climates(Schueler 1994, Weng et al. 2004),which is the key parameters of regional-urban environment research. Cause by complex heterogeneity of urban land surface, the spectrum of a image pixel may represent a combination of different components, especially for medium resolution images. Based on the vegetation–impervious surface–soil (V–I–S)model proposed by Ridd (1995), there are two kinds of main Methods of remote-sensing ISP assessment: spectral mixture analysis(SMA, Wu & Murray, 2003；Wu，2004；Lee & Lathrop, 2005； Lu & Weng , 2006; Pu, 2008) and machine study method(Yang et al. 2003; Bauer et al.2005;Mohaptra & Wu 2007; Hu & Weng,2009; Esch et al,2009).

In this paper, we propose a classification-regression methodology based on support vector machine(SVM) for regional ISP assessment using Landsat TM images. With training samples collected through classification result of high resolution image of different districts, firstly tackle the problem of binary impervious surface classification of assigning the pixels if any portion of a pixel occupies a constructed impervious surface to the impervious surface area (ISA) class; subsequently, through the SVM regression model establish the non-linear relationship between spectral feature of TM pixels and corresponding ground sample ISP value, and implied on result pixels of impervious surface classification above. In this context, our research is focused on:

(1) evaluating the classification accuracy of SVM for identifying urban surface components by binary impervious surface classification with a automatic method;

(2) evaluating the ability of SVM for assessment of ISP by using SMA;

(3) comparing the assessment accuracy of SVM model with different spectral features input.

2. STUDY AREA AND DATASETS

The study area is the region covered by a Landsat Thematic Mapper(TM+) scene (path122 row33) acquired on August 30th, 2000. It is located in Metropolitan district of Tianjin, China(Fig.1). The Landsat TM imagery with 30 m spatial resolution and 6 bands (blue, green, red, near IR and two mid IR bands) calibrated to reflectance is used as the data source in this research. Within the study area, QuickBird high spatial resolution images covered campus, residential area and factory area (shown Fig.1. a, b, c) were collected for reference data. QuickBird acquired on July, 2009, basically has time synchronization with the TM image, ensure the authenticity of the ground validation data. After the object-oriented classification and assisted visual interpretation, sample images were classified into impervious class (roads, building, parking lot et.al) and non-impervious class (vegetation, water and soil er.al) and divided into two independent ISP-sample set of training(7500 pixels) and testing(2000 pixels).

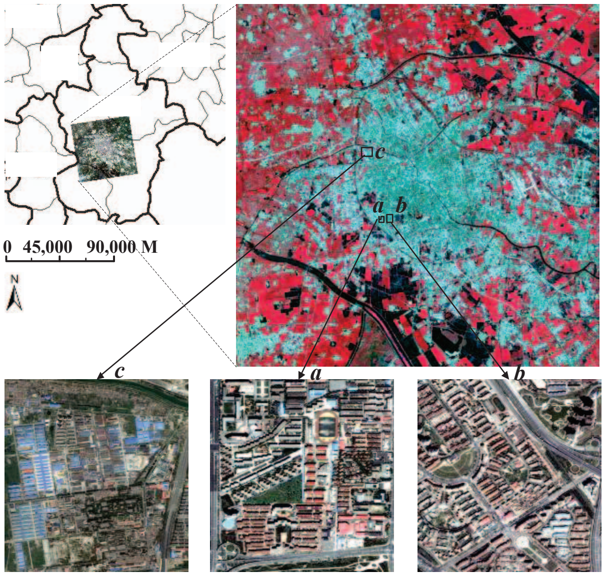

Fig.1 The location map of TianJing city, (left); the TM image, acquired on Aug.30, 2009, shows the study area, (right). and Quickbird image,shows the sample area, (bottom)

3. METHODOLOGY

3.1. Technical Route

As shown Fig.2, the general procedure of ISP information estimation can basically divided into two steps: (1) ISA extraction: extract the ISA pixels which occupies any portion of the constructed impervious class based on SVM classification (2) ISP estimation: estimate ISP of ISA pixels by using SVM regression model, build sample-ISP regression model based on various spectral features inputs and apply ISP-model for regional imperviousness mapping.

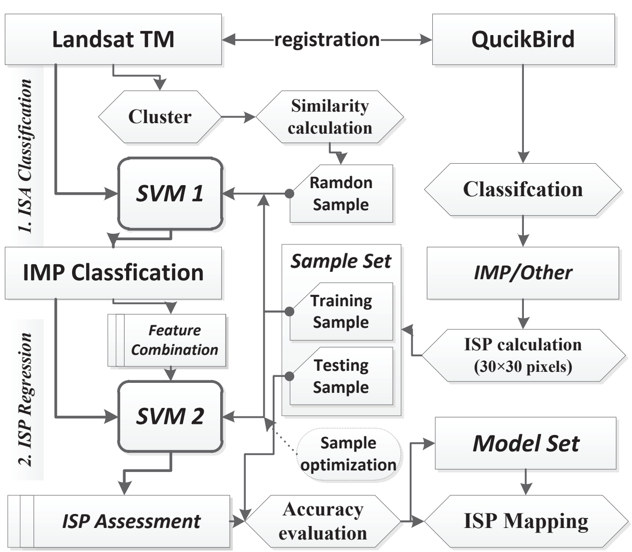

Fig.2 Technical route of ISA extraction and ISP estimation

3.2. Impervious Surface Area classification

We tackle the procession of binary impervious surface classification in first step. If any portion of a pixel occupies the constructed impervious class, that pixel is assigned to the impervious class. The motivation for a binary classification results from the imperviousness overestimation of large scale subpixel classifiers in rural areas (Homer et al., 2004; Yang et al., 2003), assigning imperviousness where no imperviousness exists. Therefore, our method does not directly compete with subpixel algorithms; instead, it complements them could succeeding perform subpixel analysis using results as a preprocessing binary filter. (Fig 2)

(1)do unsupervised classification(Isodata cluster) and get several undefined classes, ISA fields may be made up of mixtures of two or more clusters if clustering appropriately. (2)assign ISA classes and Other classes Through the spectral similarity matching index(Bhattacharyya Distance):

$$D_B = \frac{1}{8}\left(m_1 - m_2\right)^T P^{-1}\left(m_1 + m_2\right) + \frac{1}{2}\ln\left(\frac{\det P}{\sqrt{\det P_1 \det P_2}}\right)$$

Where, m_i and P_i are the mean vector and the covariance matrix of class one and two, the B-distance has widely implemented for related analysis of high-dimensional datasets; (3)generate random samples from ISA/Other classes and combined with training sample collected from High-resolution image; (4)do supervised classification using SVM with sample above to get ISA pixels; (5)dissolve the small pieces under custom minimum size in result map to refine the ISA. The ISA pixels will taken as the spatial input of ISP estimation.

3.3.Impervious Surface Percentage estimation

The ISP of pixel defined as the percentage of impervious surface in unit surface area. On the basis of previous research, the SMA and mathematical regression model both provides a characteristic ISP-estimation method. Although the lack of clear physical model derived, the mathematical regression model can acquire more accurate ISP rely on the advantages of the nonlinear fitting using machine learning algorithms. In this research, we apply SVM method to test the performance of ISP estimation.

The concept of Support Vector Machine(SVM) follows the basic idea of a supervised classification and originates from a non-parametric machine learning methodology based on the Structural Risk Minimization (SRM) principle. compared with other machine methods, SVM represent a method which facilitates comparatively accurate and fast processing – even when the number of training samples is limited – due to its good generalization ability, which is suitable for large area ISP estimation with difficult groud sample obtainment. For supervised classification SVM uses specific training and reference data in order to generate a model that estimates the targeted value of the reference based on the properties of the training data set. The background theory and detailed algorithm can be reference in relevant materials by Vapnik (1995, 1998).

Accurate training sample is the primary condition of estimation accuracy of SVM-ISP model，but the problem of image registration and building shadow sample during the sample collection process often cause errors affecting performance of SVM-ISP model. We adopt a training strategy of sample-optimization for SVM modeling(Fig.3)：Divide the ISP sample above into three groups according ISP value descending(1-0.8,0.79-0.5,0.5-0.1) and applied for SVM-ISP model training in turns, select samples with highest 90% training accuracy (RMSE) each time and combined with next group of sample set again for SVM training, avoiding the unnecessary error introduced by inaccurate samples.

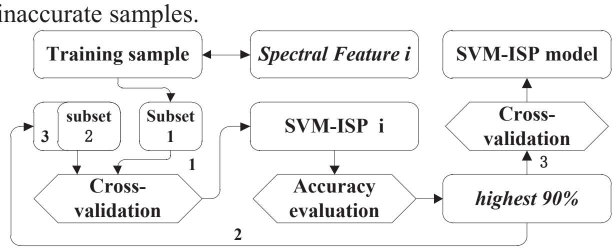

Fig.3 Training strategy of SVM-ISP model

4. RESULTS AND ANALYSIS

For the implementation of the training and modeling procedure we employed the libSVM2.9 (Chang and Lin, 2001). The classification model in this study select the C-SVC model and regression model select the epsilon SVR .The two model are with a radial basis function (RBF) as kernel type.

Considering the TM six band spectral reflectance as SVM-ISA model's feature inputs and add "greenness" fraction of TC-transform as SVM-ISP model feature inputs due to the well correlation with pixel ISP (Yang,2003). To get reliable and stable SVM model, the best SVM parameters of g and C are elected by 3-fold cross-validation(**Table.1**).

Table 1 The parameters of SVM models

SVM-Model	Inputs	parameters		training accuracy	
		g	*C*	MAE(%)	R^2
SVM-ISA	TM	0.01	10	/	/
SVM-ISP	TM	100	10	11.3	0.68
	TM+TC2	100	10	11.1	0.69

*TM：Spectral reflectance of TM image (except thermal infrared band);
TC2:"greeness" of TC transformation.

The results of both ISA classification and ISP assessment are shown in **Fig.5** (a.ISA; b:ISP);Because lack of large area ground data of ISA, we do not employ in quantitative evaluation of ISA classification. As shown in Fig.4, Compared with TM image and result of ISA classification has structured spatial distribution briefly and provides a reliable spatial information input for ISP estimation.

To verify the ISP estimation accuracy, using the above 2000 validation samples above for accuracy evaluation. The ISP value generated by SVM would potentially <0 and> 1, It is needed to designate as 0-1 to evaluation. We selected mean square deviation(MAE), RMS error(RMSE) and coefficient of determination (R^2) as the standards of fitting degree between estimated value and sample value. The accuracy evaluation is shown in **Table.2**. And Fig.4 presents scatter plots to show the agreement between ISP estimated value and sample value.

Table.3 Accuracy of ISP estimation

ISP models	Mapping accuracy		
	MAE(%)	RMSE(%)	R^2
TM	15.4	21.9	0.58
TM+TC2	12.2	17.9	0.65

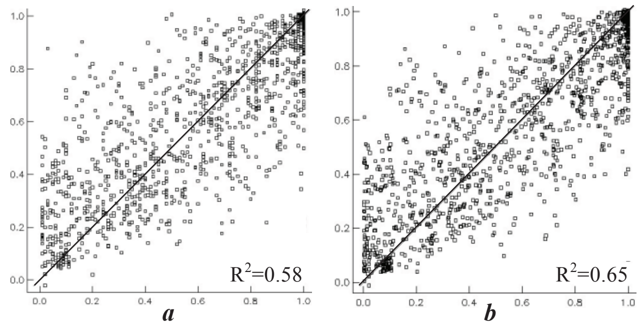

Fig.4 Correlation scatterplot chart of SVM-ISP model
(a:TM as input;b:TM+TC2 as input)

Finally, we adopt TM+TC2 model to calculate the ISP distribution in entire area of Tianjin(Fig.5), the result reasonably demonstrated the spatial distribution of ISP and the strength of construction from the visual effect, which show that the SVM model has certain potential application of the large regional ISP estimation.

5. CONCLUSION AND DISCUSSION

The experimental results show that: (1)SVM model is suitable for large area ISP mapping (of both classification and regression) without insufficient sample because of the non-linear characteristic and good performance of small-sample generalization on coarse resolution imagery (30 m Landsat ETM+ imagery); (2) By adding spectral feature vector(TC-"greenness") having significant relation with ISP can adjust the value of ISP assessment where the land cover types is lack of training sample and improve the overall accuracy of regional ISP assessment.

This study only considered the spectral features of pixel in ISA classification stage, but there is strong spectral confusion between some land cover (bail soil and high-refection build-up; building shadow and vegetation/water) on the optical remote sensing image, which still make high estimation errors of certain land covers. Further consideration of spectral association between adjacent pixels and introducing spatial statistical characteristics(Mountrakis，2011), will expected to be greatly improved accuracy of ISP estimation. The integration of multi-features and multi-classifier will be the development direction of ISA classification.(Luo & Mountrakis，2010).On the other hand, the cluster method to generate random samples using in this study is only available on the images with certain impervious surface, how to locate the ISA on images in an automatic thematic information extracting is still to be answered.

In addition, as the limitation of high resolution image access(Soe W.M,et al,2011), our study used relatively small set of ground samples. With adding sample types and introducing sampling technologies as Bootstrapp , there is still room to improve estimation accuracy. Additionally, to build a typical sample library with reliable sample matching algorithm for large-area ISP mapping(George X.2010) will be our future research directions.

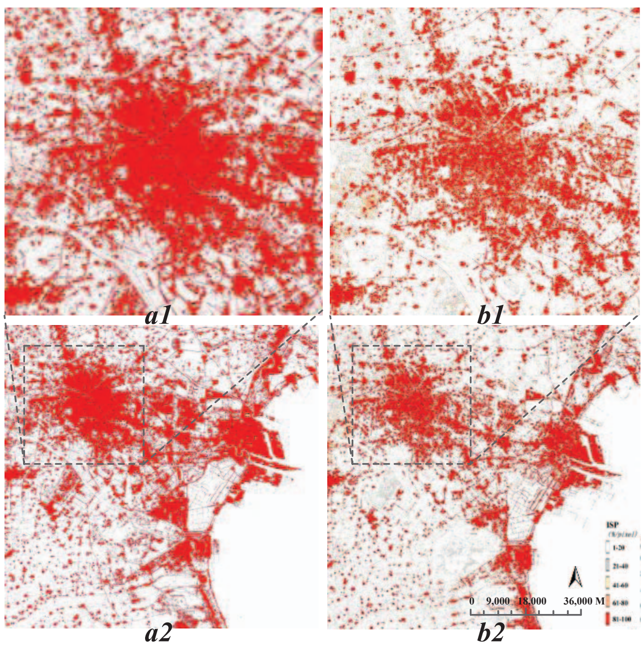

Fig.5 result images of impervious surface estimation. (a:ISA classification; b:ISP estimation)

6. REFERENCES

[1] Esch T, Himmler V,Schorcht G. 2009 .Large-area assessment of impervious surface based on integrated analysis of single-date Landsat-7 images and geospatial vector data. *Remote Sensing of Environment* ,**113**(8):1678-1690

[2] Hu X F,Wen Q H.2009.Estimating impervious surface from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural network. *Remote Sensing of Environment*,**113**():2089-1102

[3] Lee S, Lathrop R. 2005.Sub-pixel estimation of urban land cover components with linear mixture model analysis and Landsat Thematic Mapper imagery. *International Journal of Remote Sensing* , **26**(22):4885-4905.

[4] Lu D, Weng Q.2008. Use of impervious surface in urban land-use classification. *Remote Sensing of Environment*,**24**(6):1273-1286.

[5] Luo L , Mountrakis G.2010. Integrating intermediate inputs from partially classified images within a hybrid classification framework: An impervious surface estimation example. *Remote Sensing of Environment*,**114**(6):1220-1229

[6] George X & Collin H.2010. Updating the 2001 National Land Cover Database Impervious Surface Products to 2006 using Landsat Imagery Change Detection Methods. *Remote Sensing of Environment*,114():1676-1686.

[7] Giorgos Mountrakis, Li Luo.2011. Enhancing and replacing spectral information with intermediate structural inputs:A case study on impervious surface detection. *Remote Sensing of Environment,***115**():1162-1170

[8] Powell R L, Roberts D A, Dennison P E and Hess L L,2007, Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. *Remote Sensing of Environment*, **106**():253-267

[9] Pu R,Gong P,Ryo M and Todashi S.2008.Spectral mixture analysis for mapping aboundance of urban surface components from the Terra/ASTER data. *Remote Sensing of Environment* , **112**(3):939-954

[10] Soe W. M., Patricia G, Anthony B, Susanne G.C., and Weng Q.H. 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*.115():1145-1161.

[11] Vapnik V N. 1999 An Overview of StatisticalLearning Theory . *IEEE Transactions on Neural Networks*, **10**(5):988-999

[12] Wu C and Murray A T. 2003. Estimating impervious surface distri-bution by spectral mixture analysis. *Remote Sensing of Environment*, **84**(4): 493-505

[13] Wu C.2004. Normalized spectral mixture analysis for monitoring urban composition using ETM+ imagery. *Remote Sensing of Environment*, **93**(4): 480-492

[14] Yang L M, Huang C Q, Homer C G, Wylie B K and Coan M J. 2003. An approach for mapping large-area impervious surfaces-synergistic use of Landsat-7 ETM+ and high spatial resolution imagery. *Canadian Journal of Remote Sensing*, **29**(2): 230—240

[15] Yang X ,Liu Z.2005.Use of satellite-derived landscape imperviousness index to characterize urban spatial growth. *International Journal of Remote Sensing*,**29**(5):524-540.