The analysis of the classification of SPOT 5 image based on local spatial statistics

ZHU Zhongzheng, SU Wei

College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, *China*

Abstract: Influenced by the different spectrum with same object and the same spectrum with different object, when classifying remote sensing images, there will be a limitation on the improvement of the classification accuracy if merely relies on spectral signatures. However, this influence could be impaired by the characteristics of local spatial statistics, which are capable of describing and analyzing the spatial autocorrelation. In this paper, the effects which different Local Spatial Statistics indices (Moran's I, Getis-Ord G_i, Geary's C), neighborhood rule, and lags have on the classification accuracy of the high resolution remote sensing image: SPOT 5 are discussed. Firstly, band 1 is analyzed regarding to local spatial statistics, and the results are added to original spectral bands as textural ones; Secondly, supervised classification is carried out with both spectral band and textural band; Finally, test samples are selected to assess the classification accuracy and different accuracies on different conditions are juxtaposed with each other to figure out the relations and disciplines between the classification accuracy and the parameters. Results indicate that G_i statistics should be the optimum choice to raise the overall classification accuracy, with an increase from 87.74% to 95.12%.

Key words: local spatial statistics, neighborhood rule, lag, direction of spatial distribution

CLC number: P237 **Document code:** A

Citation format: Zhu Z Z and Su W. 2011. The analysis of the classification of SPOT 5 image based on local spatial statistics. *Journal of Remote Sensing*, **15**(5): 957–972

1 INTRODUCTION

In the analysis of remote sensing image, local spatial statistics plays a pushing role (Emerson , *et al*., 2005).Among the most popular local spatial statistics indices are Moran's I, Getis-Ord G_i and Geary's C. Moran's I represents the pixel clustering and a positive I means similar attribute value within nearby pixels, while a negative I means difference (Anselin, 1995). However, as an index to detect variability between spatial pixel and nearby ones, Geary's C can represent the edge between clusters and other areas with dissimilar neighboring values (Anselin, 1995). In addition, The Getis-Ord G_i index identifies hot spots, such as areas of very high or very low values that occur near one another. This is useful for determining clusters of similar values, where concentrations of high values result in a high G_i value and concentrations of low values result in a low G_i value (Getis & Ord, 1992).

Ever since the invention of this method, many researchers did a lot of experiments about the principle of it and the classification accuracy. G_i founder, Getis and Ord (1992), assert that G_i can provide researchers with a straightforward way to assess the degree of spatial association at various levels of spatial refinement in an entire sample or in relation to a single observation. Additionally, Anselin (1995) points out that Local Indicators of Spatial Association can detect local pockets of non-stationarity, or hot spots. At the same time, Emerson, *et al*. (2005) compare the effect of local variance, fractal dimension, and Moran's I, and asserted that bands which are analysed by Moran's I can be added to the stack of multi-spectral bands as texture images to lift the accuracy of classification. Furthermore, Wulder and Boots's (1998) research regarding the use of local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic also mentioned that Getis-Ord G_i can be used to assist classification.

Results mentioned above show that local spatial statistics indi-

Received: 2010-05-12; **Accepted:** 2011-03-28

Foundation: National Natural Science Foundation of China (No. 40801128)

First author biography: ZHU Zhongzheng(1989—), male, bachelor. His main research interests are remote sensing digital image processing and application. E-mail: zhucyclone09@gmail.com

Corresponding author biography: SU Wei (1979—), female, doctor. She has already published more than 14 academic papers. Her main research interests are remote sensing digital image processing and remote sensing applied in agriculture. E-mail: suwei7963@163.com

ces can improve classification accuracy. However, systematical and quantified analyses are never carried out. This paper is intended to solve this problem with experiments about different indices, neighborhood rule and lags, using he example of the high resolution remote sensing image: SPOT 5 of the agricultural transitional zone in Changping district.

2 METHODS

2.1 Study area and image data

The agricultural transitional zone in Changping district is chosen as the study area in this paper and the image which is used is the high resolution remote sensing image: SPOT 5. Image projection is UTM, the ellipsoid is Krasovsky and spatial resolution is 2.5 m. Pre-disposition includes geometric correction, atmospheric correction *etc*., Topographic map (1:10000) is used as the benchmark to correct the original sensing image and the calibration error is strictly controlled within one pixel. Atmospheric correction is carried out to eliminate the interfere of atmosphere. The resulting image is shown in Fig. 1 The analysis of local spatial statistics uses Local Spatial Statistics (ENVI), while supervised classification uses Maximum Likelihood. From Fig. 1, we can see that ,study area is mostly made up of croplands together with mixed forest, shadows, water bodies and urban and built up. Because of the same spectrum with different object between mixed forest and intensive croplands, and sparse croplands and urban and built up, merely using spectral signatures is difficult to make the classification. But luckily, characteristics of local spatial statistics can be the alternative.

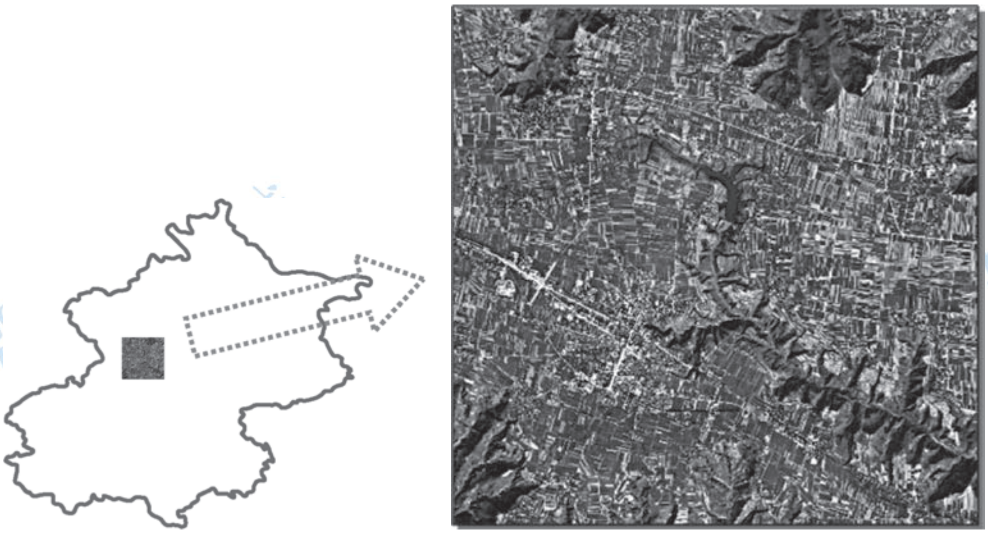

Fig. 1 SPOT 5 remote sensing image(2.5 m) of the study area

2.2 Parameter

2.2.1 *Local spatial statistics indices*

Moran's *I*: Moran's *I*, which is first put forward by Anselin (1995), is used to detect spatial objects clustering. The defined formula is

$$I(d)=\frac{\sum_i\sum_j c_{ij}\boldsymbol{w}_{ij}}{s^2\sum_i\sum_j \boldsymbol{w}_{ij}} \tag{1}$$

where s^2 means the variance of attribute values in spatial weights, d means the lag, $\boldsymbol{w}_{ij}$ means the spatial weights which is only composed of 1 and 0 to represent the impact between i and j. c_{ij} means the product of the difference between attribute value of spatial pixels, i, and average attribute value within the lag and the difference between attribute value of spatial pixels, j, and average attribute value within the lag.

Getis-Ord G_i: Getis-Ord G_i which is first put forward by Ord and Getis (1992)[2], is used to detect hot spots. The defined formula is:

$$G_i^*(d)=\frac{\sum_{j=1}^{n}\boldsymbol{w}_{ij}(d)x_j}{\sum_{j=1}^{n}x_j} \tag{2}$$

where $\boldsymbol{w}_{ij}$ means the spatial weights between i pixel and j pixel, d means the lag, x_j means attribute value of spatial pixel, j.

Geary's C: Geary's C, which is first put forward by Anselin (1995)[3], is used to detect the variability of surface features. It is useful for detecting edge areas between clusters and other areas with dissimilar neighboring values. The defined formula is

$$C_i(d)=\frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{w}_{ij}\left(x_i-x_j\right)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{w}_{ij}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2} \tag{3}$$

where the definition of $\boldsymbol{w}_{ij}$, d and x_j is the same in Getis-Ord G_i.

2.2.2 Neighborhood rule and lag

From ENVI User's Guide's Local Spatial Statistics, we can find out the explanation of neighborhood rule. It is used to define the neighboring pixels which are in comparison with the central one. Most common 7 rules are:

Rook's Case: Selects the pixels on the top, bottom, left, and right.

Bishop's Case: Selects four diagonal neighboring pixels.

Queen's Case: Selects all eight neighboring pixels.

Horizontal: Selects two neighboring pixels in the same row.

Vertical: Selects two neighboring pixels in the same column.

Positive Slope: Selects two neighboring pixels in opposite corners in a positive diagonal (45°).

Negative Slope: Selects two neighboring pixels in opposite corners in a negative diagonal (135°).

According to the number of directions, I defined multi-direction rule and mono-direction rule. The former includes Rook's Case, Bishop's Case and Queen's Case, which are all compared in more than two directions, while the latter includes Horizontal, Vertical, Positive Slope and Negative Slope, which are compared in only two directions.

The definition of lag is: define 1 in corresponding coordinate in spatial weights matrix, $\boldsymbol{w}_{ij}$, for all the pixel pairs in given distance and 0 for the rest. The neighboring pixels of 1 will affect the index value of center pixel. Therefore, with the variance of lag, local spatial statistics index/indices value of the image will change at the same time.

2.3 Introduction of experiments using local spatial statistics

In this research, the three local spatial statistics indices,7 rules and 7 lags from 1 to 7 are analyzed ,regarding to their effects on classification accuracy. Texture bands and image spectrum bands are added together to be used in later supervised classification to improve accuracy of information derived from study areas (Fig. 2).

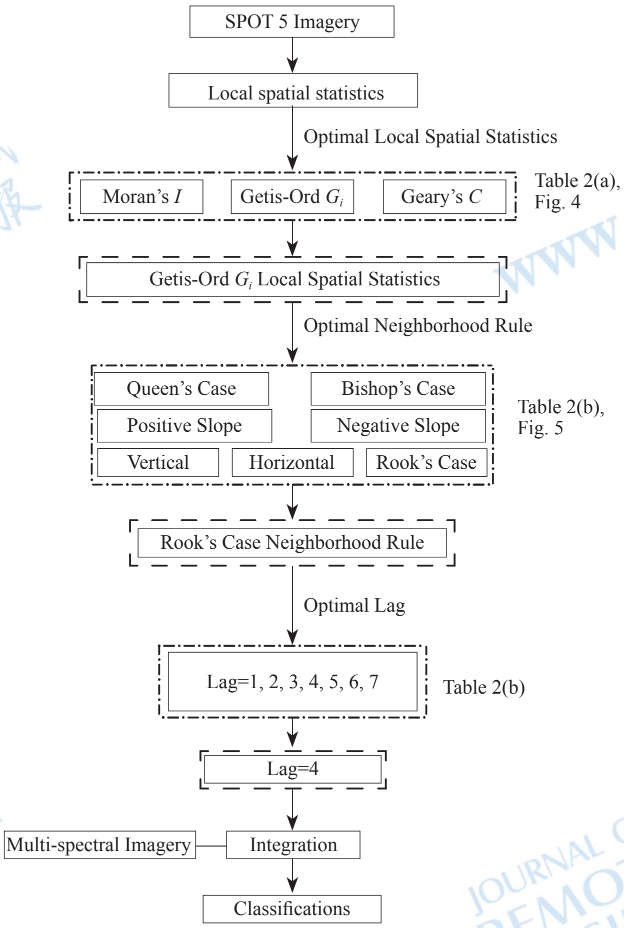

Fig. 2 Procedure of local spatial statistical analysis
(Dotted line —·— means comparison of classification accuracy under different parameters.
Dotted line — — means the best parameters of local spatial statistical analysis after evaluation.)

2.4 Image classification and accuracy assessment

Maximum likelihood classifier is used to carry out supervised classification and the probability threshold is 0.1 (the max is 1).The classification is according to IGBP-DIS classification system. Cropland is sorted into two kinds, which are sparse cropland and intensive cropland (Table 1).The image interpretation signs are shown in Fig. 3: blue part in Fig. 3 (a) is water bodies, green part in Fig. 3 (b) is intensive cropland, white part or net-like part in Fig. 3 (c) is urban and built up, white rectangles in Fig. 3 (d) is sparse cropland, green part with jagged edge in Fig. 3 (e) is mixed forest, black part in Fig. 3 (f) is shadow. The selection principle of training samples is: Uniform, randomly distributed in the whole panoramic image, in order to reduce subjective factors on classification accuracy. The quantity of training samples are: water bodies (1007), intensive cropland (6179), urban and built up (3393), sparse cropland (1046), mixed forest (1738) and shadow(1022). The selection principle of test samples is: random and not repeat. The quantity of test samples are: water bodies (1005), intensive cropland (2028), urban and built up (2223), sparse cropland (1007), mixed forest (1718) and shadow (2064). Confusion matrix is used in classification accuracy assessment.

Table 1 Land cover classification categories

Land cover	Land use classes
Water bodies	Rivers, lakes, reservoirs, bays and estuaries etc
Intensive cropland	Cultivated a following segment, crop more densely
Urban and built up	Residential, commercial and service, industrial, transportation and sport facilities etc
Sparse cropland	Cultivated a following segment, sparse crops
Mixed forest	Mixed forest land
Shadow	Interrupted data

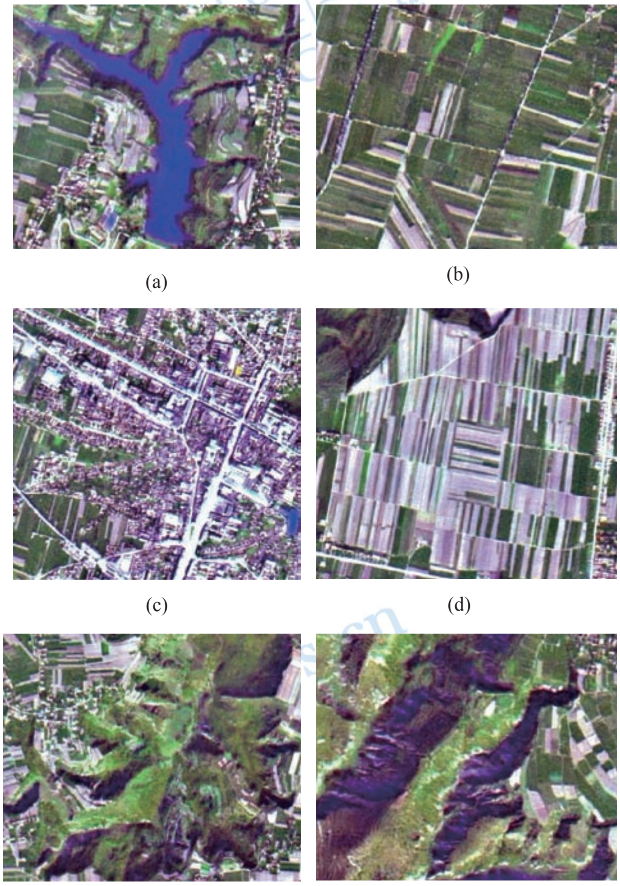
(a) (b) (c) (d) (e) (f)

Fig. 3 Classification reference
(a) Water bodies; (b) Intensive cropland; (c) Urban and built up; (d) Sparse cropland; (e) Mixed forest; (f) Shadow

3 RESULTS AND DISCUSSION

3.1 Selection of local spatial statistics indices

In order to assess the effects different indices have on classification accuracy, analysis and calculation of SPOT 5 remote sensing, band 1, are carried out under Rook's case and the lag of 1. In addition, classification accuracy all refers to production accuracy. The calculation results are shown in Fig. 4. The result of Moran's I shows that high-spatial-autocorrelation urban and built up have a high I, while low-spatial-autocorrelation mixed forest have a low one. The rest four are ordinary. Compared with Moran's I, Getis-Ord G_i can distinguish concentrations of high values of urban and built up and some sparse cropland, while water bodies, shadow, some mixed forest and most sparse cropland are concentrations of low values. Getis-Ord G_i can hardly judge whether medium values of some mixed forest cluster or not(Wulder & Boots, 1998). Geary's C mainly shows the edge of urban and built up while Cs of other objects are low.

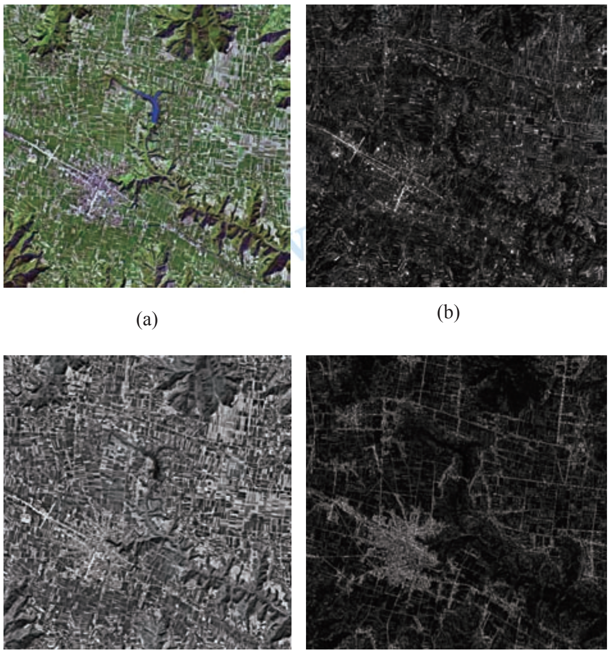
(a) (b) (c) (d)

Fig. 4 Local spatial statistics image from SPOT 5 imagery
(a) Original image; (b) Moran's I; (c) Getis-Ord G_i; (d) Geary's C

From Table 3 we can see that, of the three indices, Moran's I is only useful in intensive cropland. From Table 2 (a), Moran's I has the lowest overall classification accuracy except for horizontal. Therefore, Moran's I is the worst of the three. From Table 2 (a) we can also see that Getis-Ord G_i has the most apparent rise in overall accuracy and is most useful for mixed forest, 83.24%. This is because Getis-Ord G_i can detect the range of accumulation area precisely. (Zhang & Zhang, 2007). For the rest two indices, Geary's C, as an index to detect variability, is most useful for urban and built up, 97.03%. Ann it is proper for intensive cropland, which is most likely to be confused with urban and built up in image of only spectral signatures, with a percentage of 98.31%.

Table 2 Classification accuracies in different indices, neighborhood rules and lags

		Rook's case		Bishop's case		Queen's case		Horizontal		Vertical		Positive slope		Negative slope	
		Total accuracy/%	Kappa	Total accuracy/%	Kappa	Total accuracy/%	Kappa	Total accuracy/%	Kappa	Total accuracy/%	Kappa	Total accuracy/%	Kappa	Total accuracy/%	Kappa
(a)Three different indices (Lag =1)	Moran's I	92.79	0.9120	93.16	0.9164	93.04	0.9150	92.83	0.9124	92.99	0.9144	93.32	0.9184	93.12	0.9159
	Getis-Ord G_i	94.16	0.9285	94.39	0.9313	94.32	0.9305	93.77	0.9238	93.67	0.9226	94.21	0.9292	94.05	0.9272
	Geary's C	93.55	0.9212	93.90	0.9254	93.92	0.9256	92.54	0.9090	93.11	0.9159	93.34	0.9186	93.29	0.9180
(b)Seven different lags in Getis-Ord Gi	1	94.16	0.9285	94.39	0.9313	94.32	0.9305	93.77	0.9238	93.67	0.9226	94.21	0.9292	94.05	0.9272
	2	94.49	0.9327	94.67	0.9349	94.70	0.9352	94.14	0.9283	94.23	0.9294	94.58	0.9338	94.60	0.9340
	3	95.02	0.9391	95.00	0.9389	94.93	0.9380	94.35	0.9309	94.56	0.9336	94.62	0.9343	94.59	0.9339
	4	95.12	0.9404	94.70	0.9352	94.77	0.9361	94.25	0.9296	94.48	0.9326	94.39	0.9313	94.30	0.9302
	5	94.85	0.9371	94.20	0.9290	94.50	0.9328	94.17	0.9287	94.24	0.9295	94.14	0.9283	93.88	0.9251
	6	94.48	0.9326	93.90	0.9254	94.31	0.9304	94.20	0.9290	93.92	0.9256	93.84	0.9247	93.66	0.9224
	7	94.21	0.9292	93.74	0.9235	94.07	0.9275	94.08	0.9276	93.68	0.9227	93.36	0.9188	93.43	0.9196

Table 3 The assessment of supervised classification of different land covers under four different methods

	Only multi-spectral image		Multi-spectral +Moran's I		Multi-spectral +Getis-Ord G_i		Multi-spectral +Geary's C	
	Production/%	User/%	Production/%	User/%	Production/%	User/%	Production/%	User/%
Water bodies	83.28	95.77	93.03	93.78	93.73	96.81	93.03	95.80
Intensive cropland	95.86	89.13	97.53	86.75	96.50	91.24	96.40	87.94
Urban and built up	80.70	94.97	95.01	93.62	96.81	91.97	97.03	93.78
Sparse cropland	92.35	91.18	96.23	89.56	94.54	93.89	98.31	94.56
Shadow	96.85	98.38	97.00	98.09	98.11	98.44	97.48	98.15
Mixed forest	76.25	94.24	77.12	96.36	83.24	95.08	78.46	94.73

To sum up, Getis-Ord G_i has the best overall accuracy, Geary's C is suitable for objects of high variability, while Moran's I is the least useful one. In order to have a more accurate research of the other two parameter, Getis-Ord G_i is chosen as the only index.

3.2 Selection of neighborhood rules

After analysis, we can conclude that the classification accuracy is dependent on the similarity between the direction of comparison of pixel and the direction of distribution of surface features. Table 2 (b) shows the effect that different neighborhood and lag have on classification accuracy under the same condition of Getis-Ord G_i.

When it comes to sparse cropland, its distribution direction has a trend from north to south(Fig. 3 (d) and Fig.4 (a)).Furthermore, Fig. 5 (d) shows that its best classification accuracy is under the neighborhood of vertical. At the same time, the best accuracy of urban and built up is under the neighborhood of positive slope(Fig. 5 (c)).Therefore, we can conclude that if the chosen neighborhood is similar with the distribution direction of the objects, the classification accuracy will be higher. In addition, from Fig. 5 (c) (d) we can also see that Rook's Case is useful to classify sparse cropland and urban and built up. Thus, if the direction of comparison of pixel is similar with the distribution direction of the objects, the accuracy will also be raised.

From Fig. 5 (a) (b), we can see that ,as the object whose distribution of direction is not obvious, water bodies and intensive cropland, their classification accuracy is the highest when using Rook's Case. However, the unit areas of intensive cropland are very tiny and do not have a consistent distribution direction, so intensive cropland can only have a good accuracy under multi-direction neighborhood: Rook's Case, Bishop's Case and Queen's Case.

Although shadow is sparse, it is mainly because mountain shelter. So, its unit areas are large and have ideal spatial self-autocorrelation. From Fig. 5 (e) we can see that shadow can be easily classified except the neighborhood of vertical. Though the distribution of mixed forest and shadow is similar, the former has similar spectral signatures and characteristics of local spatial statistics with intensive cropland, so the overall accuracy is low. But Fig. 5 (e) also shows that, for large-area autocorrelation object, multi-direction neighborhood is useful. For example, Rook's Case in Fig. 5 (e).

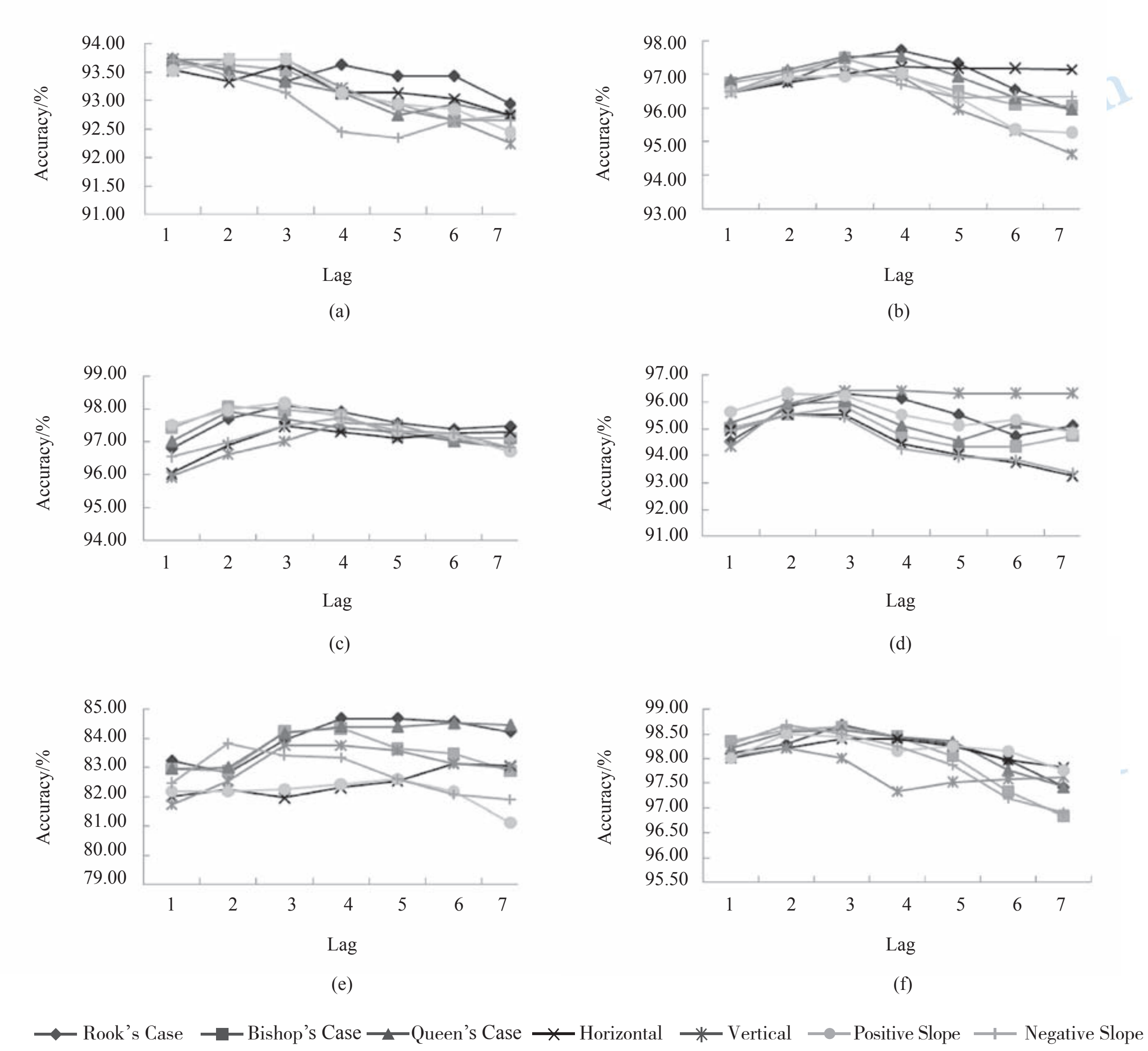

Fig. 5 The line chart of classification accuracy of six land covers under different lags and neighborhood rules
(a) Water bodies; (b) Intensive cropland; (c) Urban and built up; (d) Sparse cropland; (e) Mixed forest; (f) Shadow

In all, the impact that neighborhood rule has on classification accuracy is actually because similar neighborhood rule with the objects can reveal the spatial autocorrelation more accurately, while at the same time weaken the impact that other objects have on spatial autocorrelation. In addition, due to the more accurate spatial autocorrelation, textural band can classify different objects more accurately. Thus, the impact of the different spectrum with same object and the same spectrum with different object have on classification can be avoided or weakened.

3.3 Selection of lags

In order to analyze the relationship between lags and classification accuracy, we choose the average classification accuracy under different lags as reference, and it is shown in Table 4.From Table 4 we can see that the classification accuracy of large and high-spatial-self-autocorrelation object will be the highest under the lag of 1; However, the accuracy of small and low spatial-self-autocorrelation object will be highest if the lag is larger than 1. For example, the accuracy of sparse cropland reached 95.97% when the lag is 3 and that of mixed forest reached 83.61% when the lag is 4,which are both the optimal accuracy. In addition, the trend of accuracy with the growth of lag is: increase before the max and decrease after the max. Tiny instability may exist but that won't transcend the max number.

Table 4 Average classification accuracy of land cover under different lags

Accuracy/%	1	2	3	4	5	6	7
Water bodies	93.66	93.56	93.49	93.12	92.91	92.88	92.64
Intensive cropland	96.57	96.94	97.25	97.17	96.65	96.17	95.91
Sparse cropland	94.95	95.80	95.97	95.23	94.84	94.79	94.65
Urban and built up	96.76	97.46	97.71	97.65	97.34	97.19	97.02
Shadow	98.15	98.44	98.47	98.22	98.09	97.71	97.40
Mixed forest	82.51	82.79	83.39	83.61	83.44	83.30	82.95

3.4 Overall effects of local spatial statistics on classification accuracy

Compared with classification with only multi-spectral image, adding textural band created by local spatial statistics can impair the effect caused by the different spectrum with same object and same spectrum with different object. For example, in Fig. 6 (a) (b), mixed forest and intensive cropland, which are classified into "unclassified" because of the difficulty to be distinguished at first can now be sorted. Similarity, sparse cropland and urban and built up, as shown in Fig. 6 (c) (d),can also be better classified.

In addition, because of the effect brought by local spatial statistics, variance between different indices or parameters is not that obvious. But, it is much higher than the classification use only multi-spectral image. The max accuracy is 95.12%, which is 7.37% higher than 87.75%, which is the result of only multi-spectral image (Fig. 7).

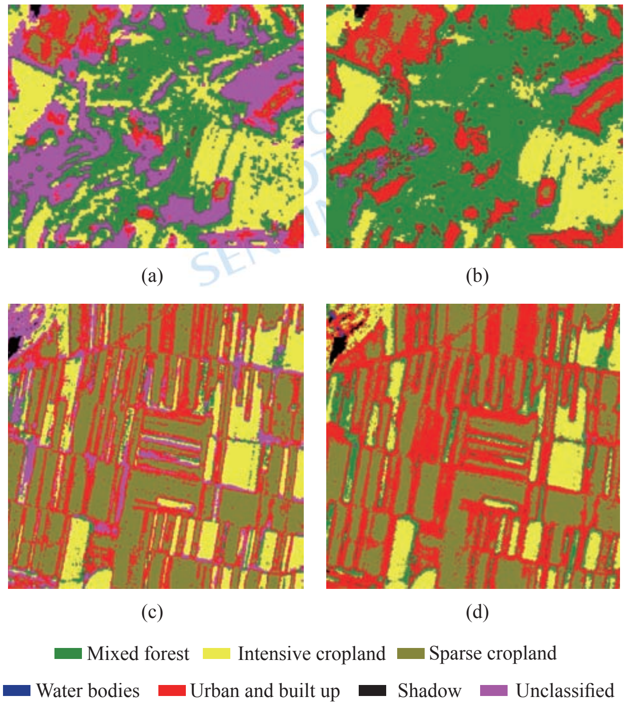

Fig. 6 The results of supervised classification (Local): Mixed forest and Intensive cropland
(a) Only multi-spectral image and (b) multi-spectral + Getis-Ord *Gi* (Rook's case, Lag = 4); Sparse cropland and Urban and built up: (c) Only multi-spectral image and (d) Multi-spectral + Getis-Ord *Gi* (Rook's case, Lag = 4)

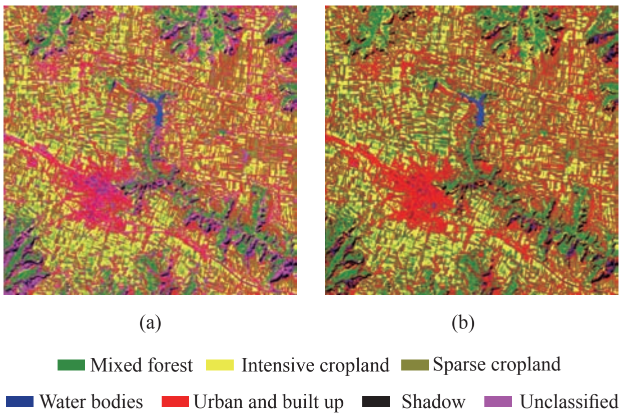

Fig. 7 The results of supervised classification (Overall):
(a) Only multi-spectral image and (b) Multi-spectral + Getis-Ord G_i (Rook's case, Lag = 4)

4 CONCLUSIONS

This paper mainly discusses the principle of impact on classification accuracy caused by indices, neighborhood and lags in local spatial statistics. The conclusions are:

(1) Getis-Ord G_i is most useful to classify objects in study area under any neighborhood and lag, which is the better than the other two indices.

(2) Neighborhood has some to do with the distribution direction of objects, but this spatial autocorrelation do not have apparent impact on sensing image. However, lags has more influence on classi-

fication accuracy than neighborhood. For example, under different lags, the average accuracy of intensive cropland varies a lot.

(3)Local spatial statistics can reduce the effects caused by he different spectrum with same object and same spectrum with different object. Therefore, using the textual band created by spectral band and local spatial statistics can improve the classification accuracy. The accuracy of the latter reached 95.12%, 7.37% higher than the former.

In addition, considering object-oriented classification has some influence on local spatial statistics, future study about it will be carried out.

REFERENCES

Anselin L. 1995. Local indicators of spatial association. *Geographical Analysis*, **27**(2): 93–115

Emerson C W, Lam N S N and Quattroch D A. 2005. A comparison of local variance, fractal dimension, and Moran's *I* as aids to multispectral image classification. *International Journal of Remote Sensing*, **26**(8): 1575–1588 DOI: 10.1080/01431160512331326765

Getis A and Ord J K. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, **24**(3): 189–206 DOI: 10.1111/j.1538-4632.1992.tb00261.x

Local Spatial Statistics. 2005. ENVI User's Guide. USA

Wulder M and Boots B. 1998. Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic. *International Journal of Remote Sensing*, **19**(11): 2223–2231 DOI: 10.1080/014311698214983

Zhang S L and Zhang K. 2007. Contrast study on the local indices of spatial autocorrelation. *Statistical Research*, **24**(7): 65–67

基于局部空间统计分析的SPOT 5影像分类

朱钟正，苏伟

中国农业大学 信息与电气工程学院，北京 100083

摘 要：受同物异谱和异物同谱现象的影响，对遥感影像进行分类时若仅利用光谱信息则分类精度的提高将会受到限制，而局部空间统计特征可以通过对地物空间聚集度的描述与分析在一定程度上减轻这种影响。本文研究了局部空间统计在不同指数(Moran's I, Getis-Ord G_i, Geary's C)、邻域规则和间隔距离下，对高空间分辨率的SPOT 5影像分类精度的影响规律。首先，对波段1进行局部空间统计分析，运算结果作为纹理波段添加到原始的光谱波段中；然后，综合利用光谱波段和纹理波段进行监督分类；最后，选取测试样本进行分类的精度评价，并比较分析不同条件下的分类精度，得到地物分类精度同参数之间的关系与规律。通过分析可以得出Getis-Ord G_i指数对于总体分类精度的提高最理想，总体分类精度从 87.74%提高到95.12%。

关键词：局部空间统计分析，邻域规则，间隔距离，空间分布方向

中图分类号：P237 **文献标志码**：A

引用格式：朱钟正，苏伟．2011．基于局部空间统计分析的SPOT 5影像分类．遥感学报，**15**(5): 957–972
Zhu Z Z and Su W. 2011. The analysis of the classification of SPOT 5 image based on local spatial statistics. *Journal of Remote Sensing*, **15**(5): 957–972

1 引 言

在遥感影像分析与解译中，局部空间统计分析对影像地物分类精度的提高起着重要作用(Emerson 等, 2005)，常用的局部空间统计分析指数有Moran's I, Getis-Ord G_i和Geary's C。其中Moran's I指数体现某一现象在空间上的聚集状态，正的I值表示该空间单元与邻近单元的属性值相似，负的I值表示属性值不相似(Anselin, 1995)；而作为检测空间单元与邻近单元之间变异性的指数，Geary's C指数能够检测出聚集区域与相邻但不相似区域的边缘(Anselin, 1995)；另外，Getis-Ord G_i指数能探测出高值聚集和低值聚集，显著的正G_i表示高值聚集，显著的负G_i表示低值聚集(Getis和Ord，1992)。

自从该方法被提出来后，许多研究者对该算法所涉及的原理及各参数对地物分类精度的影响进行了大量的研究与试验。G_i指数提出者Getis和Ord (1992)在使用距离统计指数分析空间关联的研究中提出了在完全的实例或单一实验现象的各种空间细化等级下，G_i指数能提供给研究者一种简单明了地评估空间相关程度的方法。而Anselin (1995)也在研究中指出了局部空间关联指数能检测局部空间的热点或非平稳性。同时，Emerson等人 (2005)比较分析了局部方差、分型维数和Moran's I 指数对多光谱影像分类的效果，并指出经过Moran's I 指数分析的波段可以作为纹理特征与光谱波段叠加，从而提高分类精度。另外，Wulder和Boots (1998)使用 Getis 统计指数评价遥感影像的局部空间自相关特征的研究中也提到Getis-Ord G_i指数可用于辅助分类。

上述研究结果表明局部空间统计指数能够提高分类精度，但并没有就指出局部空间统计指数、邻域规则和间隔距离的不同对分类精度的影响进行系统、定

收稿日期：2010-05-12；**修订日期**：2011-03-28

基金项目：国家自然科学基金(编号: 40801128)

第一作者简介：朱钟正(1989—)，男，理学学士，主要从事遥感数字影像处理及应用方面的研究。E-mail: zhucyclone09@gmail.com。

通信作者简介：苏伟(1979—)，女，博士，主要从事遥感数字影像处理及遥感农业应用方面的研究，已发表论文14篇，其中SCI检索4篇，EI检索4篇。E-mail: suwei7963@163.com。

量的比较分析。本文基于这个问题，以北京市昌平区农林交错带试验区的高空间分辨率SPOT 5遥感影像分类为例，通过局部空间统计分析多个指数、邻域规则与间隔距离的多组试验，考察局部空间统计分析方法各参数与影像分类精度之间的相互关系，从而为高空间分辨率遥感影像分类精度的提高提供新的途径与方法。

2　研究方法

2.1　研究区概况与数据源

试验区为北京市昌平区农林交错带，试验区影像为高空间分辨率SPOT 5影像，影像投影方式是横轴墨卡托，椭球体为克拉索夫斯基椭球，空间分辨率为2.5 m。影像的预处理工作包括几何校正、大气校正等，以试验区1∶10000地形图为基准校正原始遥感影像，校正误差严格控制在一个像元以内；为消除大气对影像分类的影响，进行大气校正，影像预处理后的结果如图1所示。局部空间统计分析使用的是ENVI的Local Spatial Statistics工具，监督分类则是用ENVI的Maximum Likelihood工具。从图1可以看出，试验区内主要以耕地为主，同时还有林地、阴影、水体和不透水层。因为林地和耕地、耕地和不透水层存在异物同谱的关系，若仅仅使用光谱信息进行分类难以将他们区分开来，幸运的是他们的局部空间统计特征可以用来解决这个问题。

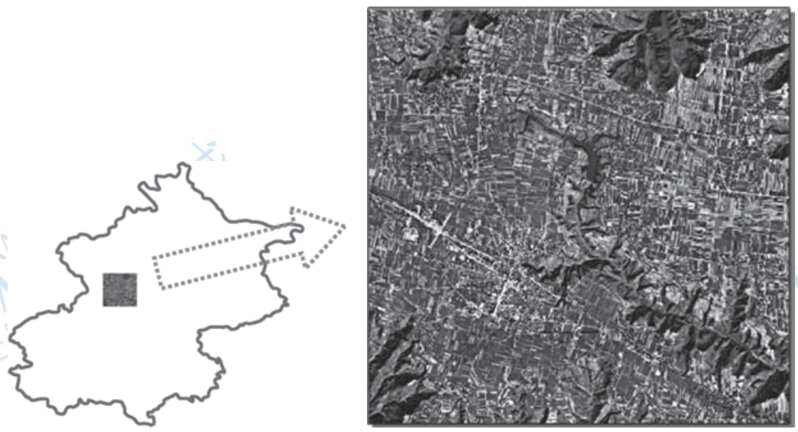

图1　研究区域的SPOT 5遥感影像(分辨率为2.5 m)

2.2　相关参数

2.2.1　局部空间统计指数

Moran's I 指数：主要用于探测空间地物聚集状况，由Anselin(1995)提出，其定义公式为：

$$I(d)=\frac{\sum_{i}\sum_{j}c_{ij}\boldsymbol{w}_{ij}}{s^2\sum_{i}\sum_{j}\boldsymbol{w}_{ij}} \tag{1}$$

式中，s^2为空间权矩阵内各个属性值的方差，d为间隔距离，$\boldsymbol{w}_{ij}$为空间权矩阵，其中只有1和0两种值，代表着空间单元i和j之间的影响程度，c_{ij}为空间单元i和j的属性值与间隔范围内平均属性值之差的乘积。

Getis-Ord G_i指数：主要用于探测热点，由Ord和Getis (1992)提出，计算公式为：

$$G_i^*(d)=\frac{\sum_{j=1}^{n}\boldsymbol{w}_{ij}(d)x_j}{\sum_{j=1}^{n}x_j} \tag{2}$$

式中，$\boldsymbol{w}_{ij}$是i、j单元之间的空间权矩阵，d为间隔距离，x_j为空间单元j的属性值。

Geary's C指数：由Anselin (1995)提出，主要用于探测地物的变异性，能够很好的探测出聚集区域的边界，其定义为：

$$C_i(d)=\frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{w}_{ij}\left(x_i-x_j\right)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{w}_{ij}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2} \tag{3}$$

式中，$\boldsymbol{w}_{ij}$、d和x_j的定义与Getis-Ord G_i指数中的相同。

2.2.2　邻域规则与间隔距离

从ENVI User's Guide的Local Spatial Statistics中关于邻域规则的相关解释中可以得知，邻域规则是为了确定与中心像元进行比较的邻域像元，常用的总共有7种，分别是：

(1)Rook's Case：选择上、下、左、右4个方向的

像元进行比较；

(2)Bishop’s Case：选择两个对角线方向的像元进行比较；

(3)Queen’s Case：选择全部8个方向的像元进行比较；

(4)Horizontal：选择同一行的像元进行比较；

(5)Vertical：选择同一列的像元进行比较；

(6)Positive Slope：选择45° 角的像元进行比较；

(7)Negative Slope：选择135° 角的像元进行比较。

依照选择方向的数量，将邻域规则分成了多方向邻域规则和单方向邻域规则，多方向邻域的解释为：中心像元比较像元的方向超过两个方向，包括Rook’s Case、Bishop’s Case和Queen’s Case；而单方向邻域的解释为：只有两个像元比较方向的邻域规则，包括Horizontal、Vertical、Positive Slope和Negative Slope。

间隔距离的定义是对于给定距离内的所有像元对都在空间权矩阵$\boldsymbol{w}_{ij}$对应的坐标中赋值1，否则是0。赋值1的邻域像元会影响到中心像元的局部空间统计指数值，赋值0的则相反，因此随着间隔距离的变化，影像的局部空间统计指数值也会改变。

2.3 局部空间统计分析实验简介

在此次研究中，3种局部空间统计指数、7种邻域规则和1到7的7种间隔距离被用来研究他们对遥感影像分类精度影响。局部空间统计分析产生的纹理波段与影像光谱波段进行叠加，共同参与后续的监督分类，以提高试验区地物信息提取的精度。实验主要流程如图2。

2.4 影像分类和精度评价

采用的监督分类方法是最大似然法，相似度阈值是0.6(最大值为1)。此次分类是依照IGBP-DIS分类体系进行确定的，其中耕地被细分为密集的耕地和稀疏的耕地，系统具体分类情况见表1。各地物的影像解译标志如图3所示：图3(a)中蓝色部分为水体，图3(b)中绿色部分为密集的耕地，图3(c)中白色且单位面积较小或呈网状分布的为不透水层，图3(d)中白色且呈矩形的是稀疏的耕地，图3(e)中绿色且边缘不整齐的是林地，图3(f)中黑色部分为阴影。训练样本的选择原则为：均匀、随机地分布于整景影像，以减少主观因素对分类精度的影响。影像分类中各地类的训练样本数量为：水体(1007像元)、密集的耕地(6179像元)、不透水层(3393像元)、稀疏的耕地(1046像元)、林地(1783像元) 和阴影(1022像元)。测试样本的选择原则是随机、不重复，具体选择数量为：水体(1005像元)、密集的耕地(2028像元)、不透水层(2223像元)、稀疏的耕地(1007像元)、林地(1718像元) 和阴影(2064像元)，采用混淆矩阵进行分类精度评价。

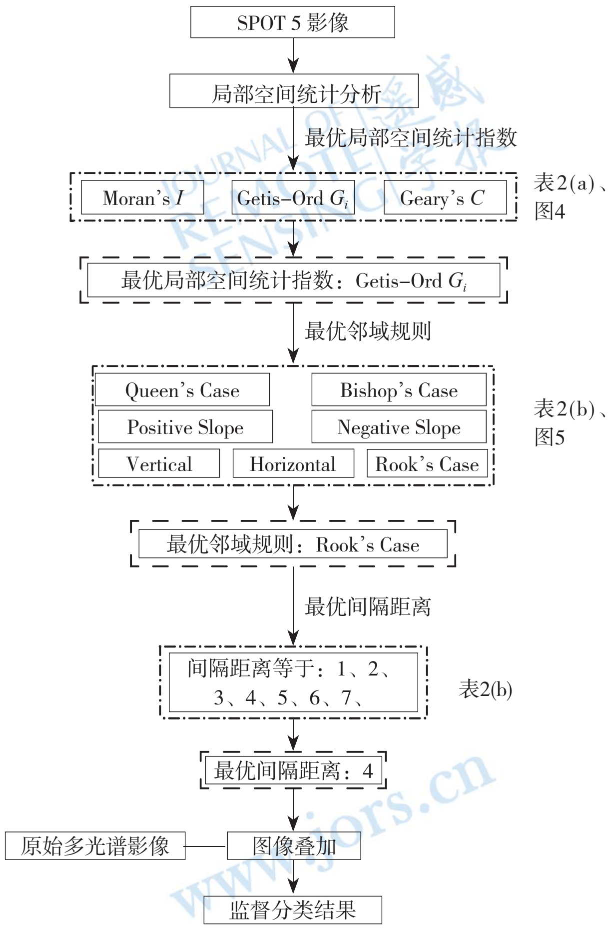

图2 局部空间统计分析流程

(-·-在不同分类参数下分类精度的对比；— —经分类精度评价得到最适宜的局部空间统计分析参数)

表1 土地覆被分类类别

土地覆被类型	土地使用分类
水体	河流、湖泊、水库、海湾和入海口等
密集的耕地	耕地细分后的一种，农作物比较密集
不透水层	包括居住地、商业用地、生产用地、交通和体育场所用地
稀疏的耕地	耕地细分后的一种，农作物比较稀疏
林地	此影像中主要为混交林
阴影	因遮挡而无法确定的土地覆盖类型

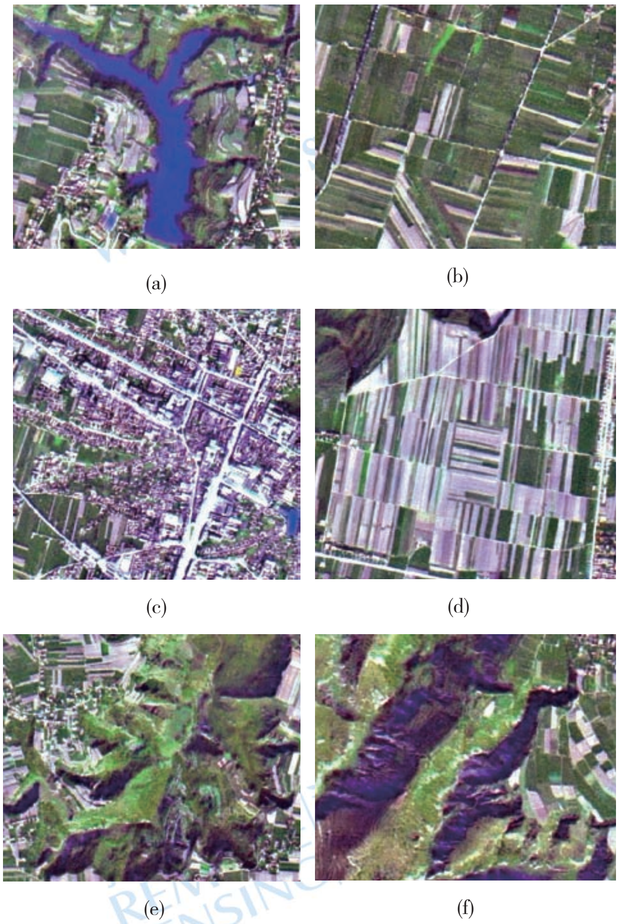

图3 地物分类参照图
(a) 水体；(b) 密集的耕地；(c) 不透水层；(d) 稀疏的耕地；
(e) 林地；(f) 阴影

3 结果和讨论

3.1 局部空间统计分析指数的选择

为了考察不同指数对分类精度的影响，在相同的领域规则(Rook's case)且间隔距离为1的情况下对SPOT 5遥感影像的波段1进行不同指数的分析和计算。为了便于分析，下文中的地物分类精度皆为生产精度。3种指数的计算结果如图4所示。Moran's I (图4(b))指数计算结果表明：空间聚集度高的不透水层地物I值较高，而空间聚集度低的林地I值则很低，其他4种地物的空间聚集度表现皆为一般。与Moran's I影像相比，Getis-Ord G_i影像(图4(c))可以分辨出不透水层和部分稀疏的耕地属于显著的高值聚集，而水体、阴影、部分林地和大部分密集的耕地则属于低值聚集，但是Getis-Ord G_i指数难以判断中值的部分林地是否聚集(Wulder和Boots, 1998)。与上面2种分析地物相似性的指数相比，Geary's C影像(图4(d))主要勾画出了不透水层的边缘，其他地物的C值则都比较低。

从表3得出，在3种指数中，Moran's I 指数只有密集的耕地分类精度最高。同时，从表2(a)得出，Moran's I 指数除了在Horizontal邻域规则下，总体分类精度均为最低。由此可知，Moran's I 指数是3种指数中分类效果最差的。从表2的(a)部分中得出，无论在任何邻域规则下，Getis-Ord G_i指数对总体分类精度的提高最显著。而相对于地物来说，Getis-Ord G_i指数对林地的分类精度最理想(表3)，为83.24%，这与Getis-Ord G_i指数能够较准确探测聚集区域的范围有关(张松林和张昆，2007)。对于其他两种指数，Geary's C作为探测地物变异性的指数，对不透水层有着最好的分类精度，为97.03%。并且它对于在只有光谱信息影像下与不透水层发生混淆的地物—密集的耕地，分类精度也是最好的，为98.31%。

综上所述，总体分类精度最理想的是Getis-Ord G_i指数，Geary's C指数对变异性高的地物分类精度较好，而Moran's I 指数是3种指数下效果最差的。为了准确地研究其他2种参数对地物分类精度的影响，下面邻域规则和间隔距离的选择和确定仅在Getis-Ord G_i指数下进行。

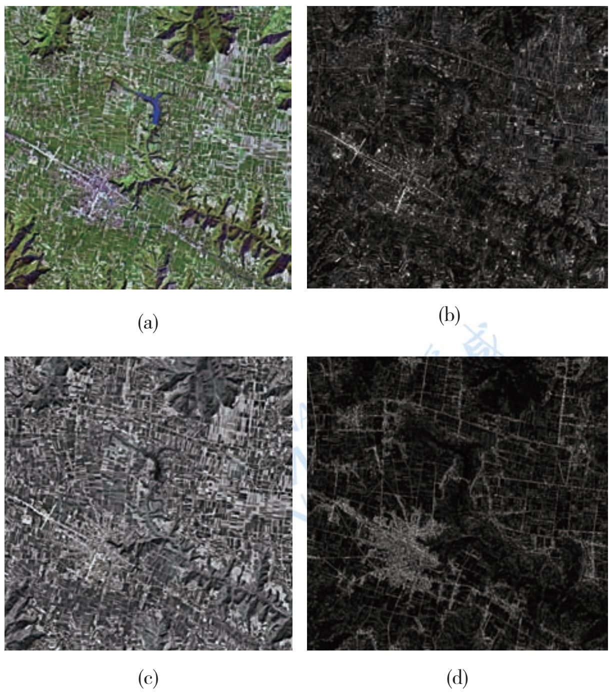

图4 从SPOT 5影像得到的局部空间分析影像
(a) 原始影像; (b) Moran's I; (c) Getis-Ord G_i; (d) Geary's C

表2 在不同指数、邻域规则和间隔距离下的分类精度

		Rook's case		Bishop's Case		Queen's Case		Horizontal		Vertical		Positive Slope		Negative Slope	
		总体分类精度/%	Kappa系数	总体分类精度/%	Kappa系数	总体分类精度/%	Kappa系数	总体分类精度/%	Kappa系数	总体分类精度/%	Kappa系数	总体分类精度/%	Kappa系数	总体分类精度/%	Kappa系数
(a)在3个不同指数下(间隔距离为1)	Moran's I	92.79	0.9120	93.16	0.9164	93.04	0.9150	92.83	0.9124	92.99	0.9144	93.32	0.9184	93.12	0.9159
	Getis–Ord G_i	94.16	0.9285	94.39	0.9313	94.32	0.9305	93.77	0.9238	93.67	0.9226	94.21	0.9292	94.05	0.9272
	Geary's C	93.55	0.9212	93.90	0.9254	93.92	0.9256	92.54	0.9090	93.11	0.9159	93.34	0.9186	93.29	0.9180
(b)在Getis–Ord G_i中的7种不同间隔距离下	1	94.16	0.9285	94.39	0.9313	94.32	0.9305	93.77	0.9238	93.67	0.9226	94.21	0.9292	94.05	0.9272
	2	94.49	0.9327	94.67	0.9349	94.70	0.9352	94.14	0.9283	94.23	0.9294	94.58	0.9338	94.60	0.9340
	3	95.02	0.9391	95.00	0.9389	94.93	0.9380	94.35	0.9309	94.56	0.9336	94.62	0.9343	94.59	0.9339
	4	95.12	0.9404	94.70	0.9352	94.77	0.9361	94.25	0.9296	94.48	0.9326	94.39	0.9313	94.30	0.9302
	5	94.85	0.9371	94.20	0.9290	94.50	0.9328	94.17	0.9287	94.24	0.9295	94.14	0.9283	93.88	0.9251
	6	94.48	0.9326	93.90	0.9254	94.31	0.9304	94.20	0.9290	93.92	0.9256	93.84	0.9247	93.66	0.9224
	7	94.21	0.9292	93.74	0.9235	94.07	0.9275	94.08	0.9276	93.68	0.9227	93.36	0.9188	93.43	0.9196

表3 在四种不同方法下，各类地物监督分类精度评价结果 /%

	只使用多光谱影像		多光谱影像+Moran's I		多光谱影像+Getis-Ord G_i		多光谱影像+Geary's C	
	生产精度	用户精度	生产精度	用户精度	生产精度	用户精度	生产精度	用户精度
水体	83.28	95.77	93.03	93.78	93.73	96.81	93.03	95.80
密集的耕地	95.86	89.13	97.53	86.75	96.50	91.24	96.40	87.94
不透水层	80.70	94.97	95.01	93.62	96.81	91.97	97.03	93.78
稀疏的耕地	92.35	91.18	96.23	89.56	94.54	93.89	98.31	94.56
阴影	96.85	98.38	97.00	98.09	98.11	98.44	97.48	98.15
林地	76.25	94.24	77.12	96.36	83.24	95.08	78.46	94.73

3.2 邻域规则的选择

经过分析，得出邻域规则对地物分类精度的影响取决于像元比较方向与地物分布方向的相似程度。表2的(b)部分是比较在Getis-Ord G_i指数下，不同邻域和间隔距离对总体分类精度的影响。

就稀疏的耕地来说，其整体分布方向有由北向南的趋势(图3(d)和图4(a))，并且从图5(d)得出，它的分类精度在Vertical邻域规则下达到最好。同时，从不透水层最理想分类精度出现在Positive Slope(图5(c))则能得出以下规律：若选择的邻域规则像元比较方向与地物分布方向相同或相似，则分类精度会提高。另外，从图5(c) (d)也得到，Rook’s Case对于稀疏的耕地和不透水层有着很好的分类效果，由此可知若邻域像元比较方向部分与地物分布方向相同或相似，分类精度也会提高。

作为分布方向不明显的地物：水体和密集的耕地，从图5(a)(b)中可以看出，他们在Rook’s Case下的分类精度最高。不过，因为密集的耕地单个面积较小，并且分布方向不一致，所以密集的耕地只能在Rook’s Case、Bishop’s Case和Queen’s Case这3种多方向邻域规则下才能达到不错的分类效果。而水体绝大部分聚集在一起，且单个面积较大，所以它在所有邻域规则下都有不错的分类精度。

阴影主要因山体遮挡而产生，虽然分散，但是每一块面积较大，空间自相关性较好。从图5(e)中可以得到，阴影除了在Vertical下分类较差，其他6种情况下都比较好。虽然林地与阴影分布情况相似，但是因其光谱特征和局部空间统计特征都与密集的耕地相似，所以总体分类精度比较低。不过从图5(e)依然可

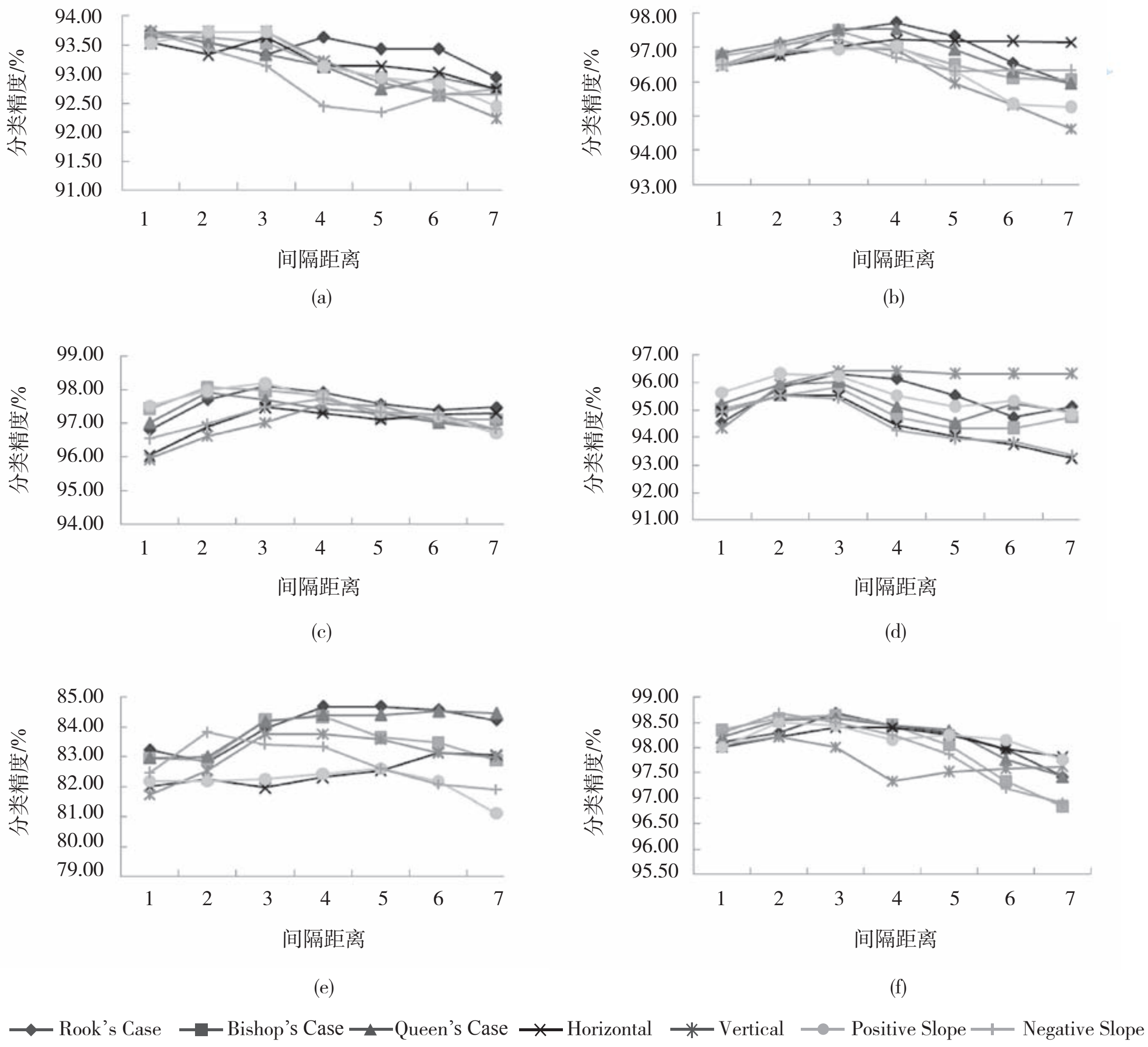

图5 6种地物在不同间隔距离和邻域规则下分类精度

(a) 水体；(b) 密集的耕地；(c) 不透水层；(d) 稀疏的耕地；(e) 林地；(f) 阴影

以得到，对于大面积聚集地物，使用多方向邻域规则分类较好(如图5(e)中的Rook’s Case)。

总之，邻域规则对分类精度的影响的本质就是，与地物分布方向相似的邻域规则像元比较方向能够更准确地得到对应地物的空间自相关性，同时能够减小其他地物对其空间自相关性的影响。因为空间自相关性的提高，分析后产生的纹理波段就能更准确地区分不同的地物，这样就能更好地避免或减轻同物异谱和异物同谱现象的对地物分类精度的影响。

3.3 间隔距离的选择

为了准确地分析间隔距离与分类精度的关系，将地物在不同间隔距离下的分类精度平均值作为标准，具体计算结果如表4。从表4可以得出，相对于其他地物较大且内部空间自相关程度高的地物，在间隔距离为1时，分类精度达到最大(如水体，在间隔距离为1时，平均分类精度为93.66%)；对于较小或内部空间自相关程度低的地物，通常间隔距离大于1分类精度才能达到最大(如稀疏的耕地和林地分别在间隔距离为3、4时，分类精度达到最大的95.97%、83.61%)。另外，地物分类精度随间隔距离增大的走势规律为：当分类精度达到最大前递增，达到最大后都呈减小趋势，可能会有小波动，但是不会超过最大值。

表4 在不同间隔距离下的地物平均分类精度

分类精度/%	1	2	3	4	5	6	7
水体	93.66	93.56	93.49	93.12	92.91	92.88	92.64
密集的耕地	96.57	96.94	97.25	97.17	96.65	96.17	95.91
稀疏的耕地	94.95	95.80	95.97	95.23	94.84	94.79	94.65
不透水层	96.76	97.46	97.71	97.65	97.34	97.19	97.02
阴影	98.15	98.44	98.47	98.22	98.09	97.71	97.40
林地	82.51	82.79	83.39	83.61	83.44	83.30	82.95

3.4 局部空间统计分析对分类精度的总体影响

相对于只用多光谱影像进行的分类来说，添加局部空间统计分析产生的纹理波段的分类能够减少同物异谱和异物同谱对分类精度结果带来的影响。对于林地和密集的耕地(图6(a) (b))，原本因难以分辨而成为未分类的地物被区分开了。同样的，对于稀疏的耕地和不透水层(图6(c) (d))，也得到了很好的分类效果。

因为局部空间统计对分类精度带来的影响，虽然不同指数或参数之间相比较分类精度相差并不大，但是相比于只用多光谱影像进行的分类来说，它们的精度有了很大的提高，最大的分类精度95.12%(图7(b))与只使用多光谱影像的分类精度87.75%(图7(a))相差了7.37%。

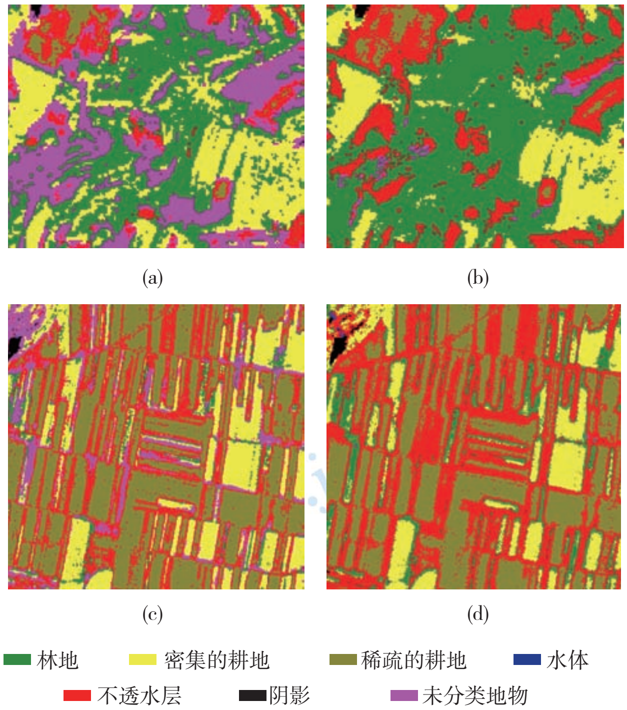

图6 监督分类结果(局部)：林地和密集的耕地

(a) 只使用多光谱影像和 (b) 多光谱影像+Getis-Ord G_i(Rook’s case邻域规则下，间隔距离为4);稀疏的耕地和不透水层：(c) 只使用多光谱影像和 (d) 多光谱影像+Getis-Ord G_i(Rook’s case邻域规则下，间隔距离为4)

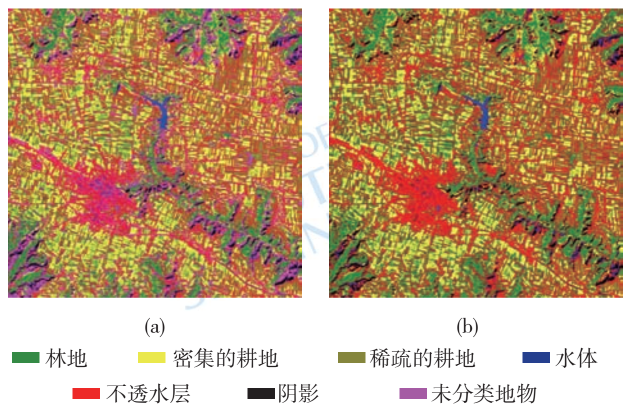

图7 监督分类结果(整体)

(a) 只使用多光谱影像；(b) 多光谱影像+Getis-Ord G_i(Rook’s case邻域规则下，间隔距离为4)

4 结 论

主要探讨了局部空间统计分析中指数、邻域规则和间隔距离对遥感影像分类精度的影响规律。结论如下：

(1)Getis-Ord G_i指数对试验区地物分类精度提高最明显，而且在任何邻域规则和间隔距离下，相比其他2种指数都是最好的；

(2)邻域规则与地物分布方向有关，但对于遥感影像这种总体空间聚集度较小的影像来说，没有太大影响。而间隔距离对于分类精度的影响则比邻域规则大，例如在不同的间隔距离下，密集的耕地最大和最小的平均分类精度相差了1.34%；

(3)局部空间统计分析能够减少同物异谱和异物同谱对分类精度带来的影响，因此，相比于只用光谱波段进行的监督分类，综合利用光谱波段和局部空间统计分析产生的纹理波段所进行的分类效果更好，后者总体分类精度最高达到了95.12%，比前者多了7.37%。

此外，考虑到面向对象的分类方法对局部空间统计分析有着相关影响，本文将对影响分类精度的面向对象分类方法进行研究。

REFERENCES

Anselin L. 1995. Local indicators of spatial association. *Geographical Analysis*, **27**(2): 93–115

Emerson C W, Lam N S N and Quattroch D A. 2005. A comparison of local variance, fractal dimension, and Moran's *I* as aids to multispectral image classification. *International Journal of Remote Sensing*, **26**(8): 1575–1588 DOI: 10.1080/01431160512331326765

Getis A and Ord J K. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, **24**(3): 189–206 DOI: 10.1111/j.1538-4632.1992.tb00261.x

Local Spatial Statistics. 2005. ENVI User's Guide. USA

Wulder M and Boots B. 1998. Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic. *International Journal of Remote Sensing*, **19**(11): 2223–2231 DOI: 10.1080/014311698214983

Zhang S L and Zhang K. 2007. Contrast study on the local indices of spatial autocorrelation. *Statistical Research*, **24**(7): 65–67

附中文参考文献

张松林，张昆．2007．局部空间自相关指标对比研究．统计研究，**24**(7): 65–67