Selection of optimal scale in remotely sensed image classification

HAN Peng[1, 2, 3], GONG Jianya[3], LI Zhilin[3], BO Yanchen[1], CHENG Liang[4]

1. *State Key Laboratory of Remote Sensing Science School of Geography*, *Beijing Normal University*, *Beijing* 100875, *China*;
2. *Xi'an Institute of Surveying and Mapping*, *Shaanxi Xi'an* 710054, *China*;
3. *State Key Laboratory of Information Engineering in Surveying*, *Mapping and Remote Sensing*, *Wuhan University*, *Hubei Wuhan* 430079, *China*;
4. *Department of Geographical Information Science*, *Nanjing University*, *Jiangsu Nanjing* 210093, *China*

Abstract: The effect of scale is continuously attracting attentions in geomatics, bionomics and environmentology. Many methods have been developed for the selection of optimal scale, including those based on local variance, variogram and transformed divergence. However, there are some problems associated with these methods, which limit their applications in practice. This paper presents a new method for optimal scale selection, based on information entropy. The novelty of this new method is that the multi-spectral information is fully used to define the optimal scale. In this method, (a) information entropy is introduced to quantify the uncertainty in image classification; (b) the spatial distribution is also taken into account. This new method has been evaluated and also compared with the existing methods, i.e., those based on local variance, variogram and transformed divergence. Two types of image are used, i.e. TM (Thematic Mapper) which has relatively low resolution and Quickbird image which has high resolution. The experimental results show that the proposed algorithm is capable of effectively determining the optimal scale for these images. In the case of classification of Quickbird image, objected-oriented classification technology is used and the results prove that the new method not only works well with traditional classifiers but also performs with object-oriented classifiers for high resolution images. A comparative analysis shows that the new method performs much better than existing methods.

Key words: optimal scale, resolution, information entropy, multi-spectral information, image classification

CLC number: TP751.1 **Document code:** A

Citation format: Han P, Gong J Y, Li Z L, Bo Y C and Cheng L. 2010. Selection of optimal scale in remotely sensed image classification. *Journal of Remote Sensing*. **14**(3): 507—518

1 INTRODUCTION

Scale effect has been considered as the first challenge in the earth observation (Raffy, 1994). The fundamental reason for the continuing interest in scale in remote sensing is that spatial resolution is the primary scale of measurement (Atkinson & Aplin, 2004). In early 1980s, Markham and Townshend (1981) pointed out that the effect of spatial resolution on the classification accuracy of remotely sensed data was related to two factors. One is the change of the number of mixed pixels which locate near the boundaries among classes and the other is the change of the spectral variations within classes. While the spatial resolution of remotely sensed data becomes finer, the number of mixed pixels will decrease, which is positive for classification accuracy. However, the spectral variation within classes will increase, which is negative for classification accuracy. The net effect of these two factors is a function of the environment of the image scene. Thus, it is necessary to explore the net effects of the change of spatial resolution on the classification accuracy of remotely sensed data (Bo *et al.*, 2005).

Much research has been conducted on identifying the 'optimal' spatial resolution for a particular purpose or investigation, where there is a prior assumption that the features under investigation are scale-dependent. Woodcock and Strahler (1987) degraded an image with relatively fine spatial resolution to successively coarser spatial resolutions to identify the most appropriate scale of observation. Variogram based method was used to identify the optimal scale (Woodcock *et al*., 1988a, 1988b; Atkinson, 1997; Atkinson & Curran, 1997; Atkinson & Tate, 2000; Treitz, 2001). The methods based on local variance and variogram make use of only a single waveband. It was found that substantial differences may exist between the statis-

Received: 2009-04-21; **Accepted**: 2009-06-22

Foundation: National Natural Science Foundation of China (No. 40901199 and No. 40871161), National 973 Project (No. 2007CB714402 and No. 2006CB701300), Programme of Excellent Young Scientists of the Ministry of Land and Resources, and Open Fundation of the State Key Lab of Remote Sensing Science.

First author biography: HAN Peng (1978—), male, Associate Professor. He received the Ph.D. degree from State Key Lab for Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, China, in 2008. He is the author or coauthor of four books and more than 20 papers in journals and conferences. His current research interests include the scale effect in image processing and pattern recognition. E-mail: zb34 xiaowu@163.com

Corresponding author: BO Yanchen, E-mail: boyc@bnu.edu.cn

tics obtained for different wavebands. This is a serious problem because usually only a single spatial resolution will be chosen for a given multi-spectral or hyper-spectral sensor and this limits their application in practice (Bo *et al*., 2005).

Recently, Bo *et al*. (2005) have developed a more complex method in which statistical sperability was used to explore the scale effect of remote sensing data classification and to determine optimal scale. In their work, Landsat TM image with 30m spatial resolution was degraded to different spatial resolutions and the samples for each land use class were taken at each scale. The transformed divergence at each spatial resolution was calculated with the multi-spectral information of the samples. The average transformed divergence was defined to describe the mean value of statistical seperability among every pair of classes in the situations with more than 2 classes. The scale with the highest averaged transformed divergence is finally selected as optimal for image classification. A similar experiment using Jeffry-Matusita (J-M) distance as the measurement of statistical seperability was also conducted. The results showed that the statistical seperability solved the problem associated with the selection of waveband in the single-waveband based methods (local variance and variogram). A severe problem is that the averaged divergence does not take into account the spatial distribution pattern of all land cover types (Sun *et al.*, 2002). Therefore, it is necessary to analyze the spatial distribution patterns in the image so as to identify the optimal scale. Indeed, it is more appropriate to develop new method on the identification of optimal spatial resolution.

The aim of this paper is to illustrate an entropy-based method for the selection of optimal scale in image classification. In this new method, entropy, which is computed with the multi-spectral information in the image, is used to describe uncertainty in image classification. The key idea of the entropy-based method is to choose optimal scale with minimal average entropy of separability from the entropies of multi-scale image data in the same region. The main advantage of the proposed method is that the spatial distribution pattern is taken into account because the entropy of all pixels in the image is obtained.

2 ENTROPY-BASED APPROACH FOR DETERMINATION OF OPTIMAL SCALE

2.1 Information entropy

In information theory, entropy is a measurement to describe uncertainty. Let X be the random message variable, if the probabilities of different message choices are $P_1, P_2,\cdots, P_i,\cdots,P_n$, the entropy of X is computed as follows:

$$H(X)=H(P_1,P_2,\cdots,P_n)=-\sum_{i=1}^{n}P_i\ln P_i \qquad (1)$$

Statistically speaking, $H(X)$ reveals how much uncertainty the variable X has on average. When the value of X is certain, P_i=1, then $H(X)$=0. $H(X)$ is at its maximum when all messages have equal probability (Li & Huang, 2002). Entropy can be used to describe the separability between classes in image classification.

Some researchers have used the entropy concept for describing separability in image classification. Let c be the total number of classes, x the pixel to be classified, $P(\omega_i|x)\overset{\text{def}}{=}p_i\,(i=1,2,\cdots,\ c)$ the posterior probability of x for each class and $\sum_{i=1}^{c}p_i=1$. The entropy of the pixel which is used to describe the separability is then as follows:

$$H_c(x)\overset{\text{def}}{=}H_c(p)=-\sum_{i=1}^{c}p_i\ln p_i \qquad (2)$$

All posterior probabilities being equal means that nothing is known about class membership, and the entropy value reaches the maximum. If, on the other hand, one of the probabilities equals 1 (and the others 0), class membership is completely determined, which is reflected in entropy value 0. Therefore, we iterate each pixel in the image and compute their entropies. The average of all these entropy values is used to express the global uncertainty of the whole image. The average entropy was calculated as follows:

$$H_A=\frac{1}{n}\sum_{i=1}^{n}H_c(x_i) \qquad (3)$$

where, H_A denotes the average entropy, n denotes the numbers of pixels in image, and $H_c(x_i)$ denotes the entropy of a pixel (i.e. i^{th}), which is the same as $H_c(x)$ in Eq. (2).

2.2 Algorithm procedure

The main advantage of average entropy is that the spatial distribution pattern of classification uncertainty in the image is taken into account because the entropy of each pixel is considered. It can be imagined that, due to the scale effect, there must be differences between the classification uncertainties of multi-scale image data in the same region. Thus, it would be sound to select the scale with lowest average entropy as optimal scale for image classification (Han *et al*, 2008).

Based on the ideas described above, a procedure is designed as follows:

(1) Image aggregation: The original image with fine resolution is degraded to successively coarser resolutions by aggregation.

(2) Selecting training samples: As application oriented, the entropy-based method requires a priori information about the type of classes to be extracted. That is, training samples for each class are selected from the image at original resolution.

(3) Estimating the parameters of the probability distribution function (PDF) for each class: The Expectation-Maximization (EM) algorithm was employed to estimate the PDF of the classes on the image.

(4) Computing the entropy which describes the seperability

between the classes: Based on the estimated PDF in Step 3, all pixels on the image are iteratively used to compute the posterior probability for each class.

(5) Averaging the entropy value of each pixel on the image to get the mean of the entropies which indicates the seperability of the whole image.

(6) Finding the scale with the lowest averaged entropy in the serial images and specify it as the optimal scale, at which the desirable overall classification accuracy can be achieved.

The flowchart of the entropy-based method is shown in Fig.1.

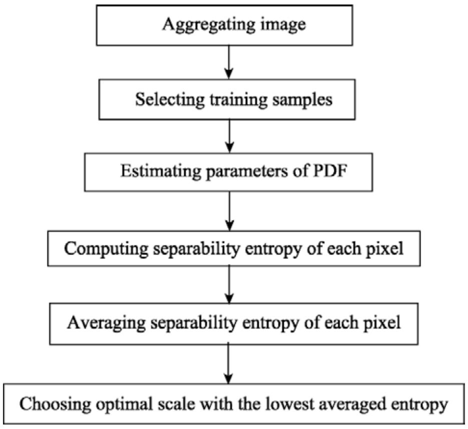

Fig. 1 Flowchart of the entropy based approach for the choice of optimal scale in image classification

2.3 Expectation-Maximization (EM) algorithm

A common task in image processing is the estimation of the parameters of a probability distribution function. In this study, the Expectation-Maximization algorithm is used to estimate the PDF of each landcover class which is assumed to belong to normal distribution, as the EM algorithm is proved to be ideally suitable for this sort of problems in many cases (Ju *et al.*, 2003; Ju *et al.*, 2005). The EM algorithm consists of two major steps, i.e. first an expectation step and then a maximization step. The expectation is with respect to the unknown underlying variables and is predicted by using the current estimate of the parameters and conditioned upon the observations. The maximization step then provides a new estimate of the parameters. These two steps are iterated until convergence (Moon, 1996) is reached. The flowchart is illustrated in Fig.2.

The convergence of the EM algorithm means that, at every iteration, the estimated parameter provides an increase in the likelihood function until a local maximum is achieved, at which point the likelihood function cannot increase any more (but will not decrease) (Moon, 1996).

For the EM algorithm, parameter initialization is a problem to be addressed. We give an initial value to each of the parameters of PDF at the beginning of the iteration, which is computed by using the selected training samples. However, the number of pixels in training areas will decrease with a decrease in spatial resolution. It is worthy of noting the small sample size problem

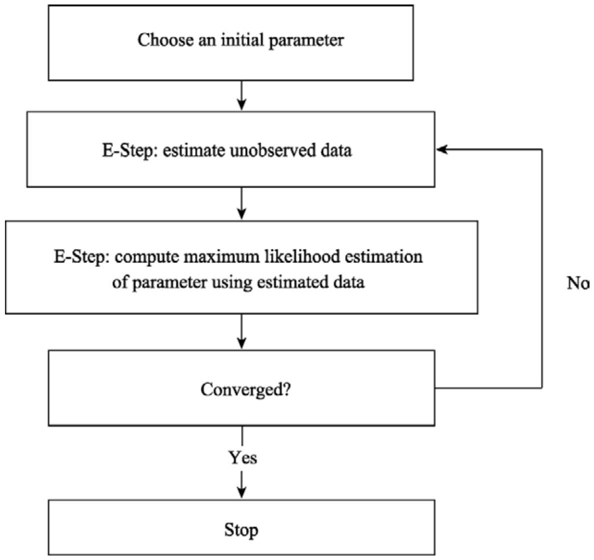

Fig. 2 Flowchart of the entropy based approach for the choice of optimal scale in image classification

will have impact on the iteration of the EM algorithm for the density estimation. At least 12 training samples for each class are required to ensure that the EM algorithm iterates ad nauseum. If not, the iteration of the density estimation would not be normally carried on.

3 EXPERIMENTAL DESIGN

To conduct experiments properly, it is essential to select some sets of appropriate image data and make proper design of the experiments.

3.1 Description of test area and data sets

In this project, we selected two types of images, i.e. TM with medium resolution and QuickBird with high resolution. The TM image covers Guangzhou and its surrounding areas (Fig.3) and was acquired on 20 November 2001 with spatial resolution of 30m. The image consists of pixels. This TM image was colleted with six bands. Six main classes of land covers are present, i.e., water, forest, farmland, built-up area, grassland and bare soil.

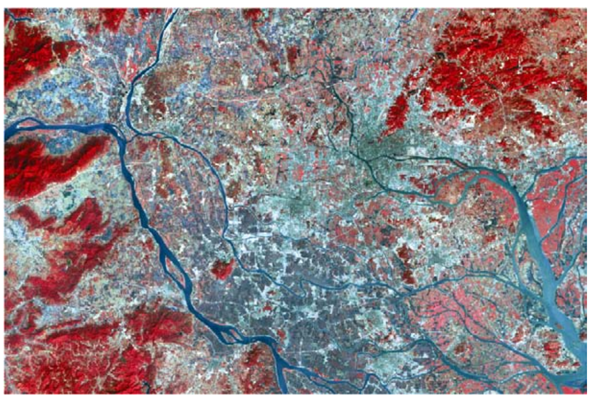

Fig. 3 Experimental TM image

The Quickbird image (Fig.4) covers an area in Hefei (China) and consists of four multi-spectral bands with 2.4m resolution. The image size is 551×599 pixels. The main land cover classes include road, country-road, water, garden plot, built-up area, farmland and bare soil.

Fig. 4 Experimental QuickBird image

3.2 Design of experiments

A number of strategies need to be considered such as the proper selection of classification methods, aggregation of images and use of ground truth.

It was decided to use the maximum likelihood (ML) classifier for the TM image classification because this is the most commonly used on medium resolution. Six classes are designed for the TM image classification, i.e. water, forest, farmland, built-up area, grassland and bare soil.

In the case of QuickBird image, a method called support vector machine (SVM) is employed as the classifier. In use of the SVM classifier for the QuickBird experiment, the kernel type was defined as Radial Basis Function; the Gamma value was set as 0.333; and the penalty parameter was 100.

The Bicubic aggregation method is used for aggregation of the images from the original resolution to new images with a range of resolutions. For the TM image, a set of new images are generated with aggregation with different window sizes, i.e. 2×2, 3×3, 4×4, 5×5, 6×6, 7×7, 8×8, 9×9, 10×10. That is, the resolutions of new images are, 60m,⋯, 300m. For the QuickBird image was gradually degraded with four different window sizes (2×2, 3×3, 4×4, 5×5). As a result, the resolution ranges from 4.8 m to 12 m.

In order to check the validity of the new method, a comparison with those which are based on local variance, variogram and transformed divergence is also carried out. In experiment on local-variance-based method, we set the size of the moving window as 3×3. In experiment on variogram-based method, the spherical model is selected to fit onto the variograms.

3.3 Experimental results

A number of strategies need to be considered such as the proper selection of classification methods, aggregation of images and use of ground truth.

3.3.1 Results on the TM image data set

The result computed by the entropy-based method from the TM image is shown in Fig.5. It can be found that the lowest average entropy is at 60m, which is indicated with rectangle point. The overall accuracy and the classification Kappa coefficient are shown in Fig.6 and Fig.7. From Fig.6 and Fig.7, it can be found that the classification accuracy increases from the original resolution 30m to 60m, with a peak at 60m. As resolution increases further, the Kappa coefficient gradually becomes lower, except at 300m at which the overall accuracy is higher than that at 270m. That is, the classification accuracy reaches the maximum when the resolution is 60m. The highest classification accuracy is highlighted with rectangle point in the figure. This test results show in an absolute sense that 60m defined as the optimal scale by the proposed method is the resolution at which highest overall classification accuracy is reached.

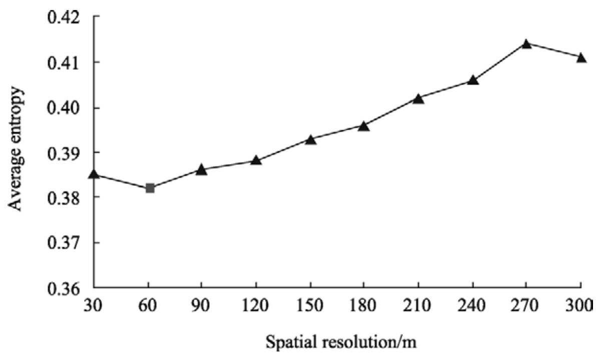

Fig. 5 Relationship between average entropy and resolution in the TM experiment

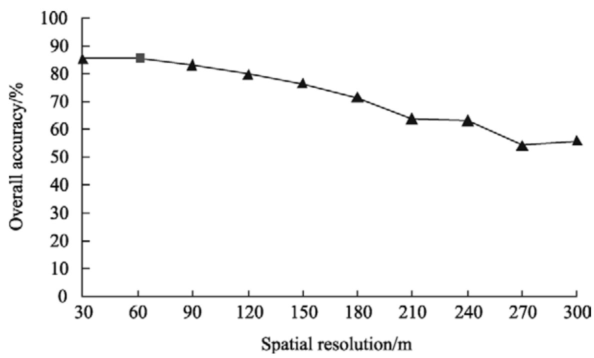

Fig. 6 Relationship between overall classification accuracy and resolution in the TM experiment

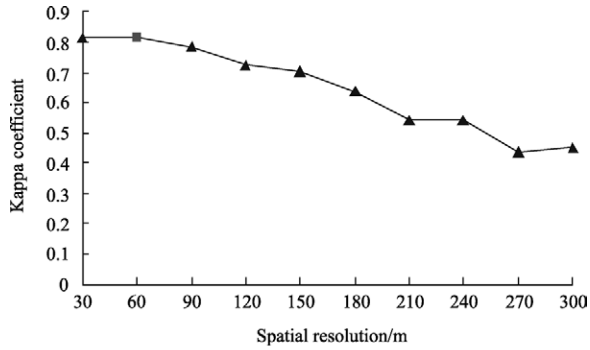

Fig. 7 Relationship between Kappa coefficient and resolution in the TM experiment

In order to assess the superiority of the proposed method, we compared the results obtained by existing methods. The experimental results are shown in Fig.8, Fig.9, Fig.10, respectively. In the case of local-variance based method, it is found that substantial differences may exist between the statistics obtained from different wavebands (Fig.8). The local variance reaches its peak at 60m in the TM 5 waveband and at 180m in the TM 3 waveband. Therefore, no unique solution could be obtained. In the case of variogram-based method, a similar phenomenon is seen to that in local variance. That is, there are differences in the variograms between wavebands. For example, the semi-variogram for TM 2 waveband reaches maximum at approximately 355m implying the size of dominant objects in the study area. For TM 4 waveband, the semi- variogram reaches maximum at approximately 280m. (Fig.9). In the case of transformed-divergence-based method, as shown in Fig.10, the average transformed divergence reaches its peak at 90m, which indicates that 90 m is the optimal scale. But this does not match the result of image classification experiment conducted above. Indeed, this result indicates that the transformed-divergence-based method would lead to wrong conclusion. On the other hand, the entropy based method is capable of predicting the optimal scale reliably, i.e. with high overall classification accuracy reached.

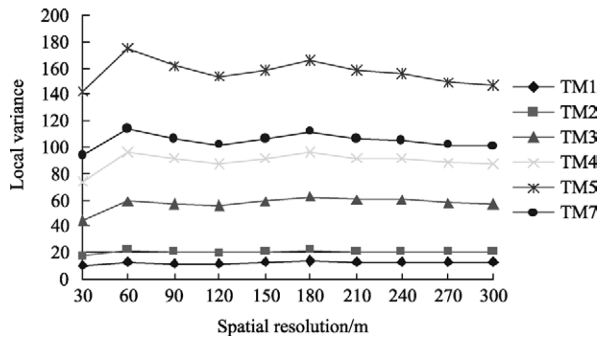

Fig. 8 Local variance as a function of resolution for the different TM band image

3.3.2 Results on the Quickbird Image Data Set

Fig.11 is obtained by the entropy-based method from the Quickbird image. It can be found that the average entropy curve reaches the lowest value at 2.4m. The results of classification of the sequential images by using SVM are shown in Fig.12 and Fig.13. From the classification accuracy figures, it can be seen that, among the five scales, the highest overall accuracy is obtained at the original resolution (2.4m). By comparing Fig.11 with Fig.12 and Fig.13, it can be noted that the accuracy curve is highly similar to entropy curve. It is clear that the proposed method is also effective in the case of classifying high-resolution images with advanced classifier, such as SVM.

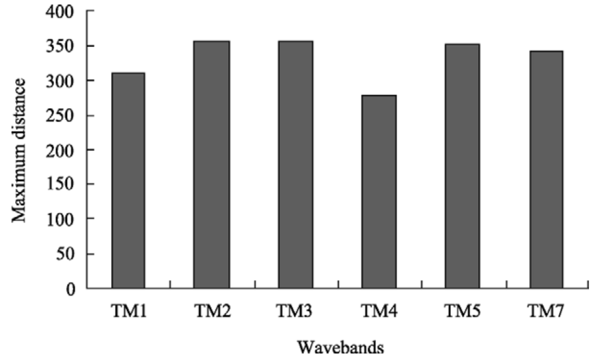

Fig. 9 Comparison of the maximum distance of variogram of the six band images

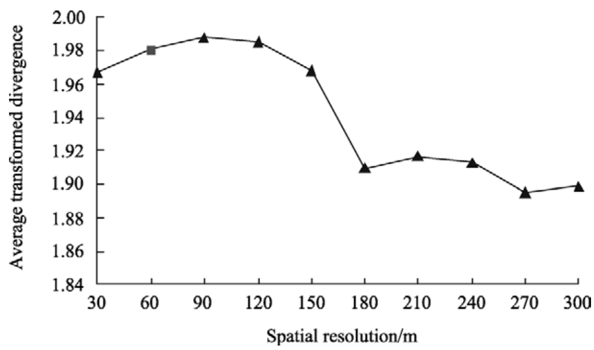

Fig. 10 Relationship between average transformed divergence and resolution in the TM experiment

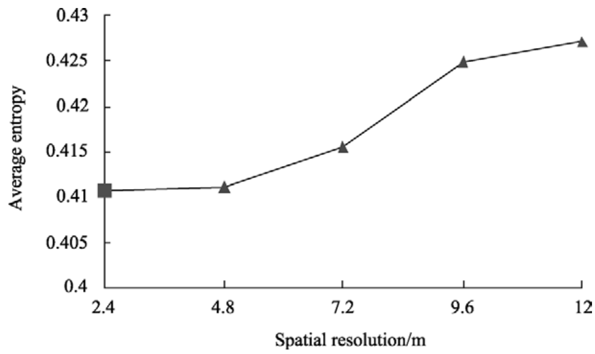

Fig. 11 Relationship between average entropy and resolution in the Quickbird experiment

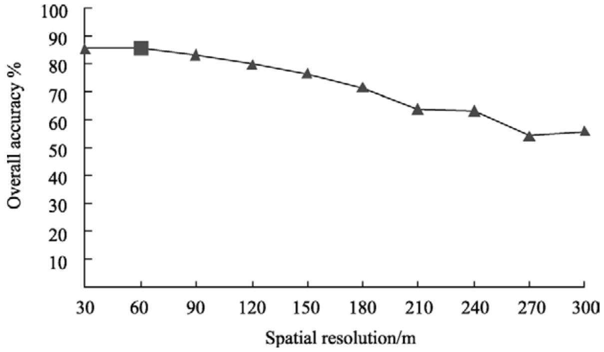

Fig. 12 Relationship between overall classification accuracy and resolution in the Quickbird experiment

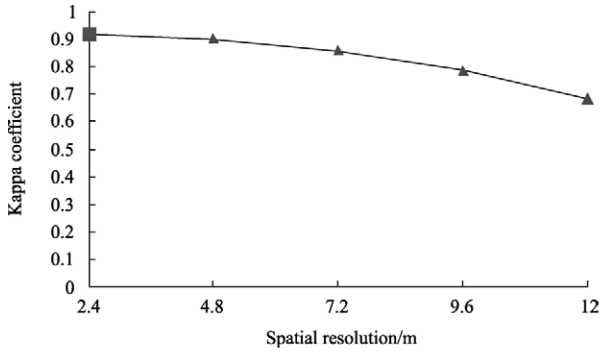

Fig.13 Relationship between Kappa coefficient and resolution in the Quickbird experiment

4 CONCLUSION

In this paper, an entropy-based method is described for the determination of the optimal scale in image classification. In this method, (a) entropy is used to describe the uncertainty in image classification; (b) multi-spectral information is used as multi-dimensional variable so as to overcome the limitation caused by existing methods which make use of only a single band; (c) spatial distribution is also taken into account by calculating the entropy of each pixel. In order to assess the effectiveness of the proposed method, two experiments were conducted. From these tests, we can conclude that (a) the new method is capable of determining optimal scale (resolution) reliably both for medium- and high-resolution images; (b) the new method produces more reliable predictions than existing methods which have the waveband selection or multi-classes problems; (c) moreover, the new method is adaptive because the optimal scale will vary with the number of classes.

Acknowledgements: The authors thank Gu Haiyan at Chinese Academy of Surveying & Mapping for her help in the Quickbird experiment.

REFERENCES

Atkinson P M and Aplin P. 2004. Spatial variation in land cover and choice of spatial resolution for remote sensing. *International Journal of Remote Sensing*, **25**(18): 3687—3702

Atkinson P M and Curran P J. 1997 b. Choosing an appropriate spatial resolution for remote sensing investigations. *Photogrammetric Engineering and Remote Sensing*, **63**(12): 1345—1351

Atkinson P M and Tate N J. 2000. Spatial scale problems and geostatistical solutions: a review. *Professional Geographer*, **52**(4): 607—623

Atkinson P M. 1997 a. Selecting the spatial resolution of airborne MSS imagery. *International Journal of Remote Sensing*, **18**(9): 1903—1917

Bo Y C, Wang J F and Li X. 2005, Exploring the scale effect in land cover mapping from remotely sensed data:the statistical separability-based method. IGARSS 2005

Han P, Gong J Y and Li Z L. 2008. A new approach for choice of optimal spatial scale in image classification based on entropy. Geomatics and Information Science of Wuhan University, **33**(7): 676—679

Ju J C, Gopal S and Kolaczyk E D. 2005. On the choice of spatial and categorical scale in remote sensing land cover classification. *Remote Sensing of Environment*, **96**: 62—77

Ju J C, Kolaczyk E D and GOPAL S. 2003. Gaussian mixture discriminant analysis and sub-pixel land cover characterization in remote sensing. *Remote Sensing of Environment*, **84**: 550—560

Li Z L and Huang P Z. 2002. Quantitative measures for spatial information of maps. *International Journal of Geographical Information Science*, **16**: 699—709

Markham B L and Townshend J R G. 1981. Land cover classification accuracy as a function of sensor spatial resolution. The 15th International Symposium of Remote Sensing of Environment

Moon T K. 1996. The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, **13**(11): 47—60

Raffy M. 1994. Change of scale theory: a capital challenge for space observation of earth. *International Journal of Remote Sensing*, **15**(12): 2353—2357

Sun J X 2002. The Principle, Modern Pattern Recognition. Beijing, Press of National University of Defense Technology.

Treitz P. 2001. Variogram analysis of high spatial resolution remote sensing data: an examination of boreal forest ecosystems. *International Journal of Remote Sensing*, **22**(18): 3895—3900

Woodcock C E and Strahler A H. 1987. The factor of scale in remote sensing. *Remote Sensing of Environment*, **21**: 311—332

Woodcock C E, Strahler A H and Jupp D L B. 1988a. The use of variograms in remote sensing I: Scene models and simulated images. *Remote Sensing of Environment*, **25**: 323—348

Woodcock C E, Strahler A H and Jupp D L B. 1988b. The use of variograms in remote sensing II: Real digital images. *Remote Sensing of Environment*, **25**: 349—379

遥感影像分类中的空间尺度选择方法研究

韩　鹏[1,2,3], 龚健雅[3], 李志林[3], 柏延臣[1], 程　亮[4]

1 北京师范大学 地理学与遥感学院, 遥感科学国家重点实验室, 北京 100875;
2 西安测绘研究所, 陕西 西安 710054;
3 武汉大学 测绘遥感信息工程国家重点实验室, 湖北 武汉 430079;
4 南京大学 地理信息科学系, 江苏 南京 210093

摘　要: 提出了一种新的基于信息熵的空间尺度选择方法。该方法充分利用了遥感影像的多光谱信息。在这个方法中, 信息熵被用于评价影像类别可分性的定量标准; 另外影像的空间分布特征也被考虑。该方法与已有方法, 即基于局部方差的方法、基于变异函数(Variogram)的方法、基于离散度的方法, 进行了比较。TM 和 QuickBird 两种影像被引入到评价中来。实验结果表明, 本方法能够准确地确定两种实验影像的最优分类精度所对应的空间尺度。QuickBird 影像采用了面向对象的分类方法进行实验, 这表明本方法不仅适合于传统的分类方法, 同时也适用于面向对象的方法。通过比较分析表明, 本文方法明确优于已有各种方法。

关键词: 最优尺度, 分辨率, 信息熵, 多光谱信息, 影像分类

中图分类号: TP751.1　　**文献标识码:** A

引用格式: 韩　鹏, 龚健雅, 李志林, 柏延臣, 程　亮. 2010. 遥感影像分类中的空间尺度选择方法研究. 遥感学报, **14**(3): 507—518

Han P, Gong J Y, Li Z L, Bo Y C and Cheng L. 2010. Selection of optimal scale in remotely sensed image classification. *Journal of Remote Sensing*. **14**(3): 507—518

1　引　言

空间尺度问题被认为是从天空观测地球的首要挑战(Raffy, 1994)。遥感空间尺度问题被持续关注的一个主要原因在于空间分辨率是一种基本的测量尺度(Atkinson & Aplin, 2004)。Markham 和 Townshend (1981) 指出遥感分类精度主要受两个因子的影响。一个因子是分类结果中类别边缘的像元, 即混合像元; 另外一个是光谱特征变异。当遥感数据空间分辨率下降时, 处于地物类别边缘处的混合像元数量增多, 分类精度将会降低; 当空间分辨率变高时, 同一地物类别内部的光谱特征变异增大, 地物内部的均质性降低, 使类别间的可分性降低, 导致分类精度降低。可以看出, 空间尺度的转换对这两个因素会有一定的影响。另外, 当空间分辨率固定时, 如果影像中存在许多类内光谱特征差异较大的地物, 则该影像的分类精度也会降低。在这个意义上, 有必要寻找一个使得这对矛盾因子达到平衡的尺度, 以获取最佳的分类结果。由此可见, 遥感影像分类中空间尺度选择的研究是非常必要的(Bo 等, 2005)。

许多学者就遥感分类中的空间尺度选择进行了大量的研究。Woodcock 和 Strahler(1987)采用局部方差选取遥感应用的最优空间尺度。Woodcock 等 (1988a, 1988b)和 Atkinson 等(1997a, 1997b, 2000)以及 Treitz (2001)先后利用变异函数(Variogram)确定遥感应用中的最优分辨率。然而, 局部方差方法和变异函数方法属于利用单波段选择最优尺度的方法, 在多光谱数据应用中存在波段选择的问题, 因此, 它们的推广受到了较大限制(Bo 等, 2005)。

柏延臣等采用变换的离散度作为特征的统计可

收稿日期: 2009-04-21; 修订日期: 2009-06-22

基金项目: 国家自然科学基金(编号: 40901199); 国家自然科学基金(编号: 40871161); 国家 973 项目(编号: 2007CB714402); 国家 973 项目(编号: 2006CB701300); 国土资源部百名优秀青年科技人才计划资助项目; 遥感科学国家重点实验室开放基金资助。

第一作者简介: 韩鹏(1978—), 男, 博士, 助理研究员, 现主要从事遥感应用中的空间尺度问题研究。已发表学术论文 20 余篇, 出版著作 4 部。E-mail: zb34xiaowu@163.com。

通讯作者: 柏延臣, E-mail: boyc@bnu.edu.cn。

分性度量，进而研究遥感最优空间尺度选择(Bo 等, 2005)，以 TM 影像的土地利用/土地覆盖分类为例，首先将原始分辨率的影像以简单平均方法进行尺度上推，得到系列分辨率的影像数据。然后，在原始空间分辨率的图像上，根据该地区土地利用图进行层次随机采样，并以原始分辨率图像上的随机采样位置为掩模，在尺度扩展后的图像上进行同样位置的采样，最后在各空间分辨率上分别计算类对间的变换离散度，并对变换的离散度随空间分辨率变化的规律进行了分析和定性解释。而基于离散度的方法，仅仅是通过度量各类间概率密度重叠程度，没有考虑到整个观测量的实际空间分布特征。另外，基于离散度的方法，只能计算一个类对之间的离散度，对于多类问题，要采用加权和的方法综合各个类对的可分性判据。但是在特征空间中，若有某两类间的离散度很大，可使平均判据变大，这样就掩盖了某些类对判据值较小的情况存在，从而可能降低总的分类正确率(孙即祥, 2002)。

本文提出了基于信息熵的保证影像分类总体精度最优的空间尺度选择方法。信息熵被用来描述影像分类中的不确定性。由于空间尺度效应对影像分类的影响，在不同空间尺度上的影像之间，其描述可分性的熵必然存在差异。基于信息熵的最优空间尺度选择方法的核心思想是利用最小平均熵来确定最优尺度。

2 基于信息熵的最优空间尺度选择方法

2.1 描述可分性的信息熵

在信息论中，熵是对信息不确定性的度量，其定义形式如下:

$$H(X)=H(P_1,P_2,\cdots,P_n)=-\sum_{i=1}^{n}P_i\ln P_i \tag{1}$$

式中，X 表示一个随机变量，X 对应的 n 种可能取值的概率为 P_1，P_2,…，P_i,…，P_n。从熵的定义可知，熵表示不确定性，熵越大不确定性越大(Li & Huang, 2002)，因此可以借用信息论中的熵来描述各类别的可分性。

对于 c 类问题，设给定的 x 各类后验概率为 $P(\omega_i|x)\overset{\text{def}}{=}p_i\,(i=1,2,\cdots,\ c)$，$\sum_{i=1}^{c}p_i=1$。则用来描述类别可分性的熵的定义如下:

$$H_c(x)\overset{\text{def}}{=}H_c(p)=-\sum_{i=1}^{c}p_i\ln p_i \tag{2}$$

由熵的性质可知，如果 $H_c(x)$较大，则说明各类的后验概率 $P(\omega_i|x)(i=1,2,\cdots,\ c)$较接近，这时分类识别的正确率可能不会太高；当各类后验概率相等，实现正确分类识别是最困难的，此时对应的描述可分性的熵值最大。由于尺度对影像分类的影响，对于同一区域的不同尺度下的遥感数据，对应的描述可分性的熵值也会有所不同，即熵值越小，可分性越强；熵值越大，可分性越差。因此，应选择使描述可分性的熵值最小的那个空间尺度下的遥感数据进行分类识别。同时，考虑到反映空间分布特征的需要，应该计算影像中每个像元对应的描述类别可分性的熵值。为了能够客观地比较多个空间尺度下的影像可分性，一般将每个空间尺度下的影像中的所有像元的可分性熵值取平均，用这个均值来选取最优空间尺度。这个平均熵值的计算如公式(3)。

$$H_A=\frac{1}{n}\sum_{i=1}^{n}H_c(x_i) \tag{3}$$

式中，n 表示当前影像的像元个数，$H_c(x_i)$表示影像上某个像元的描述可分性的熵值。

2.2 基本思路

基于信息熵的最优空间尺度选择方法，其基本思路是解算每个空间尺度下的影像数据的类别可分性的平均熵，统计计算出的系列尺度下对应的熵均值，熵均值最小的那个尺度就是总体分类精度最高的最优空间尺度。本方法的一个优点就是能够将多光谱信息联合起来选择最优空间尺度，它将多光谱数据看作多维随机变量，利用多维随机变量的概率理论计算每个像元的后验概率，避免了单波段方法的波段选择所产生的不确定性问题；同时，比较基于变换的离散度选择最优尺度的方法，本方法充分考虑了整个观测量的实际空间分布特征，改进了仅仅从各类间概率密度重叠程度进行分析的局限和不足。

基于信息熵的最优空间尺度选择方法具体的实现过程包括下面 6 个步骤。(1) 将原始影像进行尺度变换，生成系列低分辨率影像；(2) 采集原始分辨率下各个类别的样本；(3) 将遥感影像的多光谱信息视为多维随机变量，以采集的样本为基础，利用最大数学期望算法(expectation-maximization algorithm, EM)优化估计各个类别在原始分辨率影像上的多维随机变量的概率分布参数。这与 Han 等(2008)的样本采集和估计方式有所差别。Han 等(2008)分别利用每个分辨率下所采集的样本去估计分别参数，这种方法可能会随着尺度的上推发生样本不均匀的情况，从而导致参数估计的偏差，影响实验结果，

因此本文进行了上述调整；(4) 在由 EM 算法估计出来的分布空间上，统计影像上每个像元从属于每个类别的概率，并计算出描述该像元可分性的信息熵；(5) 在遍历整个影像之后，计算描述该影像可分性的信息熵的平均值；(6) 找出描述可分性的平均熵值最小的空间分辨率，即保证总体分类精度最优的空间分辨率。整个过程如图 1。

图 1　基于信息熵的最优空间尺度选择算法流程图

2.3　EM 算法

基于信息熵的遥感影像分类最优空间尺度选择方法是以像元可分性作为进行尺度选择的基础，具有很强的统计学背景。因此，与各种基于统计学的方法一样，基于信息熵的遥感影像分类最优空间尺度选择方法首先也要对待分类的各种地物的概率密度函数的参数进行估计，在此基础上计算像元的可分性。本文假设遥感影像的各类地物的概率密度函数符合高斯分布，采用在较多文献(Ju 等, 2003, 2005)中被使用的最大数学期望算法来进行概率密度函数分布参数的估算。

最大数学期望算法是对不完整数据问题进行最大似然估计的一种常用算法，它无需任何外来数据和先验知识，仅从观测数据本身得到参数的估计值。标准的 EM 算法是一个迭代的过程，每次迭代由求期望值和期望最大化两个步骤组成，前者根据待估计参数的当前值，从观测数据中直接估计概率密度的期望值，后者通过最大化这一期望来更新参数的估计量，这两步在整个迭代过程中依次交替进行，直至迭代过程收敛(Moon, 1996) (图 2)。

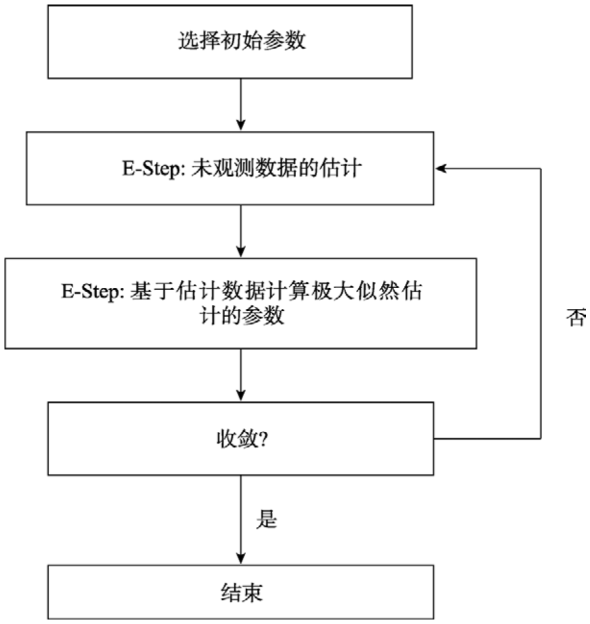

图 2　EM 算法流程图(Moon, 1996)

3　实验与分析

3.1　实验数据

为了对本文所提出的算法进行评价，这里将 Landsat TM 和 QuickBird 两种影像引入到评价中来，以检验该方法的效能。

实验所采用的 TM 影像位于珠三角区域的广州地区。该影像的拍摄时间为 2001-12-20。实验中，选取其中一个 3600×2400 的区域(图 3)。实验区域以山地为主，林地覆盖率高，且有部分草地；研究区域内还有大片水域，由于处于冬季枯水期，出现很多旱滩地带。本实验的分类种类包括水体、林地、耕地、城市建筑、草地和裸土 6 个种类。

图 3　实验用 TM 影像

实验所采用的 QuickBird 影像位于合肥地区，其为 2.4m 分辨率的多光谱影像(图 4)。实验区域包括高速道路、乡村道路、水体、耕地、城市建筑、草地和裸土 7 种类型。

图 4　实验用 QuickBird 影像

3.2　实验设计

对于第一个实验，采用传统的极大似然分类法对 TM 影像进行分类。由于该分类方法对于高分辨率影像不太适宜，我们采用 SVM 分类器对 QuickBird 影像进行分类(其中 Kernel Type 采用 Radial Basis Function, the Gamma value 被设置为 0.333，而 Penalty Parameter 被设置为 100)。

本文采用 Bicubic 方法对两组实验影像进行空间尺度上推。其中，对于 TM 影像，以 30m 为间距，用 Bicubic 方法分别生成了由 90—300m 共 9 个空间尺度下的系列分辨率影像。同时，为了将原始分辨率下的影像数据进行比较，故将原始 30m 分辨率的 TM 影像也作为一个尺度的数据参与分析比较，该实验共包括 10 个空间尺度下的数据。对于 QuickBird 影像，采用 4 个尺度的转换窗口(2×2，3×3，4×4，5×5)进行空间尺度转换。根据 ENVI 所采集的系列分辨率下的样本，利用 EM 算法分别优化估计出原始分辨率下各个类别的多维随机变量的概率分布特征。在 EM 算法迭代出来的概率分布下，计算出每个分辨率下描述每个像元对应的各个类别的可分性的信息熵，并计算出每个分辨率下对应的熵值的均值。

为了检验本文提出的基于信息熵的遥感分类最优空间尺度选择方法的有效性和应用性，将基于信息熵的方法与局部方差法、变异函数法、变换离散度法 3 个方法进行了比较。这 4 种方法分别进行了实验测试。其中，基于信息熵的方法和变换离散度法采用的样本是一样的。基于局部方差的方法所采用的滑动窗口尺寸设置为 3×3；对于基于变异函数的方法，采用球状模型(spherical model)被用来模拟 TM 影像的变异函数，并且步长设为 30m。

3.3　实验结果与分析

3.3.1　TM 影像的实验结果

TM 影像基于信息熵的最优空间尺度选择实验的结果如图 5，可以看出描述可分性的平均熵在 60m 达到最小值。图 6 和图 7 分别反映了总体分类精度、Kappa 系数与分辨率之间的关系。从图 6 可以看出，影像分类的总体精度在 60m 达到最优，之后随着空间分辨率的增大而逐渐变小(除了在 270m 处有些差异)。同样的情况也出现在图 7 中，Kappa 系数与分辨率的变化关系与图 6 一致。由此看出该 TM 影像的最优分辨率位于 60m。

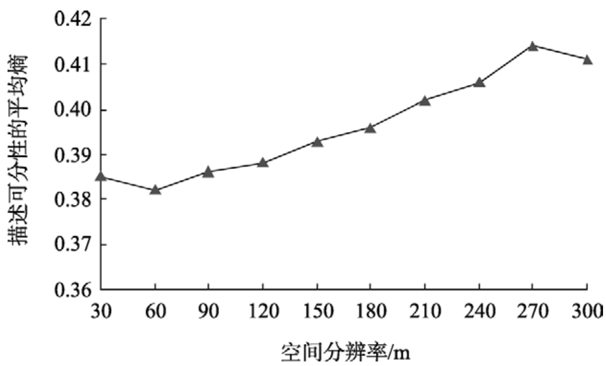

图 5　TM 影像上描述可分性的平均熵随空间尺度变化图

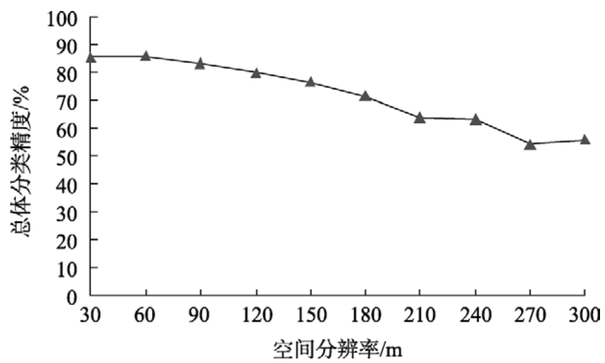

图 6　TM 影像上总体分类精度随空间尺度变化图

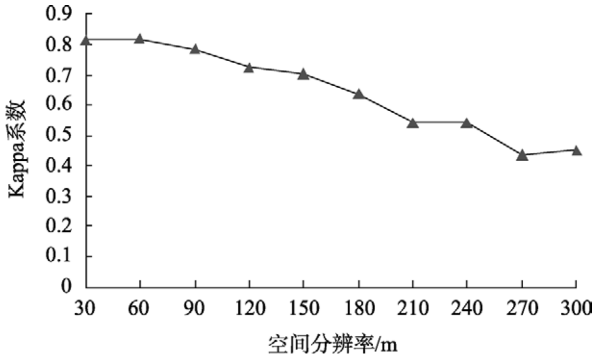

图 7 TM 影像上 Kappa 系数随空间尺度变化图

为了进一步验证本文所提方法的优越性，基于局部方差的方法、基于变异函数的方法和基于变换离散度 3 种方法也分别进行了实验，实验结果分别如图 8、图 9 和图 10。对于基于局部方差的方法，情况比较复杂，不同波段的局部方差随空间分辨率的变化曲线有差异，不同的波段选择的最优空间尺度有所不同(图 8)。例如，利用 TM5 所确定的最优空间尺度在 60m，但是 TM3 所确定的最优空间尺度却出现在 180m，这就让研究人员很难确定到底选用哪个波段的结果。对于基于变异函数的方法，也存在与局部方差方法一样的问题(图 9)。利用 TM2 所确定的最大变程是 355m，而 TM4 所确定的最大变程却出现在 280 m，选择不同的波段出现了不同的结果。对于基于变换离散度的方法(图 10)，虽然它没有出现像局部方差法和变异函数法所存在的波段选择问题，但是它所对应的描述类别可分性的平均离散度是在 90m 处达到最大值，即它所确定的最优空间尺度是 90m，这与影像分类实验所得到的结果不一致。

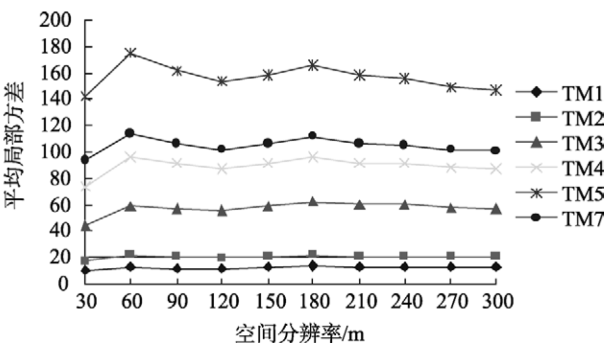

图 8 TM 影像上各个波段的平均局部方差随空间尺度变化图

3.3.2 QuickBird 影像的实验结果

QuickBird 影像的基于信息熵的最优空间尺度选择实验的结果如图 11，可以看出描述可分性的平均熵在 2.4m 达到最小值。图 12 和图 13 分别反映了总体分类精度、Kappa 系数与分辨率之间的关系。从图 12 可以看出，影像分类的总体精度在 2.4m 达到最优，之后随着空间分辨率的增大而逐渐变小。同样的情况也出现在图 13 中，Kappa 系数与分辨率的变化关系与图 12 一致。由此可以看出该 QuickBird

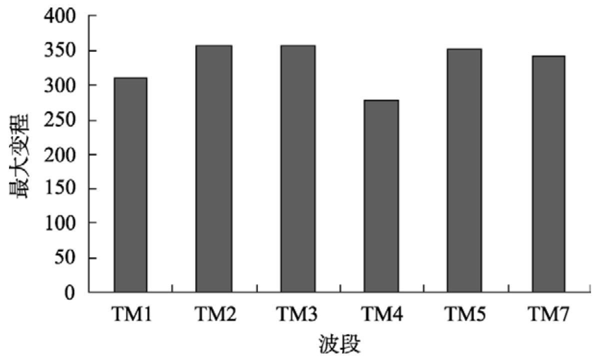

图 9 TM 影像上各个波段的变异函数的最大变程随空间尺度变化图

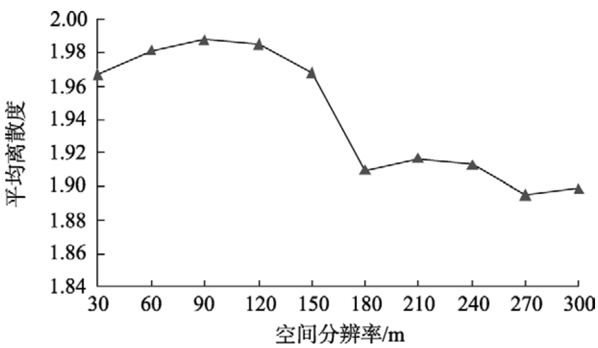

图 10 TM 影像上描述类别可分性的平均离散度随空间尺度变化图

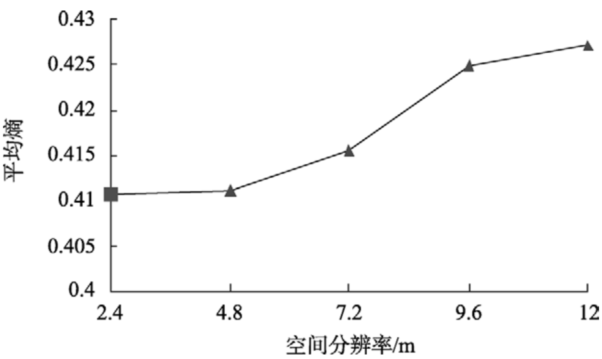

图 11 QuickBird 影像上描述可分性的平均熵随空间尺度变化图

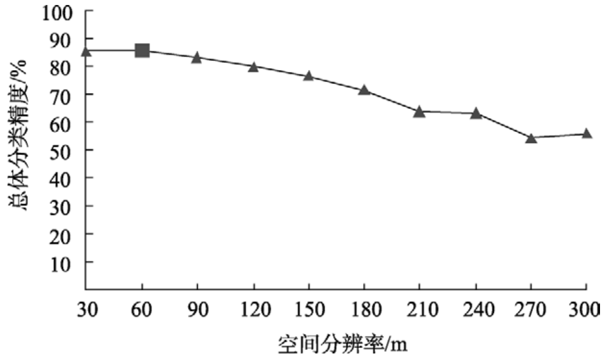

图 12 QuickBird 影像上总体分类精度随空间尺度变化图

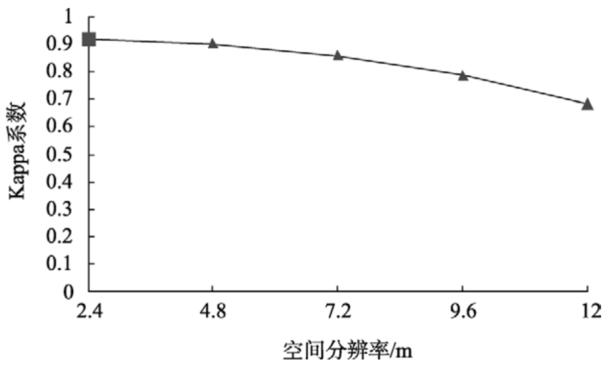

图 13 QuickBird 影像上 Kappa 系数随空间尺度变化图

影像的最优分辨率位于 2.4m。可以看出，本文提出的方法也适合面向对象的高分辨率影像分类中的最优空间尺度选择。

4 结 论

本文所提出的基于信息熵的遥感分类最优空间尺度选择方法，能够比较准确地选择出保证遥感影像分类的总体精度达到最优时所对应的空间尺度。该方法首先利用将多光谱信息作为多维随机变量来确定最优空间尺度，解决了局部方差法和变异函数法在多光谱影像处理时所遇到的波段选择问题；同时，它的计算过程遍历了整个影像，充分考虑了影像上的空间分别特征，避免了基于变换离散度的方法仅仅根据人工采集各个类别的样本进行尺度选择的不足。实验结果表明，基于信息熵的遥感分类最优空间尺度选择方法不仅适合传统的分类方法，还适用于面向对象的高分辨率影像分类，能够在一定程度上指导实际遥感分类中的空间尺度选择。

致 谢 本文中的高分辨率影像分类实验得到了中国测绘科学研究院顾海燕女士的帮助，在此表示最诚挚的感谢。

REFERENCES

Atkinson P M and Aplin P. 2004. Spatial variation in land cover and choice of spatial resolution for remote sensing. *International Journal of Remote Sensing*, **25**(18): 3687—3702

Atkinson P M and Curran P J. 1997 b. Choosing an appropriate spatial resolution for remote sensing investigations. *Photogrammetric Engineering and Remote Sensing*, **63**(12): 1345—1351

Atkinson P M and Tate N J. 2000. Spatial scale problems and geostatistical solutions: a review. *Professional Geographer*, **52**(4): 607—623

Atkinson P M. 1997 a. Selecting the spatial resolution of airborne MSS imagery. *International Journal of Remote Sensing*, **18**(9): 1903—1917

Bo Y C, Wang J F and Li X. 2005, Exploring the scale effect in land cover mapping from remotely sensed data:the statistical separability-based method. IGARSS 2005

Han P, Gong J Y and Li Z L. 2008. A new approach for choice of optimal spatial scale in image classification based on entropy. Geomatics and Information Science of Wuhan University, **33**(7): 676—679

Ju J C, Gopal S and Kolaczyk E D. 2005. On the choice of spatial and categorical scale in remote sensing land cover classification. *Remote Sensing of Environment*, **96**: 62—77

Ju J C, Kolaczyk E D and GOPAL S. 2003. Gaussian mixture discriminant analysis and sub-pixel land cover characterization in remote sensing. *Remote Sensing of Environment*, **84**: 550—560

Li Z L and Huang P Z. 2002. Quantitative measures for spatial information of maps. *International Journal of Geographical Information Science*, **16**: 699—709

Markham B L and Townshend J R G. 1981. Land cover classification accuracy as a function of sensor spatial resolution. The 15th International Symposium of Remote Sensing of Environment

Moon T K. 1996. The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, **13**(11): 47—60

Raffy M. 1994. Change of scale theory: a capital challenge for space observation of earth. *International Journal of Remote Sensing*, **15**(12): 2353—2357

Sun J X. 2002. The Principle, Modern Pattern Recognition. Beijing, Press of National University of Defense Technology

Treitz P. 2001. Variogram analysis of high spatial resolution remote sensing data: an examination of boreal forest ecosystems. *International Journal of Remote Sensing*, **22**(18): 3895—3900

Woodcock C E and Strahler A H. 1987. The factor of scale in remote sensing. *Remote Sensing of Environment*, **21**: 311—332

Woodcock C E, Strahler A H and Jupp D L B. 1988a. The use of variograms in remote sensing I: Scene models and simulated images. *Remote Sensing of Environment*, **25**: 323—348

Woodcock C E, Strahler A H and Jupp D L B. 1988b. The use of variograms in remote sensing II: Real digital images. *Remote Sensing of Environment*, **25**: 349—379

附中文参考文献

韩鹏，龚健雅，李志林．2008．基于信息熵的遥感分类最优空间尺度选择方法．武汉大学学报．信息科学版，**33**(7)：676—679

孙即祥. 2002．现代模式识别．长沙：国防科技大学出版社