

# THE STUDY ON THE CHOICE OF OPTIMAL SCALE IN IMAGE CLASSIFICATION

P. Han<sup>a,\*</sup>, J.Y. Gong<sup>a</sup>, Z.L. Li<sup>a</sup>, L. Cheng<sup>a</sup>

<sup>a</sup> State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China – zb34xiaowu@163.com

**KEY WORDS:** Remote Sensing, Scale, Resolution, Multiresolution, Classification

## ABSTRACT:

In this paper, a new approach is presented to select the optimal scale for image classification obtaining desirable overall classification accuracy. The objective of the new approach is to choose optimal scale with minimal average entropy of separability. The concept of information entropy is introduced to quantify the uncertainty in image classification. Moreover, the spatial distribution is also taken into account for the choice of optimal scale. Since local variance, variogram and transformed divergence are introduced to select optimal scale in many cases, we design a comparative experiment to compare our approach with the existing ones. By using an image sequence with different spatial resolutions, the experiment show that the new approach is more meaningful than traditional ones, thus benefiting the choice of optimal scale in image classification and many other applications.

## 1. INTRODUCTION

Scale is an important factor in the application of remote sensing which directly influences the effect of analysis. The scale effect was 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, 2004). As an increasing number of sources of multi-spectral remotely sensed data become available, the selection among different spatial resolutions becomes more difficult. To select an image with appropriate spatial resolution for a study, one should examine the characteristics of scene content, especially the changing pattern of a scene as a function of changes in analytical scale and resolution (Chen, 2001). 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 locates near the boundaries among classes, another is the change of the spectral variations within classes. While the spatial resolution of remotely sensed data become finer, the number of mixed pixels will decrease which are positive for high classification accuracy. However, the spectral variation within classes will increase which is negative for high classification accuracy. The net effect of these two factors is the 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).

Woodcock and Strahler (1987) investigated the effect of spatial resolution on the local variance observed in remotely sensed imagery. Woodcock et al. (1988a; 1988b) and Atkinson et al. (1997a; 1997b; 2000; 2004) used variogram to choose an appropriate spatial resolution for remote sensing investigations. A recent development is a method based on the statistical separability to explore the scale effect of remote sensing data classification and to determine optimal resolution (Bo et al., 2005). However, there still exist various problems in defining

the optimal scale for image classification. The problems will be discussed below in detail.

In this paper, a new approach based on information entropy is proposed to select an optimal scale in image classification with desirable overall classification accuracy. Entropy is used to describe uncertainty in image classification. Due to the scale effect on image classification, there must be difference between the entropies of multi-scale image data in the same region. The key idea of the entropy based approach is to choose optimal scale with minimal average entropy of separability. The key contribution is that the spatial distribution is taken into account. Based on the new approach, an algorithm is also put forward to implement choice of optimal scale, thus benefiting the image classification and many other applications.

## 2. RELATED WORK

There are many methods which seem to be useful for selecting optimal scale, including 1) local variance based method, 2) variogram based method, 3) transformed divergence based method. These methods can be divided into two types: 1) methods based on single waveband; and 2) those which combine with multi-spectral information. The local variance method and variogram method are based on single waveband, and this limits the application of them in multi-spectral remote sensing data because different results may be obtained from the multi-bands. Bo et al. introduced the method based on the transformed divergence to explore the scale effect of remote sensing data classification and to determine optimal resolution (Bo et al., 2005). This method combined with multi-spectral information which solved the uncertainty in choosing waveband. But, transformed divergence only focuses on the probability density function of the manly selected class samples and ignores the posteriori probability. The spatial distribution of the image is not taken into account.

It is desirable to consider an alternative approach to the problem of choosing optimal scale, so we propose an entropy

---

\* Han Peng, PHD candidate in Wuhan University, China. His research interests are scale effect in remote sensing. E-mail: zb34xiaowu@163.com

based approach. In the new method, multi-spectral information is used as multi-dimension variable and this can avoid the limitation resulted from single band. Furthermore, the new approach takes into account the spatial distribution and overcomes the shortcomings of traditional statistical separability.

### 3. CHOICE OF OPTIMAL SCALE BASED ON ENTROPY

#### 3.1 Information Entropy

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

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

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 and Huang, 2002). Entropy can be used to describe the separability between classes in image classification.

Many researchers have adopted the entropy concept for separability in image classification. Let  $C$  be the total number of classes,  $x$  the pixel to be classified,

$P(\omega_i|x) \stackrel{def}{=} p_i (i = 1, 2, \dots, c)$  the posterior probability of  $x$  for each class and  $\sum_{i=1}^c p_i = 1$ . The entropy of the pixel to

describe the separability is then as follows:

$$H_c(x) \stackrel{def}{=} H_c(p) = -\sum_{i=1}^c p_i \log_{10} p_i \quad (2)$$

According to the characters of entropy, the entropy in (2) will become larger when the posterior probabilities are close to each other, and this indicates that it is difficult to get high classification accuracy. If the posterior probabilities are equal to each other, the classification accuracy will be lowest and the entropy will reach the maximum. Due to the scale effect on image classification, there must be difference between the entropies of multi-scale image data in the same region. It should be sound to select the scale with lowest entropy as the optimal scale for image classification. In this study, entropy is used to choose the optimal scale for image classification. Moreover, for the purpose of taking account of spatial distribution, the entropy of each pixel in the image is computed. Furthermore, it is better to select the average entropy of all pixels on each scale as the measurement for choosing optimal scale, which the differences between the pixel numbers of each scale can be avoided. The average entropy can be calculated as follows:

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

where,  $H_A$  denotes the average entropy;  $n$  denotes the numbers of the pixel in image; and  $H_c(x_i)$  denotes the entropy of one pixel, which is the same to  $H_c(x)$  in (2).

#### 3.2 The Line of Thought

The objective of the new method is to choose optimal scale with minimal average entropy of separability. To find out the scale with minimal average entropy, the samples for the classes with same spatial position should be select on the consecutive spatial scales. In order to use the method as we describe, the component probability densities  $p_i$  must be known or learned beforehand. In this paper, we specify them to be themselves Gaussian mixtures and fit the corresponding parameters for each component to pure-class training data using a standard Expectation-Maximization (EM) algorithm, as described in (Ju and Kolaczyk, 2005; Ju et al., 2003). Taking these inferred  $p_i$  as fixed, we then analyze an image at a given scale by computing the average entropy for all pixels. Finally, the scale with minimal average entropy will be selected from the consecutive spatial scales, that is, the optimal scale.

## 4. EXPERIMENT AND ANALYSIS

#### 4.1 Image Sets

The study site is in Guangzhou city which is in the Pearl River Delta (PRD), China. For the study area, Landsat-TM imagery shown in Figure 1 with spatial resolution of 30 m was used to analyze the scale effect of remotely sensed data classification. The TM image was acquired on November 20, 2001, with the size of 3200 × 2400. The TM image was collected in six bands. Six land cover classes are present in this study, i.e., water, forest, farmland, built up, grassland, and bare ground.



Figure 1. The experimental image of TM in color composite with bands 4,3,2 in R,G,B

We aggregated the initial TM image based on the Average up-scaling algorithm. The aggregations were sequential, from 30 m to 300 m resolutions, a total of 10 iterations (30 m increment per iteration). The sample of the six land cover classes were selected by ENVI software at the original resolution and saved as masks to ensure that the images with coarser resolution have training samples at same location. Based on the samples, the

Gaussian probability distribution of the six land cover classes on each resolution is estimated by using EM algorithm. Then the average entropy of each resolution can be obtained.

#### 4.2 Image Classification

In order to test applicability of the entropy based algorithm, we conducted image classification on each resolution with the classification scheme described above, using the maximum likelihood classifier. The overall classification accuracy and kappa coefficient of these images are shown in Figure 2 and Figure 3, respectively.

Figure 2 represents that the overall classification accuracy increases from the original resolution, 30 m to 60 m and a peak is observed at 60 m. As resolution increases past this peak, the overall classification accuracy gradually becomes lower on the order of resolution, except 300 m, at which the overall accuracy is higher than 270 m. The classification accuracy reaches maximum when the resolution is 60 m. The same is true in Figure 3 which shows the relationship between kappa coefficient and resolution. In the consecutive resolutions, 60 m can be defined as the optimal scale with highest overall classification accuracy.

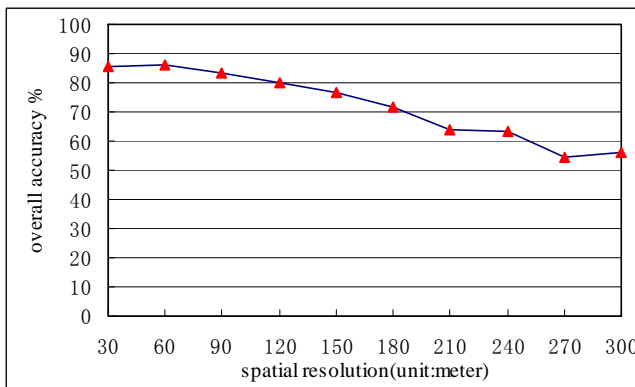


Figure 2. Relationship between overall classification accuracy and resolution

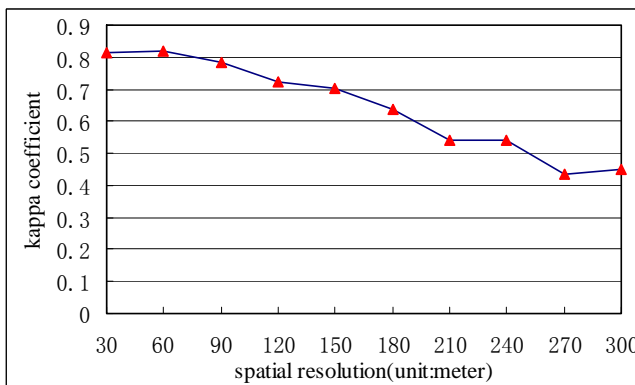


Figure 3. Relationship between kappa coefficient and resolution

#### 4.3 Image Classification

Since local variance, variogram and transformed divergence are respectively introduced in many cases (Woodcock and Strahler, 1987; Woodcock et al., (1988a; 1988b); Atkinson et al., (1997a; 1997b; 2000; 2004); Bo et al., 2005) we make a comparison

between entropy based method, local variance, variogram, and transformed divergence, to check validity of the new algorithm. In the comparative experiments, the four methods are conducted respectively. The entropy based method and the transformed divergence method are processed with the samples picked in the image classification experiment mentioned above. In the local variance experiment, we set the size of the moving window as  $3 \times 3$ . In the variogram experiment, the spherical model is selected to fit to the variograms of the experimental TM image and the lag size is set as 30 m. The experimental results of the four methods are shown in Figure 4, Figure 5, Figure 6 and Figure 7, respectively. The following conclusions are therefore summarized as below.

1. Using the entropy based method, the optimal scale with desirable overall classification accuracy can be selected exactly. Figure 4 has shown that the average entropy decreases from the original resolution, 30 m to 60 m and a foot is observed at 60 m. As resolution increases past this foot, the average entropy gradually becomes higher on the order of resolution, except 300 m, at which the average entropy is lower than 270 m. The shape of average entropy is a mirror of Figure 2 and Figure 3. Based on the entropy theory, 60 m can be selected as the optimal scale with highest overall classification accuracy in the consecutive resolutions.

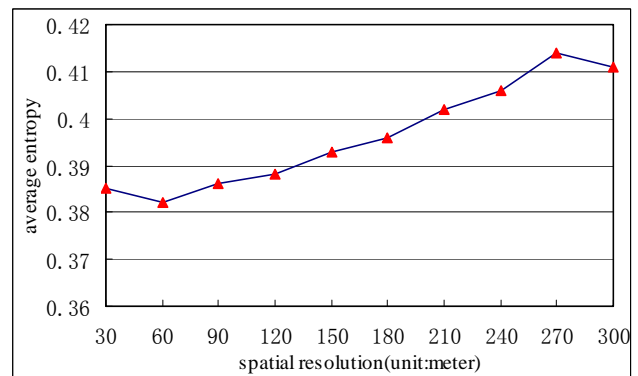


Figure 4. Relationship between average entropy and resolution

2. Using the local variance based method, it is found that substantial differences may exist between the statistic obtained for different wavebands. In Figure 5, the local variance indicates a peak at 60 m in the TM 5 waveband and one at 180 m in the TM 3 waveband. This is a difficult question to answer since the same scene properties have resulted in very different characters of spatial variation in the separate wavebands. This problem is important because generally, a single spatial resolution will be chosen for a given multi-spectral or hyper-spectral sensor (Atkinson, 2004).

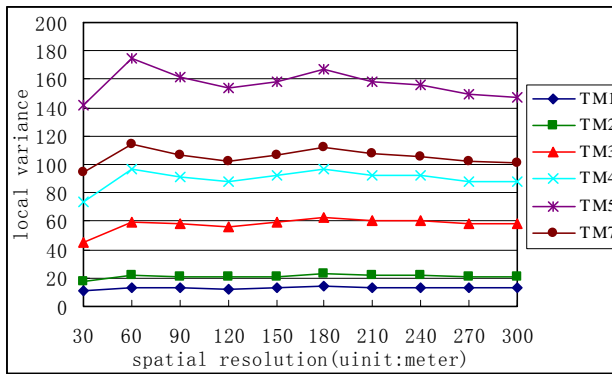


Figure 5. The graph of local variance as a function of resolution for the different TM band image

3. Using the variogram based method, the TM image exhibits differences in the variograms between wavebands, which is similar to the local variance based method. In Figure 6, the maximum distance of variogram in each waveband is shown. In contrast, the semivariance for TM 2 waveband reaches maximum at approximately 355 m implying the dominance of objects of around this size in the study area. For TM 4 waveband, the semivariance reaches maximum at approximately 280 m.

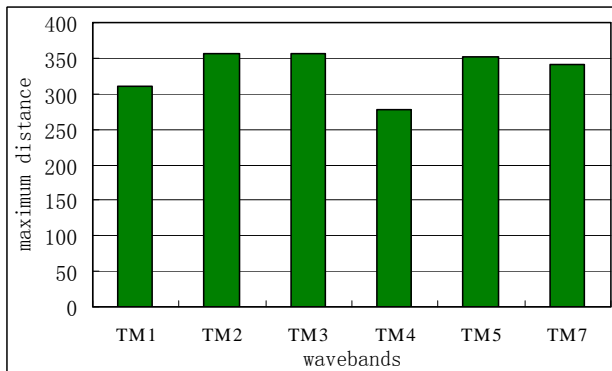


Figure 6. Comparison of the maximum distance of variogram of the six band images

4. Using the transformed divergence based method, the selected optimal scale would not correspond to the result of image classification. Figure 7 shows that the average transformed divergence reaches peak at 90 m, which indicates that 90 m is the optimal scale for desirable overall classification accuracy. But this does not match the result of image classification experiment conducted above. This means that transformed divergence based method would lead to wrong result.

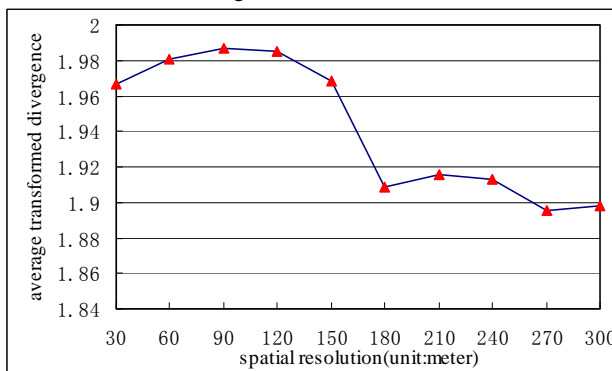


Figure 7. Relationship between average transformed divergence and resolution

## 5. CONCLUSION

A new approach is put forward to select the optimal scale in image classification for desirable overall classification accuracy. The core of the approach is that entropy is used to describe the uncertainty in image classification. In the entropy based algorithm, multi-spectral information is used as multi-dimension variable and this can avoid the limitation resulted from single band. Moreover, by calculating the entropy of each pixel, the spatial distribution is also taken into account. Experiments were conducted to compare the entropy based method, local variance based method, variogram based method and transformed divergence based method. The results show that the entropy based method is more meaningful than the existing ones.

## ACKNOWLEDGEMENTS

The work was supported by the National Basic Research Program of China under Grant 2006CB701300.

## REFERENCES

- Atkinson. P. M., 1997 a. Selecting the spatial resolution of airborne MSS imagery. *International Journal of Remote Sensing*, 18(9), 1903–1917.
- 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), pp. 1345–1351.
- Atkinson. P. M., and Tate. N. J., 2000. Spatial scale problems and geostatistical solutions: a review. *Professional Geographer*, 52(4), pp. 607–623.
- 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), pp. 3687–3702.
- Bo. Y.C., Wang. J.F., Li. X., 2005, Exploring the Scale Effect in Land Cover Mapping from Remotely Sensed Data:the Statistical Separability-based Method. *IGARSS*, pp. 3884-3887.
- Chen. D. M., 2001. Multi-resolution image analysis and classification for urban land use/cover mapping using high resolution imagery. Doctor Dissertation, University of Californian, Santa Barbara/ San Diego State University, USA.
- 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, pp. 550–560.
- 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, pp. 62–77.
- Li. Z. L., and Huang. P. Z., 2002. Quantitative measures for spatial information of maps. *International Journal of Geographical Information Science*, 16, pp. 699–709.

Markham. B.L., and Townshend. J.R.G., 1981. Land Cover Classification Accuracy as a Function of Sensor Spatial Resolution. In: The 15<sup>th</sup> International Symposium of Remote Sensing of Environment, pp. 1075-1090.

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

Woodcock. C. E., and Strahler. A. H., 1987, The factor of scale in remote sensing. *Remote Sensing of Environment*, 21, pp. 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, pp. 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, pp. 349–379.

