# Land cover classification of remotely sensed image with hierarchical iterative method*

LI Peijun** and HUANG Yingduan

(Institute of Remote Sensing and GIS, Peking University, Beijing 100871, China)

Received August 11, 2004; revised August 31, 2004

**Abstract** Based on the analysis of the single-stage classification results obtained by the multitemporal SPOT 5 and Landsat 7 ETM+ multispectral images separately and the derived variogram texture, the best data combinations for each land cover class are selected, and the hierarchical iterative classification is then applied for land cover mapping. The proposed classification method combines the multitemporal images of different resolutions with the image texture, which can greatly improve the classification accuracy. The method and strategies proposed in the study can be easily transferred to other similar applications.

**Keywords:** multitemporal, iterative classification, texture, land cover, image classification.

Land cover classification is one of the most widely used applications of remote sensing. The use of multitemporal remote sensing data in land cover classification is one of the effective methods of obtaining accurate land cover/land use data. For a particular image, different land cover types often show a similar spectral response and are difficult to separate. The advantage of using multitemporal data is that different vegetation types show different spectral characteristics in spectral bands acquired at different dates[1]. Among the multitemporal classification methods, the hierarchical classification is one of the most often used methods[2,3]. The hierarchical classification method is also called layered classification[2], multi-stage classification[2] or stepwise hybrid classification[3]. This classification method uses stepwise classification strategy. In every step (or level), one or more land cover types are extracted using different data combinations, which can more effectively take advantage of different data sets and classification algorithms and thus higher classification accuracy can be achieved than single-stage classification. The classification strategies used in different classification experiments are varied, depending on the image data used and land cover types to be classified. In the previous hierarchical classification, single parameters (e.g. vegetation indices) were used as the threshold for a certain type[2]. In these studies, spatial correlation between pixel and its neighbors was ignored, and the data used usually have the same resolution. The purpose of this paper is to perform land cover classification of multisource dataset, including the multitemporal images with different resolutions as well as the textural information, using a hierarchical iterative approach.

## 1 Study area and data

The area surrounding the Hengshui Lake in Hebei Province was selected as the study area (Fig. 1). This area is the only state-level wetland in North China, and it is an extremely flat agricultural area. The main crops in the area include cotton, winter wheat and corn, and the winter wheat and corn are planted in rotation. The other less common crops include apple and pear, and other vegetation types include woodland, reed and grassland. In early July, corn is planted, the ground is not fully covered by the corn. Other vegetation types flourish at this time. In September, different vegetation types reveal different growing conditions.

Following the crop phenology and the growing season of the vegetation in the study area, the SPOT 5 multispectral data (10 m/pixel) acquired in September 2002 and the LANDSAT 7 ETM+ multispectral data (30 m/pixel for bands 1—5 and band 7) acquired in July 2001 were used in this study. The data with different resolutions could provide complementary information for land cover classification.

The SPOT 5 image was first geometrically cor-

* Supported by National Natural Science Foundation of China (Grant No. 40372130)
** To whom correspondence should be addressed. E-mail: pjli@pku.edu.cn

rected using topographic maps of scale 1∶10000. The ETM+ image was coregistered with georeferenced SPOT 5 image and resampled to 10 m/pixel. The RMS error of the registration is less than 0.35 pixels (3.5 m). No atmospheric correction was conducted, since the classification was performed on composite data acquired in different dates[4]. A subscene of 2180×2060 pixels was used in the study (Fig. 1).

Fig. 1. SPOT 5 image of the study area (Band 4).

The study area is divided into 10 land cover/land use types (Table 1) according to the vegetation types and land use in the study area.

Table 1. Main land cover classes in the study area and the training and test samples used

| | Land core types | Training samples | Test samples |
|---|---|---|---|
| 1 | Rural residence | 3303 | 7657 |
| 2 | Bare land | 1801 | 2194 |
| 3 | Cotton | 3681 | 4105 |
| 4 | Water body | 5083 | 4745 |
| 5 | Corn | 7950 | 12777 |
| 6 | Reed | 5322 | 5774 |
| 7 | Grassland | 299 | 134 |
| 8 | Woodland | 608 | 572 |
| 9 | Orchard | 2750 | 2536 |
| 10 | Urban | 3414 | 3004 |
| Total | | 34211 | 43498 |

## 2 Methods

The hierarchical classification method proposed in this paper can be summarized as follows: The maximum likelihood (ML) classifier was first used for the classification of different data combinations. The sequence of the final hierarchical classification was then determined based on the analysis of results of the single-stage classifications of different data combinations. The data sets used in single-stage classification process include the original spectral data and the image texture derived from the spectral data. Moreover, since the data with different resolutions were used, it would be better to choose the appropriate resolution for the classification of each class based on the single-stage classification results using the images of different resolutions.

Improved spatial resolution does not always lead to better classification results for land cover mapping with conventional spectral classification[5]. The classification accuracy is a function of two counteracting factors. The first factor is that finer spatial resolution can result in an increase in the number of pure pixels and a decrease in the number of mixed pixels. This factor can increase the classification accuracy. On the other hand, the finer the spatial resolution, the larger the number of detectable sub-class elements. This implies that high within-class spectral variance of land cover classes decreases their spectral separability and results in lower classification accuracy. As a result, as the spatial resolution becomes finer, the classification accuracies may increase for some land cover types, but the classification accuracies may decrease for the land cover types with high within-class variability. Thus, for those classes with high within-class variability, the classification using low resolution data may produce higher accuracies. The overall effect on classification depends not only on the spatial resolution of the image but also on the land cover type to be classified.

The incorporation of image texture into image classification is one of the important methods for improving the classification accuracy. In this paper, variogram in geostatistics is used as texture measure, which is described as[6]

$$\gamma(\boldsymbol{h}) = \frac{1}{2N(\boldsymbol{h})}\sum_{i=1}^{N(\boldsymbol{h})}[DN(x_i+\boldsymbol{h}) - DN(x_i)]^2, \tag{1}$$

where $N(\boldsymbol{h})$ is the number of pixel pairs at distance of $\boldsymbol{h}$ (lag distance), $DN(\circ)$ is the gray level values of pixels at the locations $x_i$ and $x_i + \boldsymbol{h}$. Equation (1) quantitatively describes the spatial correlation between pixels and has been widely used in extraction of texture and image classification[7,8].

There are three parameters to be determined in the extraction of geostatistical texture: the moving window size, the size and direction of the lag distance. The lag distance of 1 pixel and the average of the four main directions (NS, EW, NW-SE, NE-SW, called omnidirectional) were used in the study. The window size was determined by the trials. The window sizes of 3×3, 5×5, 7×7, and 9×9 were investigated for texture computation. The results indicate that the error (misclassification) at edge is still very large even the window size of 3×3 pixels was used. In order to reduce the edge effect, four pixels surrounding the central pixel (EW, NS) are considered for image texture calculation. The method is called 4-neighbor methods, and it can be described by

$$\begin{aligned}\gamma_{i,j} = \frac{1}{2\times 4}[&(DN_{i,j} - DN_{i+1,j})^2 \\ &+ (DN_{i,j} - DN_{i,j+1})^2 + (DN_{i,j} - DN_{i-1,j})^2 \\ &+ (DN_{i,j} - DN_{i,j-1})^2],\end{aligned} \tag{2}$$

where $DN_{i,j}$ is the digital value of pixel at location $(i, j)$, $\gamma_{i,j}$ is the texture value of pixel at location $(i, j)$. Since the SPOT 5 image has higher spatial resolution and correspondingly more spatial information, SPOT 5 image was used to compute the texture.

The data combinations used in the land cover classification include single-date SPOT 5 multispectral image (4 bands), single-date ETM+ multispectral image (bands 1—5 and 7, total 6 bands), composites of SPOT and ETM+ images, multitemporal classifications incorporating image texture, and principal component analysis based multitemporal classification. The training and test samples used for classification (Table 1) were separately selected from the image using field data, existing land use maps and topographic maps (at scale 1∶10000). The classification accuracy was computed using the confusion matrix.

The hierarchical iterative classification method is determined based on the single-stage classification results using the above data combinations. The basic principle can be summarized as follows: All the single-stage classification results obtained were first compared according to the classification accuracy and the data combinations having the highest accuracy for single class are chosen. The hierarchical classification was then performed by a sequence, which is decided by comparing the classification accuracies of the different land cover types. In the hierarchical iterative classification approach, land cover types are classified one or two at a time, and the data combination used in each iteration is different. At each iteration, the data combination that provides the best result (highest accuracy) among the classes that remain to be classified is selected. At each step the best data combination for a certain class was used for the classification of that class. The class classified in each step was extracted and masked out from the image. The process is done for the remaining classes step by step until the iterative method can not improve the classification accuracies of these types any more. Finally, ML classification is done for the finally remaining classes using an appropriate data combination.

The best data combinations for individual classes were selected using the following rule. Since the classification accuracy for individual land cover types includes the producer's accuracy and the user's accuracy, the data combination with both the highest producer's accuracy and the highest user's accuracy was selected for the class. Otherwise, the combination with the highest sum of the producer's accuracy and user's accuracy was chosen.

The rules for determining the iterative classification sequence are as follows:

1) The class with both the higher producer's accuracy and the higher user's accuracy is selected to be classified first;

2) the class with higher sum of the producer's accuracy and the user's accuracy is first extracted;

3) if the sums of the producer's accuracy and the user's accuracy for two classes are equal or their difference is less than 1%, the data combination with higher producer's accuracy is first used for classification.

To obtain higher classification accuracy, it is desirable to use spatial smoothing or filtering. A 3×3 pixels majority filtering is applied to the classified image.

## 3 Results and discussions

The confusion matrix for the classification with 4-band SPOT 5 multispectral image alone is shown in Table 2. From the table, the overall accuracy is 74.14%. The accuracies for individual classes varied considerably. The classes cotton and water body have the highest accuracies, since they have unique and relatively homogeneous spectral features. Other vegetation classes have different producer's and user's accuracies since the vegetation classes have similar spectral responses. The class woodland is small in size and have similar spectral feature to other vegetation classes, thus it has low classification accuracy. The confusion between rural residence and bareland is significant in this classification.

The classification accuracy using ETM+ is 81.75%, slightly higher than that using SPOT 5 alone. However, the accuracies for some classes like cotton and water body in ETM+ classification are much lower than those in SPOT 5 classification. It is interesting that the classification accuracy of the class urban in ETM+ classification is significantly higher than that in SPOT 5 classification, 3% higher in the producer's accuracy, and 14% higher in the user's accuracy. The results suggest that the ETM+ data may be superior to SPOT data for discriminating class urban. The main reason for this is that the class urban is highly heterogeneous, which has high within-class variability. Greater pixel size of ETM+ data (compared to pixel size of SPOT 5 data) can alleviate some of the spatial/spectral heterogeneity caused by within-class variability. As a result, the class urban has higher accuracy in ETM+ classification. Other classes have similar accuracies in the two classifications.

Table 2. Confusion matrix for the SPOT 5 classification[a)] (Unit: %)

| | Rural residence | Bare land | Cotton | Water body | Corn | Reed | Grassland | Woodland | Orchard | Urban | Producer's accuracy | User's accuracy |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rural Residence | 61.81 | 15.00 | 0 | 0 | 0 | 0.07 | 0 | 0 | 0 | 8.09 | 61.81 | 89.15 |
| Bare land | 23.85 | 82.13 | 0.02 | 0 | 0.05 | 3.79 | 0.75 | 0.17 | 0.32 | 0.60 | 82.13 | 46.41 |
| Cotton | 0.04 | 0.09 | 94.45 | 0 | 0.09 | 0.03 | 0 | 0.35 | 0 | 0 | 94.45 | 99.49 |
| Water body | 0 | 0 | 0 | 99.30 | 0 | 0.09 | 0 | 0 | 0 | 0 | 99.30 | 99.89 |
| Corn | 0 | 0 | 2.63 | 0 | 56.37 | 0.33 | 0 | 22.03 | 14.67 | 0 | 56.37 | 92.02 |
| Reed | 0.26 | 0.91 | 2.27 | 0.27 | 14.54 | 84.90 | 1.49 | 39.51 | 2.60 | 0 | 84.90 | 68.08 |
| Grassland | 0.25 | 0.23 | 0 | 0 | 0.10 | 5.51 | 97.01 | 4.20 | 0.55 | 0 | 97.01 | 24.86 |
| Woodland | 0.29 | 0 | 0.61 | 0 | 12.19 | 2.82 | 0.75 | 26.40 | 2.56 | 0.43 | 26.40 | 7.56 |
| Orchard | 0.85 | 0.36 | 0.02 | 0 | 16.65 | 2.44 | 0 | 7.34 | 79.30 | 0.03 | 79.30 | 45.75 |
| Urban | 12.66 | 1.28 | 0 | 0.42 | 0 | 0.02 | 0 | 0 | 0 | 90.85 | 90.85 | 72.83 |

a) Overall accuracy: 74.14; kappa coefficient: 70.00

The classification accuracies for the composites of the SPOT 5 and ETM+ images are higher than those for the single-date SPOT 5 or ETM+ image. For example, the classification accuracy for the 10-band combination (4-band SPOT image and 6-band ETM+ image (bands 1—5 and 7)) is 12.86% higher than the single-date SPOT image. For individual classes, all classes have higher accuracies in multitemporal classifications with the exception of the class cotton, which indicates that multitemporal data can effectively discriminate among vegetation classes. In multitemporal classification, the producer's and user's accuracies for the class urban raise 3.79% and 8.17%, respectively. The accuracies for rural residence and bareland are similar in single-date and multitemporal classifications.

The principal component analyses were performed separately on the SPOT 5 image (4 bands) and 4 bands from the ETM+ image (bands 2—5). A combined 5-band database, called PC-8-5 combination, was formed by combining the first three principal components from the SPOT 5 image with the first two principal components from the ETM+ image, based on scores of the principal components. The classification results indicate that the PC-8-5 combination has almost the same accuracy as the SPOT+ETM combination (4-band SPOT and 4-band ETM+ (bands 2-5)), which suggests that principal component analysis can compress most available information to few principal components, which can reduce the data volume and does not significantly lower the classification accuracy.

The datasets including bitemporal SPOT and ETM+ images and the texture images computed on different SPOT 5 bands using variogram and 4-neighbor method were classified. The classification results indicate that the data combination of the bitemporal data and the texture images from SPOT band 1 and band 3 has the highest accuracy among all data combinations. This multitemporal classification incorporating texture images from SPOT bands 1 and 3 obtained accuracy 2.10% higher than the multitemporal classification with spectral information alone. Moreover, the classes with significant texture (variability), such as rural residence, bareland and woodland, have higher accuracies in the classification incorporating the texture than in the classification with spectral information alone (Table 3).

Table 3. Classification accuracy for the combinations of multitemporal data and texture[a)] (Unit: %)

| Data combination | SPOT+ETM6+T(G)+T(NIR) | |
|---|---|---|
| | Producer's accuracy | User's accuracy |
| Rural residence | 69.58 | 97.49 |
| Bareland | 95.26 | 45.26 |
| Cotton | 85.63 | 95.44 |
| Water body | 97.13 | 99.83 |
| Corn | 94.39 | 97.45 |
| Reed | 90.21 | 98.06 |
| Grassland | 82.09 | 88.71 |
| Woodland | 81.12 | 40.49 |
| Orchard | 97.87 | 99.56 |
| Urban | 96.24 | 78.90 |

a) SPOT: 4-band SPOT 5 multispectral image; ETM6: ETM+ bands 1—5 and 7; T(G): Texture from SPOT band 1; T(NIR) Texture from SPOT band 3. Overall accuracy: 89.10; Kappa coefficient: 87.00

It can be found from the above results that when using single-stage ML classification method, each data combination has its own advantage. Even for single-stage classification with multitemporal data, a significant improvement in the classification accuracy can be achieved for some classes. However, the improvement is still marginal for other classes. Therefore, it is desirable to adopt the hierarchical classification (multi-stage classification) strategy to combine all the best combinations and to improve the overall accuracy and the accuracies of individual classes.

The best combinations for individual classes and the sequence of hierarchical classification are shown in Table 4. Table 5 presents the confusion matrix of the hierarchical iterative classification.

Table 4. The sequence of hierarchical classification[a)] (Unit: %)

| Order | Class | Best data combination | Original producer's accuracy/user's accuracy |
|---|---|---|---|
| 1 | Water body | SPOT | 99.30 / 99.89 |
| 2 | Orchard | SPOT+ETM6+T(G)+T(NIR) | 97.87 / 99.56 |
| 3 | Cotton | SPOT | 94.45 / 99.49 |
| 4 | Corn | PC-8-5+T(G)+T(NIR) | 95.96 / 98.76 |
| 5 | Reed | SPOT+ETM4 | 90.04 / 98.54 |
| 6 | Urban | ETM6 | 93.71 / 86.72 |
| 7 | Grassland | SPOT+ETM6+T(G)+T(NIR) | 82.09 / 88.71 |
| 8 | Rural residence | SPOT+T(G)+T(NIR)+T(R)+T(SWIR) | 74.22 / 91.78 |
| | Bare land | | 84.41 / 59.00 |
| | Woodland | | 81.12 / 40.49 |

a) SPOT: 4-band SPOT 5 multispectral image; ETM6: ETM+ bands 1—5 and 7; ETM4: ETM+ bands 2—5; PC-8-5: 5 principal components from SPOT 5 and ETM4; T(G): texture from SPOT band 1; T(R): texture from SPOT band 2; T(NIR): texture from SPOT band 3; T(SWIR): texture from SPOT band 4.

It can be seen from Table 5 that the overall classification accuracy for the hierarchical classification is 93.14%, which is 4.34% higher than the single-stage classification using bitemporal image and texture, 20% higher than the single-date SPOT 5 image. Both the producer's accuracies and the user's accuracies for most classes significantly increase: the producer's accuracies for individual classes range from 82.84% to 99.30%, while the user's accuracies for most classes are higher than 88.45%. The user's accuracy for class bareland is 67.26%, 8.26% higher than that for single-date SPOT image. The only exception is the class woodland having low user's accuracy of 35.14%. This is due to the small size of the woodland and the interplant with corn and cotton, as well as the similar spectral features, so there are many mixed pixels for this class in the image. Considering this situation, higher resolution images can be used to reduce the mixed pixels and improve the classification accuracy.

Table 5. Confusion matrix for hierarchical iterative classification[a)] (Unit: %)

| | Rural residence | Bareland | Cotton | Water body | Corn | Reed | Grassland | Woodland | Orchard | Urban | Producer's accuracy | User's accuracy |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rural residence | 86.89 | 10.44 | 0 | 0 | 0 | 0.02 | 0 | 0 | 0 | 5.89 | 86.89 | 94.12 |
| Bareland | 8.75 | 88.38 | 0.19 | 0.19 | 0.32 | 3.78 | 2.24 | 0.35 | 0 | 0.07 | 88.38 | 67.26 |
| Cotton | 0 | 0.05 | 93.52 | 0 | 0.05 | 0 | 0 | 0.87 | 0 | 0 | 93.52 | 99.69 |
| Water body | 0 | 0 | 0 | 99.30 | 0 | 0 | 0 | 0 | 0 | 0 | 99.30 | 100.00 |
| Corn | 0 | 0 | 2.44 | 0 | 96.56 | 0 | 0 | 2.97 | 0.16 | 0 | 96.56 | 99.03 |
| Reed | 0 | 0 | 0.15 | 0.30 | 0 | 90.61 | 0 | 10.31 | 0.04 | 0 | 90.61 | 98.49 |
| Grassland | 0 | 0 | 0 | 0 | 0 | 0.17 | 82.84 | 0.52 | 0 | 0 | 82.84 | 89.52 |
| Woodland | 0.03 | 0.55 | 3.70 | 0.17 | 3.08 | 5.04 | 14.93 | 84.97 | 0.75 | 0 | 84.97 | 35.14 |
| Orchard | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 99.05 | 0 | 99.05 | 100.00 |
| Urban | 4.34 | 0.59 | 0 | 0.04 | 0 | 0.38 | 0 | 0 | 0 | 94.04 | 94.04 | 88.45 |

a) Overall accuracy: 93.44; Kappa coefficient: 92.18

## 4 Conclusions

The results presented in the paper indicate that the hierarchical iterative classification method is one of the effective methods for achieving highly accurate land cover information. The method proposed in the study combines the temporal information, different resolution and spatial texture information, which uses the best combination for each class and reduces the confusion between different classes, thus improving the overall accuracy and the accuracies of individual classes. The method and strategy proposed in the paper can be easily applied to other applications.

## References


1 Guerschman J. P., Giallorenzi M. C. and Pacin F. Land cover classification in the Argentine Pampas using multi-temporal Landsat TM data. International Journal of Remote Sensing, 2003, 24 (17): 3381—3402.

2 Wolter P. T., Mladenoff D. J., Host G. E. et al. Improved forest classification in the Northern Lake states using multi-temporal Landsat image. Photogrammetric Engineering & Remote Sensing, 1995, 61(9): 1129—1143.

3 Turner M. D. and Congalton R. G. Classification of multi-temporal SPOT-XS satellite data for mapping rice field on a West African floodplain. International Journal of Remote Sensing, 1998, 19(1): 21—41.

4 Song C., Woodcock C. E. and Seto K. C. et al. Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? Remote Sensing of Environment, 2001, 75(2): 230—244.

5 Shaban M. A. and Dikshit O. Improvement of classification in urban areas by the use of textural features: the case study of Lucknow city, Uttar Pradesh. International Journal of Remote Sensing, 2001, 22(4): 565—593.

6 Matheron G. The theory of regionalized variables and its applications. Centre de Morphologie Mathematique de Fontainebleau. 1971, 1—211.

7 Chica-Olmo M. and Abarca-Hernandez F. Computing geostatistical image texture for remotely sensed data classification. Computers & Geosciences, 2000, 26(4): 373—383.

8 Berberoglu S., Lloyd C. D., Atkinson P. M. et al. The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean. Computers & Geosciences, 2000, 26(4): 385—396.