THE STUDIES ON COMPUTER AIDED CLASSIFICATIONS FOR GRASS RESOURCES INVESTIGATIONS IN NORHTERN TIBET BY USING SPACELAB COLUR INFRARED IMAGES

Shu Ning

Wuhan Technical University of Surveying and Mapping
Avenue Lo-Yu, 39 Wuchang
Wuhan CHINA

Commission VII

Abstract

This paper discussed the theories and methods of computer aided classifications for the grasslands investigations by using the Space labcolour infrared images in accordance to the grasslands classification principles and the characteristics of information space used in the vizual interpretations of remote sensing images. It takes the several kind of experiments as the instances of utilizing auxiliary data and image data composite technologies. The resutts have been evoluated by comparing those with the results of experts' interpretations according to the on-the-spot surveying records. It shows us that it is flexible to use this kind of remote sensing data for resources investigations, and one can get the optimum effects by employing landsat MSS biomass index, DTM images, slope and its orientation data.

I. Introduction

Infra-red color photos (scale 1:820,000) taken on the air of Tibet, China, on December 2 1983 with Spacelab 1 metric cameras installed on space shuttles have aroused great interest of a large number of people who are working with photogrammetry or remote-sensing application. It lies in that these photos like aerial photographs are stereoscopic and available for map--making. On the other hand they are images from space remote sensor with higher resolution, so they can provide another novel means for investigating material resources and natural environment in the distant regions. The Department of Aerial Photogrammetry and Remote Sensing of the Wuhan Technical University of Surveying and Mapping has selected an area of a topographic map with a scale of 1:100,000 covered by the image (An area in Bango County) as a test site. The altitude of this region is between 4,555 m and 5,632m . The geomorphologic distribution mainly consists of river wide valleys in plateaus, mountains and hills. There are essential types of grass, such as kobresia littledalei(K), Kobresia sp-Stipa purpurea (K-S), Kobresia pygmaca-Androsace tapete÷Arenaria musciformis (K-A÷A), Stipa purpurea (S), Stipa purpurea -- Leontopodium pusillum + Saussurea humilis + Ortropis mollis + Astragalus tanguticus + Potentilla bifurca (S-L+S÷O÷A+P), Stipa purpurea - Carex stenophylla + Kobresia sp + Androsace tapete (S-C+K+A) etc. For surveying the grassland, visual interpretation is being made simultaneously with computer classifying process, and both are combined. Spacelab images, however, are only infrared photos with single band and without digital tape record. So it is necessary, first of all, to digitalize them. There are two methods for digitalization, one is scanning with color separation and three "bands" will be got, the other is single color scanning, and only one "band" is to be had. Color separation scanning will result in, owing to its having several "bands", automatic classification on the image processing system ARIES-III; whereas single color scanning, because of its having only one band, cannot directly make automatic classification on that system. In general, there are such ways for doing with this single color band as the following:

1. Segmentation.
2. Further with texture analysis.
3. Complex analysis with other image information or auxiliary data.

The segmentation method is actually curving up regions by gray scale in digitals. Textural analysis is local transformation of operators. Concerning the experiment with spacelab images, the latter two methods have been employed for automatic classification, and digital terrain models (DTM) have been introduced, being put on grey scale, are to be used as an "uxiliary band". ---Analysis of the classification results is based on the information from on-the-spot sampling investigation. The Gansu Grassland Ecology Institute has offered, as a coordinated unit of ours, a large amount of on-the-spot information, so the basis for the selection of automatic classification training area and the evaluation of automatic classification results has been provided.

II. Information space. data composition and grasslands classification principles

In the past computer aided classification procedures, we usually take the multispectral data as objectives, and employ a certain kind of classifier. There exist a number of limitations when using this kind of method which purely depends upon the seperation degree of different classes in the spectral space. It can't do any thing when there are different spectral reflections for a same kind of ground feature, or similar spectral reflection for different kinds of ground features.

According to the general regularities of artificial intelligences, one must divide the train of the thoughts and the actions of some experts into several steps when implementing the human's thinkings and behaviours by a certain kind of machine, then realize those steps one by one.

Generally, visual interpretations of remote sensing images are carried out in a multi-dimensional information space, which not only involves grey level data of pixels in images and the coordinates of image systems and ground systems, but also a number of auxiliary informations , referencing documents, and on-the-spot surveying data. Furthermore, one needs different kind of image data (for example, color infrared photographs, MSS images or SPOT images, etc.) , if possible. The types of auxiliary data are more plentiful, for example, soil type distribution map, agricultural climate data, digital terrain model, etc. When implementing the visual interpretation, one must use so many informations for synthetic analysis, logical judgements, etc. But while computer works, all of those informations must be composited, i.e. being made identical in geometry. The digitalization of various auxiliary data is necessary to be done, and their coding must be simplified, or in form of grey level. Then, those data must be composited with image data, and the realization in digit is to ensure the identification of ground coordinates of start point, same ground interval between the neighbour pixels, same number of lines and columns, and identical ground coordinates of every corresponding pixels. In case of the composition of different kind of data, the dimension number of information space is properly increased for the computer processing to implement the many-sided analysis, and to escape from the limitations of only depending on the spectral informations, so raising the "intelligent quotient" of computer processing.

As to the auto-classifications for the grasslands resources investigations by using Spacelab images, one must implement three levels of surveying for the classes, sub-classes and groups in accordance with grasslands classification principles (synthetic sequential approaches). Class is the first level, and has water and heat index as the main factor in consideration, which changes with the variations of longitude, latitude and elevation. Sub-class is the second level, which mainly takes into considerations the geomorphological subareas and soil conditions. Group is the third level, for which one must consider the vegetation conditions, for example, the type of grasses, the superiorities degrees and ground environments. For those reasons, it is necessary to introduce the elevation informations, soil type distribution maps, geomorphological units distribution maps and the image data with higher spectral resolutions. Simultaniously, the enhancements of image data, selections of features, or some kind of transformations must be taken into practice, to realize the grasslands classification principles.

All of the above have been discussed in detail in reference (1), and must be follwed in computer aided classifications for grasslands investigations.

III. Several experiments for use of the Spacelab images

1. Utilization of DTM

The grasslands on the Tibetan Plateau is apparently characterized of vertical distribution. As the temperature gradient in the mean latitude area is 0.6℃ each 100m, it is minus function of the altitude, and humidity is plus function of the altitude, both the temperature and the humidity, constituting the index of the hotness and water, become important parameter of distinguishing the types of grasslands. Hence, it is bound to introduce altitude data. Strictly speaking, altitude data have to be transformed into data of slopes and slope inclination, and put into the automatic classification process as auxiliary data. Because the gradients of temperature and humidity are different on different slope inclinations, there is variation in the vertical distribution of grassland. Work has already been underway.

Simply introducing altitude data will to a certain degree satisfy the requirements of automatic classifications. There are two ways for introducing altitude data. One is to use them directly as auxiliary data without changing its value, but when automatically classifying, there appear a vast mount of data which occupy considerable space of storage. The other is to turn them into grey scale, i.e., in accordance with the difference in height of the regions under study, the altitude is changed to value of grey scale, so as to produce "DTM images" according to formula as follows:

$$Vp = \frac{Ve - Vmin}{Vmax - Vmin} \cdot 255$$

Where Ve represents the altitude on P point. Vmax and Vmin are the maximum and minimum values of the altitude of the regions in question. Vp is the result of the altitude data on P point being transformed into grey scale.

Images, when automatically classifying, have been geometrically corrected. If the origin point of the image grid data is in conformity with that of the "DTM images" (ground coordinates) and DTM is interpolated to a network which corresponds with the image grids being consistent in ground intervals, then "DTM images" will be a composite of the remote sensing images.

"DTM images" can roughly reflect the changes of physical features of the place, such as the run of the mountain ridges and the distribution of river valleys, etc. , just like a "space image". The variation of grey scale is also that of the altitude.

If the DTM images are only compounded with the single band Spacelab images, automatic classification can be made. In employing texture images or compunding with other image data (such as Landsat data) or classifying Spacelab color separation scanning images, it is possible to add to it DTM images. Analysis of the results of classification will be seen later.

The classification results of DTM images and single band Spacelab images show that the distribution of different classes is reasonable and fairly reasonable, particularly, (K-S), (S),(S-C+K+A) and flowing stone slope, their distribution being quite appropriate. The types of grass on the mountains and in the river valleys are completely set apart, giving expresion to the restrictive function of DTM. The main areas of distribution of (K-A+A) and (S-L+S+O+A+P) reveal themselves, only that the area of the former is simply large in one or two places but rather small in very few cases. Class (K) are less visible, while the area of (A) is rather larger, showing the effect of shortage of image spectrum information in winter.

This can be obtained by simply adding to it DTM images, and it illustrates that DTM has a decisive effect on the classification of grasslands on the mountains and plateaus.

As the single band numeralized images are images taken in winter, though various arrangements of grasslands can be roughly distinguished, the information, after all, is not sufficient enough. Here is only an attampt to utilize such kind of space remote sensing materials, it has realized to a certain degree the aims and purpose of investigating grasslands.

2. Classification of color separation scanning images

Look first at the images got from color separation scanning.

Three "bands" are produced from color separation scanning, of them the first and second "bands" give larger amount of information, while the third "band" gives less. It will be found that after the composition of the first and second "bands" the amount of information will be increasing greatly than that of infrared single band. Where there were primarily a number of dark spots in a single band several shades of color can be made out, so it is undoubtedly very useful for classification.

Automatic classification still has DTM added. The result shows that (S) and (K-S) can be better distinguished; class (K) appears much more than that classified in the single band with DTM; (K-A+A) and (S-L+S+O+A+P) are reasonably classified, and so are (S-C+K+A). But (A) appears rather less, while flowing stone slopes become more. Similarly it shows that the spectrum information is insufficient.

To produce images of several "bands", color separation scanning of the infrared band negatives can be made digitalized. It can reflect, to a certain degree, the information of surface features and also can be automatically classified. So it makes up for the shortage of data of the single band. As a substitude method, it has achieved some results in classification of grasslands, and so it seems available for reference.

3. Classification of texture information used for auxiliary information

Texture information is the measurement for space frequency of the grey scale. Part of the texture information represents the linear composition of the original partial information, but on the whole they are not strougly correlated with the original data. So such information can be used as auxiliary data to analyze original data. They can be taken as one "band", the same as the images of DTM.

Such information can be obtained from convolution of local operators. Simple operators have partial mean values, mean spuare errors, partial edges, partial contrast enhancement and partial gradient, etc.

Partial mean values and partial contrast enhancement are used in the experiment. The effect of partical contrast enhancement makes the shades of color and the fragmentary information much richer in comparison with the original data.

Similarly the result of utilizing DTM is essentially the same as that of single band with DTM, and yet on some fragments the shades of color seem much more than that of the former, showing the effect of information increment on the fragments.

4. Composite with LANDSAT images

To resolve the inadequacy of the infrared single band scanning data, it is quite natural to resort to other space remote sensing image materials. At present, the materials available are Landsat images. The Spacelab images are taken in December and the spectram information of the grasslands are quite little. So Landsats MSS images in summer are selected (July, 1983).

The pixel resolution of these two kinds of images is different. The pixel resolution of Spacelab images is 40m while that of Landsat is 80m. It is possible to visual interpretation of each of them, but it is necessary to have identical resolution for surface and to ensure that the corresponding surface objective of each pixel should be the same, if these two kinds of images are to be compounded and automatically classified. Thus it is necessary, when performing geometrical correction, to assure to have the same size of the corresponding surface of the pixel (after correction). Obviously it would be preferable to make correction according to the pixel size of spacelab images in test regions. This method of image composition is extremely beneficial for automatic classification.

In order to select information that there is no mistaking what they are aiming at when images are to be composited, it is essential to have the MSS bands enhanced first and then make classification, but not rest content with adding directly several bands of MSS to them. The method first employed here is to calculate the index of biomass. Other methods feasible are the transformation of principle components and 7/5 bands ratio. In this way one "band" will be obtained respectively, for example, as in the transformation of principle components, only the first component is to be selected, and DTM images are still to be added to after being composited with spacelab images.

It is quite satisfactory to see the results of classification of the composition of biomass indexes, which best coincides with the actual situation. The composite images clearly display the several shades of grasslands, especially (K). Primarily on MSS images there covered by layers of cloud on the northern part of the test region. Biomass indexes having been taken, the interference of clouds disappeared. The parts covered by them, when classifying, are the other two "bands", i.e., Spacelab image and DTM image, and they work, no clouds are seen on the results of classification. However, it is unavoidable when MSS bands are used for classification. So it follows that the technique of image composition is quite significant for making use of those useful data from a large size of image to counteract the interference of partial clouds.

If composited with the principle component images and ratio images, it can also avoid separating types of could. The principle component images having been compounded, class (S-C+K+A) can be distinguished quite well, and (S),(K-S),(K-A+A) and (S-L+S+O+A+P) can also be mostly separated well. Class (K) can be partly distinguished, flowing stone slope better separated, and (A) partly distinguished.

The classification results of the composition with ratio images are close to that of biomass index composition, particularly class (K), (K-S), which are seen mainly from the whole situation of distribution. Class (A) are more separated.

Here mention has to be made of the classification using alone MSS images plus DTM. Because of images in summer, the spectram information of grassland is abundant, class (K), (K-S), (S-C+K+A), (K-A+A) are well distinguished, but flowing stone slopes and class (A) appear rather more. This has something to do with the influence of the spectrum of surface features which play a dominent part. Cloud as a type can be accurately separated. Since there are clouds, quantitative statistics of the area of grasslands is left undone.

5. Classification results by means of division of elevation belts, slope and its orientation data

The vertical distribution of grasslands varies with the variation of slope and its orientations, specially with the later. So the data of slopes and their orientations must be considered in auto-classifications. DTM image data only plays a part of roles, though the classification results are improved by its efficiency. This is because it is treated as a "band" and doesn't have the weight much great. In fact, no weight has been added, and DTM data is only one dimension in the information space. If using

DTM as the dicision factor after employing some classifiers, it will have the weight much great. But the softwares of image processing system couldn't be changed immedially. In order to emphasis the function of DTM, it is necessary to utilize the elevation belts. The boundary surface of vertical distribution of a certain kind of grass is a curved surface or a inclined plane in entirety. But in a local area, it can be considered as a plane. For those reasons, after the division of several regions in terms of natural zones, the controlling of elevation belts can mostly restrict some negative functions of spectral informations in auto-classification. So, in our test area which is a local area, the division of elevation belts and the employment of slope and its orientation data are introduced for the considerations.

Two elevation belts have been determined in the test area. One is from 4700m to 5100m, the other is of 5100m up. When classification will be carried out in a elevation belt, we assign zero for the image data in other belt with the information of the concerned belt remained. The slope and its orientation data are obtained by the calculations based upon the DTM data, and then transformed into the form of grey level, same as the DTM data.

In this case, we have MSS biomass index, Spacelab image, slope and its orientation data, plus DTM image for the classifications. The results show us that the distinguishment of grass groups are more exquisite, more correct and the grasslands distributions on the northern and southern slopes are more distinct. The grass group (K-A+A) growing on the mountains of 5100m up is reasonably classified, and the distributions of stone slope is more correct. On has found that the class of (S-C+K+A) is classified quite well and the grass group (K) and (A) are very satisfactory. All of those tell us that the division of elevation belts and the utilization of slope and its orientation data are very reasonable and very correct.

6. The statistics and analysis for the classification results

After the grasslands investigations, the statistics in area must be carried out for the planning and managements of grasslands. In order to complete the quantitative analysis for assessing the effects resulted from the different auto-classification approaches, the areas of all classes in different kind of result theme files have been calculated and compared with those obtained after visual interpretations and drawing into theme maps by the technologists of Gansu Grasslands Ecolonogy Institute (GGEI) who have fulfiled the on-the-spot sampling surveying. Then, the assessments have been given out according to the synthetic grade methods.

Tab. 1 is the results of the comparisons of statistics. Because of the differences between the classification system of GGEI and ours, the comparisons are carried out in terms of three groups, of which one is the grass groups of (K) as the main group, the second is that of (S) as the main, the third is that of (S-C+K+A) including stone slopes. The unit of area is square kilometer.

	GGEI	(1)SBDS	(2)SBD	(3)SD	(4)SSD	(5)STD
(K)	188.32	198.80 5.7%	171.80 9.3%	191.20 1.5%	155.60 17.4%	164.70 12.5%
(S)	61.44	61.48 0.07%	64.40 4.8%	66.20 7.7%	52.60 14.4%	46.70 24%
(S-C+K+A)	343.24	341.47 0.52%	346.80 2.2%	339.00 1.2%	376.10 9.5%	345.70 1%
mean error		2.05%	5.4%	3.5%	13.8%	12.5%
accuracy in area		97.95%	94.6%	96.5%	86.2%	87.5%
marks for the distributions of all classes		10	9.5	8.2	9.0	8.5
Synthetic marks		0.992	0.948	0.878	0.886	0.860

Tab. 1

The percentages in the first three lines of Tab. 1 are those of the area statistics differences between GGEI and ours to the areas calcutated by GGEI. SBDS expresses the classification results of the composition of Spacelab image (Sp), MSS biomass index (Bi), DTM image (D) plus slope and its orientation data (Sl). SBD represents the results of the composition of (Sp), (Bi),(D). SD is the result of (S) and (D). SSD represents the result of Band 1, Band 2 of the colour seperation scanning images and (D). STD is the effect of (Sp), its texture image and (D). One can see in Tab. 1 that the accuracy on area statistics of SBDS is the highest, that of SBD and SD are of the second level, that of SSD and STD are also of 85% up. Here SD is the second, but the distributions of the classes in the result theme file are not very satisfactory. So, in this case, the assessment only depending upon the area statistics is not completely reasonable. Generally, the reasonableness of the grass group distribution in the theme map must be firstly taken into account for the assessment of grasslands investigation auto-classifications. Based on the reasonableness of distributions, calclate the areas to see how are the accuracies of statistics, then one can get the correct conclusions. For that reason, we have taken the synthetic marks evoluation as the assessing method after the experts of GGEI gave out the marks for all of the results. The sixth line in Tab. 1 is the marks given out by them, which is consistant with the analysis in the above discusions for the various tests. The synthetic mark evoluations are given out according to the formula as follows;

$$Ms = 0.6(g/10) + 0.4a$$

here g represents the marks given out by the experts, a is the accuracy of area calculation, Ms is synthetic marks which can be seen in the last line of Tab. 1. The synthetic marks indicate the reliabilities of area statistics, and accord with the principles of assessment. As the distributions of all grass groups are considered as the most essential condition for evoluations according to the experts, so it has the weight of 0.6. The statistics about auto-classification results have the weight of 0.4 as there are some fortuities during the computer processing. From the synthetic evoluations, we can see that SBDS is the most satisfactory approach, which results in the most reasonable distributions of all classes, and reaches to the highest accuracy of area statistics. That is the effect of efficient considerations for the grasslands classification theories and taking those into practice by doing all one can. SBD takes the second place, which is simpler in processing without the considerations of elevation belts and slope data. However, the distributions of the classes and the area statistics are all satisfactory. SSD comes the third, resulted from colour seperation scanning images, and the grass groups distributions are better than that of SD and STD with a slight improvement. As mentioned above, STD and SD are almost the same in distributions of classes, but here the accuracies of area statistics are different. It can be seen that the area statistics reliabitity of SD is not as great as the accuracy of area calculation, because the qualitative analysis show us that STD is better than SD. Consequently, the synthetic marks are more reasonable.

The last three columns are the results of Spacelab images and DTM without other kind of images. As to the utilizations of Spacelab images, now that the images in winter can have the reliability of 85% up to the auto-classification effects, if there were the images in summer, it could be more satisfactory. For that reason, the utilizations of this kind of images can meet the needs for the resources investigations.

IV. Conclusion

Spacelab colour infrared digital image as one kind of space remote sensing image offers us a new kind of meterial for the resources investigations in distant regions. When using this kind of images for the computer classification processings, one has to resolve the problem of single band, and utilize various kind of informations and data composite technologies to ensure the sufficient informations in auto-classifications. In accordance to the classification principles for the grass resources investigations, the introduction of DTM can make auto-classification reliabilities reaching to 85% up. The utilizations of texture information (from the view of grass group distribution marks) or the method of colour seperation scanning for digitalizations can improve the classifying accuracies. If there were the images in summer, it could be prodicted to obtain better effects. The compositions with Landsat images combine the higher resolution in spectrum with the higher resolution in space of Spacelab images. As to the selection of features, biomass index looks like a good tool. Specially, when taking into considerations the slope, its orientation data and different elevation belts, the accuracy of classifications has been much improved. Those facts show us the important significance of data composite technologies and taking into practice the grasslands classification principles, and the flexibilities of making computer aided classifications more intelligent or realizing the experts' thinkings.

Reference:

(1) Shu Ning: "The realizations of the grasslands classification principles in computer processing for the grasslands resources investigations", the 16th International Congress of Photogrammetry and Remote Sensing, 1988.

(2) "The manual on grassland jobs", edited by the Grasslands Department of Gansu Agricultural University, 1979.