INT. J. REMOTE SENSING, 10 SEPTEMBER, 2004,
VOL. 25, NO. 17, 3529–3539

Data fusion and multisource image classification

D. AMARSAIKHAN† and T. DOUGLAS*‡

†Department of Geoinformatics, Institute of Informatics and RS, Mongolian Academy of Sciences, Ulaanbaatar-51, Mongolia;
e-mail: amarsaikhan_d@arvis.ac.mn
‡School of Applied Sciences, University of Northumbria, Newcastle upon Tyne, England, UK; e-mail: terry.douglas@northumbria.ac.uk

(*Received 21 November 2001; in final form 12 March 2003*)

Abstract. The aim of this study is to explore different data fusion techniques and compare the performances of a standard supervised classification and expert classification. For the supervised classification, different feature extraction approaches are used. To increase the reliability of the classification, different threshold values are determined and fuzzy convolutions are applied. For the expert classification, a set of rules is determined and a hierarchical decision tree is created. Overall, the research indicates that multisource information can significantly improve the interpretation and classification of land cover types and the expert classification is a powerful tool in the production of a reliable land cover map.

1. Introduction

Data fusion and multisource image classification are of increasing interest in the present development of digital image processing. Remote sensing images taken in the optical range of the electromagnetic spectrum contain information on the reflective and emissive characteristics of the Earth surface features, while the Synthetic Aperture Radar (SAR) images contain information on the surface roughness, texture and dielectric properties of the natural and human objects. It is evident that a combined use of the optical and SAR images will have a number of advantages because a specific object or class that is not seen on the passive sensor image might be seen on the active sensor image and vice versa because of the complementary information provided by the two datasets (Amarsaikhan *et al*. 2000, 2001).

Over the years, integrated approaches employing both optical and radar images have been widely used for interpretation and enhancement of different surface features. Many authors have proposed and applied different techniques to combine optical and SAR images in order to enhance various features and they all judged that the results from the fused images were better than the results obtained from the

* Author to whom correspondence should be addressed.

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0143116031000115111

individual images (Harris *et al.* 1990, Yėsou *et al.* 1993, Wang *et al.* 1995, Pohl and Van Genderen 1998, Prakash *et al.* 2001, Ricchetti 2001).

For several decades, digital methods of classification of remote sensing images have been effectively used for land cover mapping. The early methods mainly involved supervised and unsupervised methods and hence many techniques have been developed (Amarsaikhan and Ganzorig 1997). Multisource data have proved to offer better potential for discriminating between different land cover types. A number of authors have assessed the potential of multisource images for the classification of land cover classes (Srinivasan and Richards 1990, Munechika *et al.* 1993, Serpico and Roli 1995, Benediktsson *et al.* 1997, Hegarat-Mascle *et al.* 2000). In remote sensing applications, the most widely used multisource classification techniques are statistical methods including Dempster–Shafer theory of evidence and neural networks (Solberg *et al.* 1996). Any of these techniques can include different rules to involve human expertise in the final decision-making to increase the classification accuracy (Amarsaikhan *et al.* 2000).

The aim of this research is (a) to compare different data fusion techniques for the enhancement of spectral variations of urban and non-urban features, later to be used for training sample formation, (b) to compare the results of a standard supervised classification method (maximum likelihood classifier (MLC)) and expert classification. For the supervised classification, different feature extraction approaches have been applied. For the actual analysis, SPOT-XS and ERS-2 SAR images of the urban area in Mongolia have been used. The analysis was carried out using PC-based ERDAS Imagine 8.4.

2. Study area and data sources

The central part of Ulaanbaatar, the capital city of Mongolia was selected as a test site. In order to perform efficient data fusion and avoid radar geometric distortions, a flat area was chosen. The area is about 9.5 km × 5 km and is characterized by such classes as urban (buildings and gers—the Mongolian national house), vegetation, soil and water.

The data used consisted of a SPOT XS image of 19 June 1997 and an ERS-2 SAR image of 25 September 1997. The SAR image is a 16 bit PRI (precision image) product with 12.5 m pixel spacing and ground resolution of about 25 m. In addition, a topographic map of 1984, scale 1 : 50 000, and a general urban planning map were available. Figure 1 shows the target area in a SPOT XS image and some examples of its land cover.

3. Geometric correction of the multisensor images

In order to perform successful data fusion, a high geometric accuracy and good geometric correlation between the datasets are needed. Initially, the SPOT image was geometrically corrected to a UTM map projection using a topographic map of the study area, scale 1 : 50 000. The ground control points (GCPs) have been selected on clearly delineated crossings of roads, streets and other clear sites. In total, 12 points were selected. For the transformation, a second-order transformation and nearest-neighbour resampling approach were applied and the related root mean square (rms) error was 0.54 pixel.

In order to correct the SAR image, 18 more regularly distributed GCPs were selected comparing the locations of the selected points with other information such as the SPOT XS image and the topographic map. The image was geometrically corrected to a UTM map projection using the topographic map of the study area.

Figure 1. SPOT XS image of the central part of Ulaanbaatar (red = band 3, green = band 2, blue = band 1). The size of the displayed area is about 9.5 km × 5 km.

For the actual transformation, a third-order transformation was used. As a resampling technique, the nearest-neighbour resampling approach was applied and the related rms error was 0.98 pixel.

4. Speckle suppression of the SAR image

The SAR images have a granular appearance due to the speckle formed as a result of the coherent radiation used for radar systems; the reduction of the speckle is a very important step in further analysis. The analysis of the radar images must be based on the techniques that remove the speckle effects while considering the intrinsic texture of the image frame (Ulaby *et al*. 1986). In this study, four different speckle suppression techniques such as local region, lee-sigma, frost and gamma-map filters (ERDAS 1999) of 3 × 3, 5 × 5 and 7 × 7 sizes were compared in terms of delineation of urban features and texture information. The 3 × 3 gamma-map filter created the best image in terms of delineation of different features as well as preserving content of texture information. In the output image, speckle noise was significantly reduced with very low degradation of the textural information.

5. Data fusion

After applying error correction and georeferencing procedures, data from multiple sources can easily be integrated. The image fusion is the integration of different digital images in order to create a new image and obtain more information than can be separately derived from any of them (Pohl and Van Genderen 1998, Ricchetti 2001). In the case of the present study, for the urban areas, the radar image provides structural information about buildings and street alignment due to the corner reflector effect, while the optical image provides the information about the spectral variations of the urban features.

Image fusion can be performed at pixel, feature and decision levels (Abidi and Gonzalez 1992, Pohl and Van Genderen 1998). In this study, data fusion has been performed at a pixel level and the following techniques were applied: (a) intensity–hue–saturation (IHS) transformation, (b) sum normalization, (c) band ratioing,

(d) principal component analysis (PCA). Each of these techniques is discussed in detail below.

IHS: The IHS method is the most widely used data fusion technique. This method assumes that the H and S components contain the spectral information, while the I component represents the spatial information (Pohl and Van Genderen 1998, Mather 1999, Ricchetti 2001). A detailed review of this approach is given in Mather (1999). For the IHS transformation, bands 2 and 3 as well as the SAR image have been used and the latter was considered as the I. When the IHS image was transformed back to the RGB colour space, contrast stretching has been performed to the H and S channels.

Sum normalization: In this method, each of the sum-normalized multispectral bands (B_1, B_2 and B_3) was multiplied by the SAR image as follows:

$$\begin{aligned} B_1' &= (B_1/(B_1+B_2+B_3))\,\mathrm{SAR} \\ B_2' &= (B_2/(B_1+B_2+B_3))\,\mathrm{SAR} \\ B_3' &= (B_3/(B_1+B_2+B_3))\,\mathrm{SAR} \end{aligned} \tag{1}$$

Band ratioing: In this method, a ratio between the SAR image and each of the SPOT XS bands (B_1, B_2 and B_3) has been defined and then multiplied by a compensation factor of 127 to keep the data range between 0 and 255. It can be formulated as follows:

$$\begin{aligned} B_1' &= (\mathrm{SAR}/B_1)\,127 \\ B_2' &= (\mathrm{SAR}/B_2)\,127 \\ B_3' &= (\mathrm{SAR}/B_3)\,127 \end{aligned} \tag{2}$$

PCA: The most common understanding of the PCA is that it is a data compression technique used to reduce the dimensionality of the multidimensional datasets (Richards 1993). It is also helpful for image encoding, enhancement, change detection and multitemporal dimensionality (Pohl and Van Genderen 1998). PCA has been performed using all available bands and the results are shown in table 1. As can be seen from table 1, PC1 and PC3 are dominated by the variance of the optical bands, while in the PC2 the SAR image has a very high negative loading. Moreover, it can be seen from table 1 that the first three PCs contain 98% of the total variance. The inspection of PC4 indicated that it contained noise from the total dataset.

In order to obtain a good colour image that can illustrate spectral and spatial variations of the classes in the selected image frame, different band combinations

Table 1. Principal component coefficients from three bands of the SPOT XS and ERS-2 SAR.

	PC1	PC2	PC3	PC4
XS1	0.68	−0.05	−0.22	−0.69
XS2	0.66	0.01	−0.19	0.72
XS3	0.29	0.30	0.90	−0.02
SAR	0.06	−0.95	0.29	0.03
Variance (%)	56	26	16	2

have been used. Most of the methods created good images on the basis of visual interpretation. Specifically, the colour image created by the RGB components transformed back from the IHS method showed a better result than many other combinations. The combinations of the normalized and ratio bands created the outputs that had similar interpretation, however the ratio image filtered by a 3×3 edge-enhancement filter gave a superior image in terms of the spatial separation between different objects and classes. The PC image demonstrated very much the characteristics of the radar image because of the high negative loading of the radar image on PC2. Figure 2 shows the comparison of the images obtained by different fusion methods.

6. Classification of the images

6.1. *Feature determination and supervised classification*

In order to have the same range of data used for the classification, initially the 16 bit SAR image was transformed to an 8-bit image. Then, a 3×3 average filtering was applied to the 8-bit SAR image to increase the spatial homogeneity of the data. From the (enhanced) fused images, one to four areas of interest (AOI) representing the five selected classes (built-up area, ger area, vegetation, soil, and water) have been selected through thorough analysis using a polygon-based approach. For the classification the following input bands (features) were used:

1. The original spectral bands of the SPOT XS.
2. The SPOT XS and ERS-2 SAR.
3. The median-filtered SPOT XS and ERS-2 SAR (because median filters remove random noise).
4. The first three PCs of the PCA (here, PCA was performed using SPOT XS

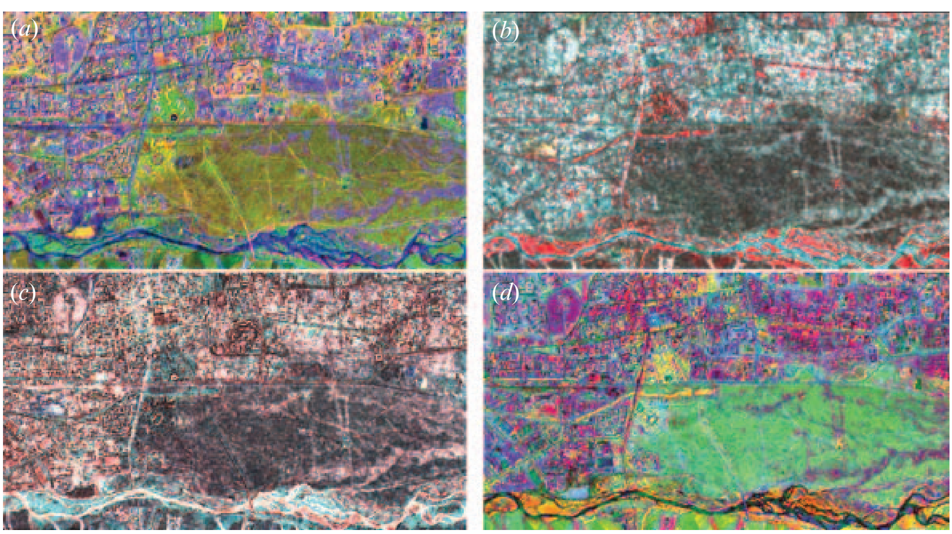

Figure 2. Comparison of the fused images: (*a*) the enhanced through IHS method RGB image; (*b*) a false colour composite (FCC) image created by a combination of the sum-normalized multispectral bands multiplied by the SAR image; (*c*) a FCC image created by a combination of the ratio images; (*d*) PC image (red = PC1, green = PC2, blue = PC3).

and the average filtered 8-bit SAR images (because average filters reduce random noise)).

5. Multiple bands including the original (SPOT XS and SAR) and eight other derivative bands (i.e. PC1, outputs of 5×5 high-pass and 5×5 Wallis adaptive filters (ERDAS 1999), four ratio bands (B_3/B_1, B_3/B_2, SAR/B_2 and $B_3\times SAR/B_1$) and one sum-normalized band ($B_3\times SAR/(B_1+B_2+B_3)$)).

For each of these combinations, the (1–4) training samples were selected on the basis of the previously defined AOI. The separability of the training signatures was firstly checked in feature space and then evaluated using transformed divergence (TD) (Richards 1993). Then the samples that demonstrated the greatest separability were chosen to form the final signatures. The final signatures included about 112–224 pixels. The plot of the mean values for the chosen signatures in the selected bands are shown in figure 3. For the actual classification the standard statistical MLC has been used assuming that the training samples have the Gaussian distribution (Richards 1993).

To increase the reliability of the classification, threshold values were determined assuming that 5% of the overall pixels were misclassified. To perform this, a confidence level (0.05), which is the percentage of the pixels that are believed to be misclassified, was determined for each class and the related χ^2 (chi-square) parameters were determined, thus cutting the tail of the training data histogram. Based on the size of the areas created by the misclassified pixels, window sizes for the fuzzy convolution have been selected and the most appropriate sizes were 3×3 and 5×5 size windows. The fuzzy convolution creates a thematic layer by calculating the total weighted inverse distance of all the classes in a determined window of pixels and assigning the centre pixel the class with the largest total inverse distance summed over the entire set of fuzzy classification layers, i.e. classes

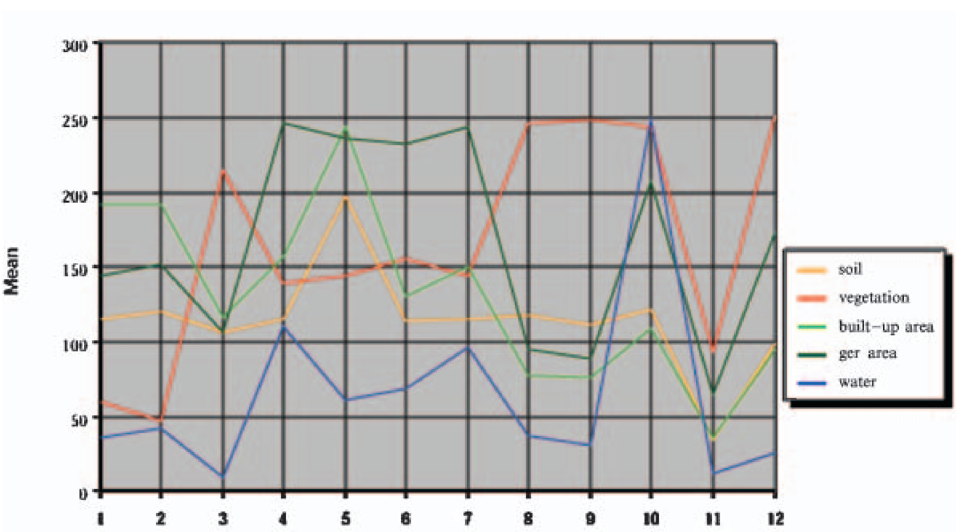

Figure 3. The plot of the mean values for the selected classes. Along the x-axes: 1–4 = SPOT XS and SAR bands; 5 = PC1; 6 = 5×5 high-pass filtered band; 7 = 5×5 Wallis adaptive filtered band; 8 = B_3/B_1; 9 = B_3/B_2; 10 = SAR/B_2; 11 = sum-normalized band; 12 = $B_3\times SAR/B_1$.

with a very small distance value will remain unchanged while the classes with higher distance values might change to a neighbouring value if there are a sufficient number of neighbouring pixels with class values and small corresponding distance values (ERDAS 1999). The visual inspection of the fuzzy convolved images indicated that there are some improvements on the borders of the neighbouring classes that significantly influence the separation of the decision boundaries in multidimensional feature space.

The images classified by each of the above-mentioned feature determination approaches are shown in figure 4(*a–e*). As shown in figure 4(*a–e*), there is a high overlap between two urban classes (built-up area and ger area) on the image classified using only optical bands, but this overlap decreases on other images for the classification of which SAR and optical bands as well as fused images have been used. Although the multisource datasets produced a better result than the single source image, it could not separate urban classes in the eastern part of the city. An attempt to separate these classes will be made in the expert classification section.

To verify the accuracy of the final MLC, the overall performance was used. This approach creates a confusion matrix in which reference pixels are compared with

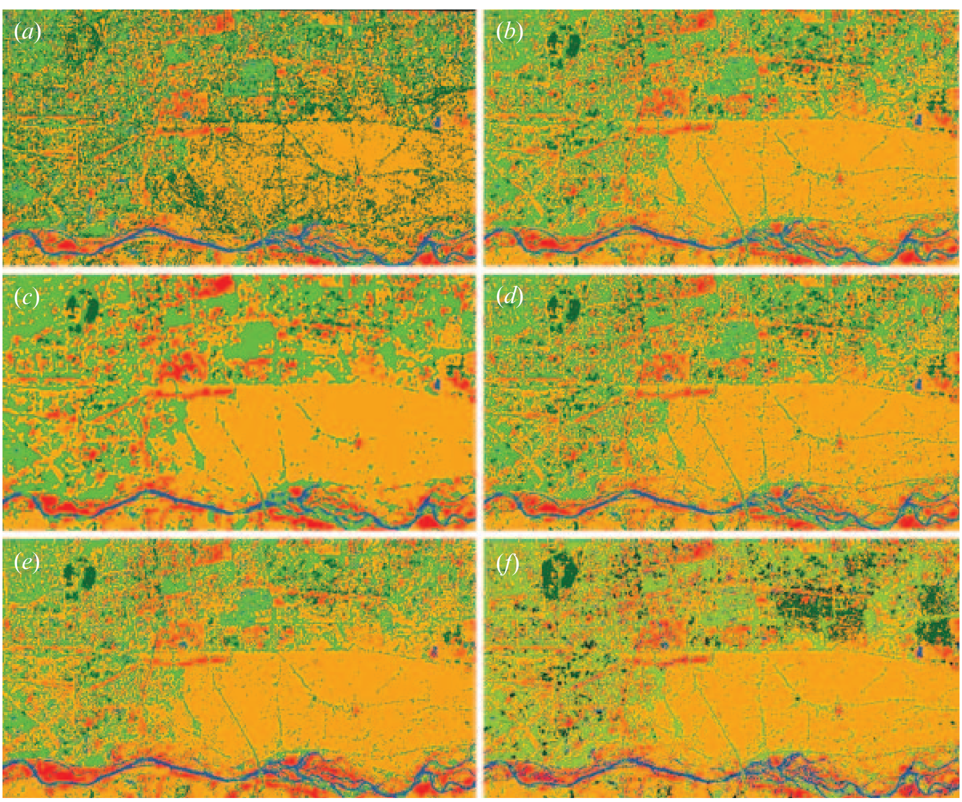

Figure 4. Comparison of the classification results for the selected classes (soil—orange, vegetation—red, built-up area—light green, ger area—dark green, water—blue). Classified images (*a*) using SPOT XS bands, (*b*) using SPOT XS and ERS-2 SAR bands, (*c*) using median-filtered SPOT XS and ERS-2 SAR, (*d*) using the first three PCs, (*e*) using multiple bands, and (*f*) using expert classification.

the classified pixels and as a result an accuracy report is generated indicating the percentages of the overall accuracy (ERDAS 1999). The stratified random pixels (i.e. more pixels are selected for the evaluation of the larger classes such as built-up area, ger area and soil classes than the smaller classes such as vegetation and water classes) were selected as the potential sample points and overall 60 points were determined. The overall classification accuracies for the selected classes are shown in table 2.

6.2. *Expert classification*

Application of a knowledge-based approach has more and more usage in the digital classification of remote sensing images. The knowledge in image classification can be represented in different forms depending on the type of knowledge and necessity of its usage (Amarsaikhan and Ganzorig 1999). The most commonly used techniques for knowledge representation are a rule-based approach and neural network classification (Amarsaikhan *et al.* 2000).

Expert classification uses a rule-based approach in which a hierarchy of rules or a decision tree describing the conditions under which a set of low-level primary objects gets abstracted into a set of the high-level object classes. The primary objects contain the user-defined variables and include geographical objects represented in both raster and vector structures, external programmes, scalars and spatial models (ERDAS 1999).

It is evident that multisensor images integrated with other ancillary datasets or thematic information will significantly improve the expert classification. However, within the framework of this study, in order to make a comparison with other results obtained through the use of the supervised method, only the satellite images were used for the expert classification. Based on the knowledge about spectral and scattering characteristics of the selected five classes and other intermediate classification results, a set of rules (mainly constraints on spatial and spectral parameters and threshold values) were determined and a hierarchical decision tree was created. The image classified by this method is shown in figure 4(*f*). As seen from the image, the rules to extract specifically the ger area information which was very difficult in the MLC produced a good result. The overall classification accuracy has been evaluated using the same number (60) of sample points as in the previous classification, and it demonstrated an improvement to 91.53%. A general diagram of the expert classification is shown in figure 5.

7. Conclusions

The studies on data fusion and multisource image classification were carried out using SPOT XS and ERS-2 SAR images. For the data fusion, the IHS, sum normalization, band ratioing and PCA were carried out. The band combinations

Table 2. The overall classification accuracy of the classified images.

The bands (features) used for the MLC	Overall accuracy (%)
SPOT XS (three bands)	72.08
SPOT XS and ERS-2 SAR (four bands)	86.67
The median-filtered SPOT XS and ERS-2 SAR (four bands)	87.92
The first three PCs of the PCA	86.67
Multiple bands	87.26

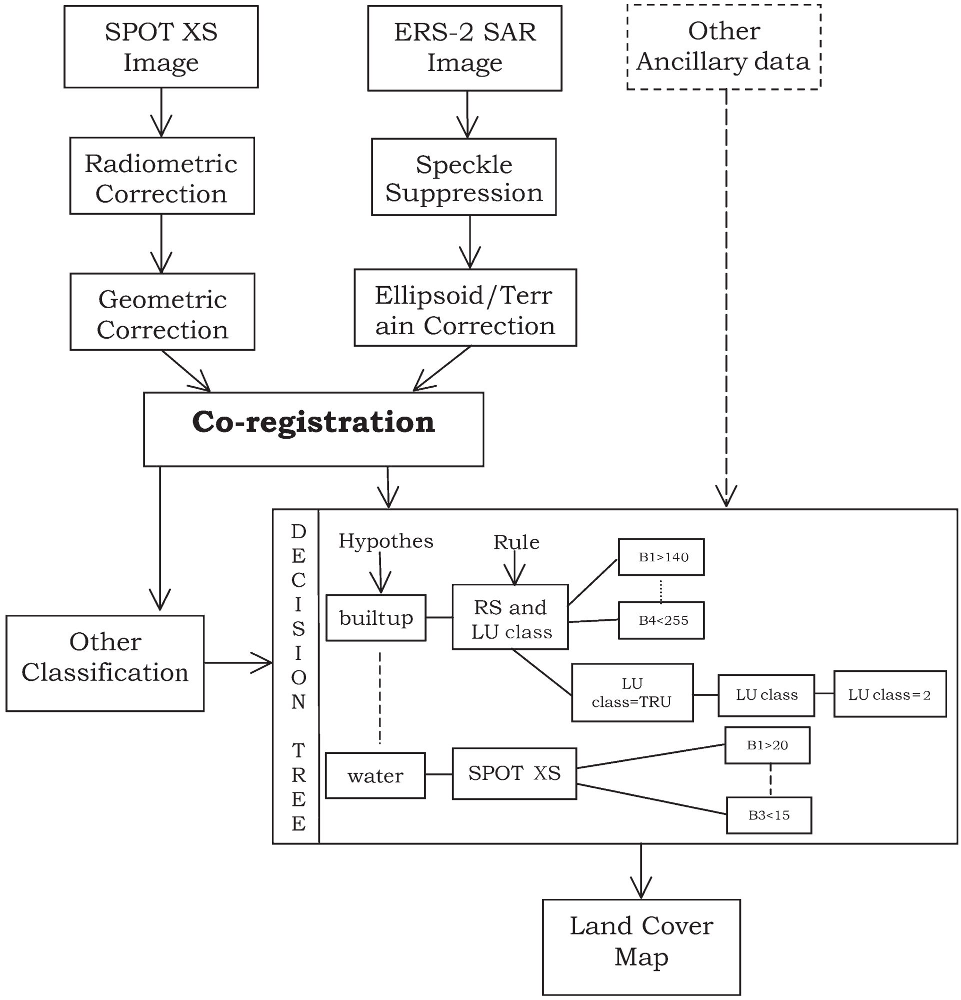

Figure 5. A general diagram for the expert classification.

of either of these methods created good images in terms of delineation of spectral and spatial variations of different objects, but the image created by the RGB components transformed back from the IHS method demonstrated a better visual result than many other combinations.

For the multisource classification, a standard MLC and expert classification were compared. For the MLC, different bands (original and derivative) were used and to increase the reliability of the classification, different threshold values and fuzzy convolutions were applied. For the expert classification, a set of rules was determined and a hierarchical decision tree was created. When the results of the classifications were compared, the expert classification resulted in a higher accuracy.

Overall, the study demonstrated that multisource information can significantly improve the interpretation and classification of land cover types and the expert classification is a powerful tool to produce a reliable land cover map. In addition, it could be seen that the method can produce a map that is ready-to-use for decision-making, when there is a set of sophisticated rules applied to multisource datasets in which optical and radar images as well as other ancillary information are integrated.

Acknowledgment

The authors would like to acknowledge the Royal Society in awarding Dr D. Amarsaikhan funding to participate in an academic exchange at the University of Northumbria.

References

Abidi, M. A., and Gonzalez, R. C., 1992, *Data Fusion in Robotics and Machine Intelligence* (New York: Academic Press).

Amarsaikhan, D., and Ganzorig, M., 1997, Comparison of different pattern recognition techniques. Scientific Papers of Informatics Centre, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia, pp. 71–75.

Amarsaikhan, D., and Ganzorig, M., 1999, Different approaches in feature extraction for hyperspectral image classification. *Proceedings of the 20th Asian Conference on RS, Hong Kong, China, 22–26 November 1999* (Ulaanbaatar: Mongolian Academy of Sciences), pp. 434–438.

Amarsaikhan, D., Ganzorig, M., and Enkhtuvshin, B., 2000, Application of spectral and scattering knowledge for interpretation of active and passive sensor data. Institute of Informatics and RS, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia.

Amarsaikhan, D., Ganzorig, M., and Batbayar, G., 2001, Forest ecosystem change study using active and passive sensor data. *Proceedings of the Open Symposium on Change and Sustainability of Pastoral Land Use Systems in Temperate and Central Asia, Ulaanbaatar, Mongolia* (Ulaanbaatar: Mongolian Academy of Sciences), pp. 74–80.

Benediktsson, J. A., Sveinsson, J. R., Atkinson, P. M., and Tatnali, A., 1997, Feature extraction for multisource data classification with artificial neural networks. *International Journal of Remote Sensing*, **18**, 727–740.

ERDAS 1999, *Field Guide*, 5th edn (Atlanta, Georgia: ERDAS, Inc.).

Harris, J. R., Murray, R., and Hirose, T., 1990, IHS transform for the integration of radar imagery with other remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, **56**, 1631–1641.

Hegarat-Mascle, S. L., Quesney, A., Vidal-Madjar, D., Taconet, O., Normand, M., and Loumagne, 2000, Land cover discrimination from mutitemporal ERS images and multispectral Landsat images: a study case in an agricultural area in France. *International Journal of Remote Sensing*, **21**, 435–456.

Mather, P. M., 1999, *Computer Processing of Remotely-sensed Images: an Introduction*, 2nd edn (Chichester: John Wiley & Sons).

Munechika, C. K., Warnick, J. S., Salvaggio, C., and Schott, J. R., 1993, Resolution enhancement of multispectral image data to improve classification accuracy. *Photogrammetric Engineering and Remote Sensing*, **59**, 67–72.

Pohl, C., and Van Genderen, J. L., 1998, Multisensor image fusion in remote sensing: concepts, methods and applications. *International Journal of Remote Sensing*, **19**, 823–854.

Prakash, A., Fielding, E. J., Gens, R., Van Genderen, J. L., and Evans, D. L., 2001, Data fusion for investigating land subsidence and coal fire hazards in a coal mining area. *International Journal of Remote Sensing*, **22**, 921–932.

Ricchetti, E., 2001, Visible-infrared and radar imagery fusion for geological application: a new approach using DEM and sun-illumination model. *International Journal of Remote Sensing*, **22**, 2219–2230.

Richards, J. A., 1993, *Remote Sensing Digital Image Analysis—An Introduction*, 2nd edn (Berlin: Springer-Verlag).

Serpico, S. B., and Roli, F., 1995, Classification of multisensor remote sensing images by structural neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 562–578.

Solberg, A. H. S., Taxt, T., and Jain, A. K., 1996, A Markov random field model for classification of multisource satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **34**, 100–112.

Srinivasan, A., and Richards, J. A., 1990, Knowledge-based techniques for multisource classification. *International Journal of Remote Sensing*, **11**, 505–525.

Ulaby, F. T., Kouyate, F., Brisco, B., and Williams, T. H. L., 1986, Textural information in SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, **24**, 235–245.

Wang, Y., Koopmans, B. N., and Pohl, C., 1995, The 1995 flood in the Netherlands monitored from space—a multisensor approach. *International Journal of Remote Sensing*, **16**, 2735–2739.

Yėsou, H., Besnus, Y., Rolet, J., and Pion, J. C., 1993, Merging Seasat and SPOT imagery for the study of geologic structures in a temperate agricultural region. *Remote Sensing of Environment*, **43**, 265–280.