ASSESSING RESULTS OF SATELLITE IMAGE CLASSIFICATION: THINGS TO BE OMITTED

Pham Bach Viet[1], **Lam Dao Nguyen**[2] **and Akiko Tanaka**[3]

[1]Deparment of Cartography-Remote Sensing-GIS, Faculty of Geography, HCMC USSH, 10-12 Dinh Tien Hoang Str. Dist 1, HCMC, Vietnam. Email: phambachviet@gmail.com
[2]GIS and Remote Sensing Research Center - HCMC Institute of Resource Geography - VAST, Vietnam
[3]Geological Survey of Japan, AIST Tsukuba Central 7, Japan

ABSTRACT

Assessment of classification results is a critical task, it is an obligation as using remotely sensed data for land cover/ land use classification. Results of assessment normally include of parameters as overall accuracy and Kappa that are sufficient to be evaluate accuracy of classification in normal cases. However, in cases of different types of imagery or different classification techniques, it is required that statistical analysis has to be paid much more attention. This report present a case of using images of Landsat, SPOT and ALOS PALSAR in Ca Mau area for classifying independently each other and fusion between Landsat, SPOT and ALOS PALSAR. The accuracy of outcomes is assessed with more parameters than usual and the analysis results show that if important assessment parameters were omitted it could lead to an unreasonable conclusion of the classification results.

Key words: Accuracy assessment, Ca Mau, Mangrove, SAR fused-Optic images..

1. INTRODUCTION

General applications of remote sensing technique are focused on using satellite imagery to classify land use/ land cover, its results are depended on types and characteristics of data and other factors from the user side such as classification technique, training data sampling. Optical and radar images are the two types of remote sensing data, whose the features have their own advantages; optical and radar imagery can be fused together to enhance the information. This fusion technique has been applied in many fields and on various objects, including of identification of land cover types and mangroves (Claudio, 2000; Hussin, 2007; Sheoran *et al.*, 2009).

Outcomes of the classification are evaluated by using technique of accuracy assessment. Normally, error matrix is taken with known parameters of accuracy and KHAT analysis for cases of using one type of imagery, e.g., optical or radar, similar spatial and spectral resolution. In cases of different data, accuracy assessment requires evaluation with more parameters. Accuracy assessment in classification using remotely sensed data is widely applied, which had been considered under different points of view in terms of statistics, synthetically used with more complex techniques in practice (Congalton, 1991; Congalton and Kass, 2008). Accuracy calculation is not a simple task for comparison and evaluation in case of different areas, images types or classification techniques.

This report focuses on comparison of performance of optical image and radar SAR fused-optic images in land cover classification. Landsat 5-TM, SPOT 5 and ALOS PALSAR

imagery are taken to examine the performance on classifying land cover, mangrove types over the area. Each type of data is independently classified and each of optical images is fused with radar SAR data to compare classification results of prior and after fusion. Study site covers approximately 43,000 ha; it is a part of Ca Mau province, southernmost of the Mekong delta, Viet Nam. Land cover of this area is characterized of a coastal zone with mangroves, shrimp farms mixed with mangroves, crops and built-up. Mangroves are in dense, medium, open forests and shrubs (Figure 1: right image is band ratio B4/B3 of Landsat, mangroves are in white to bright grey, water is in black to dark grey).

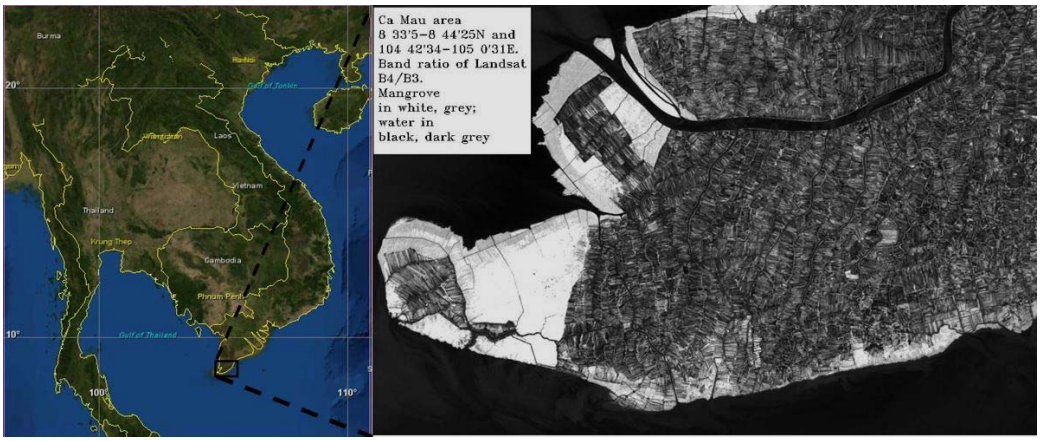

Figure 1. Study area ($8^0$33'5 - $8^0$44'25 N and $104^0$42'34 - $105^0$0'31 E)

2. METHOD

2.1 Data

Landsat 5 TM and SPOT 5 data are used in this study. Radar SAR image of ALOS PALSAR (hereafter abbreviated SAR) with wavelength of band L (23cm) known to be appropriate for forest study in this case, land cover mostly is forests. Used and processed images are as the follows.

(Imagery, Acquired date, Mode/ Polarization, Spatial resolution)

1. SPOT 5, 11/01/2009, XS - 10m
2. Landsat 5 TM, 14/01/2009, Multispectral - 30m
3. ALOS PALSAR, 05/01/2010, 20/02/2010, HH, 12.5m (pixel size 6.25m)
   08/10/2010, HH and HV, 25m (pixel size 12.5m)
4. SPOT fused with ALOS PALSAR, pixel size are resampled at 10m
5. Landsat fused with ALOS PALSAR, pixel size are resampled at 10m

2.2 Method

Total number of land cover types classified is 12 categories with seven statuses of mangroves from open to dense, contiguous and sparse distribution, including shrimp farms mixed inside forest patches, three categories of shrimp farms, one of built-up land and one of water bodies (streams). Samples of training data are taken from field survey, distributed over the study area and these are used for all cases of prior and after image fusion. Supervised classification technique is applied for both cases of fusion and not. Classification of Landsat

and SPOT images are applied at its original spatial resolution. In case of fusion, pixel sizes of all images are resampled at 10 m to yield the equal size.

Samples of validation data are randomly distributed with each type approximately 30 locations of sampling used for accuracy check. Error matrix (or confusion matrix) is used to assess classification accuracy, by statistical calculation of the matrix to yield overall accuracy and Kappa coefficient (Congalton and Kass, 2008). Error matrix is a cross table, constructed based on reference samples. The value of Kappa, or KHAT statistics is defined by the subtraction of the chance of agreement from the observed accuracy divided by 1 minus the chance of agreement, it indicates the agreement degree between referent and classification data in the error matrix. In case if the two classification results need to be compared, Kappa variances, z-test and Margfit analysis are applied (Congalton and Kass, 2008). Margfit analysis is a technique of standardization of value in the matrix, that yield a sum of each row and column equal 1; total sum of column (reference data) and that of row (classification data) is equal number of classified categories (Congalton and Kass, 2008). This is done for testing if two independent error matrices are significantly different, applied for comparison assessment of images in pair, prior and after fusion (Case 1: Landsat and Landsat + SAR; Case 2: SPOT and SPOT + SAR).

3. RESULTS AND DISCUSSION

3.1 Overall accuracy and Kappa coefficient (K) of classification results

Table 1. Error matrix for Landsat classification prior to fusion

Classes	**Reference**												**Total**
	# 1	**# 2**	**# 3**	**# 4**	**# 5**	**# 6**	**# 7**	**# 8**	**# 9**	**# 10**	**# 11**	**# 12**	
# 1	**32**	2											34
# 2		**21**	3										24
# 3			**23**	2									25
# 4				**24**	8								32
# 5					**19**								19
# 6	3					**25**	5						33
# 7						1	**29**		6				36
# 8								**10**			11		21
# 9								8	**24**	4			36
# 10								2	1	**20**			23
# 11											**27**	1	28
# 12												**26**	26
Total	35	23	26	26	27	26	34	20	31	24	27	27	337

Overall accuracy = 83 % Kappa coefficient = 0.8151

Cross table 1 and table 2 present results of accuracy assessment by using classification result in row and reference data in column to calculation related parameters. Twelve categories are classified and checked for classification errors. In brief, user accuracy, producer accuracy, omission and commission error are not presented here as these are relevant to accuracy of categories. Mangroves are classified as open, medium and dense forests based on cover rate. Number 1 and 2 are medium forests; number 3, 4, and 5 are medium to dense forests; number 6 and 7 are open forests mixed with shrimp farms (forests are dominant). Number 8, 9 and 10 are shrimp farms mixed with mangroves at different cover rate (shrimp farms are dominant); number 11 is built-up land and number 12 is water body (streams).

Based on the calculation from the Table 1 and 2 above, this accuracy assessment could conclude that classification of the Landsat image prior to fusion had a lower accuracy than that of the SAR fused-Landsat image because the former overall accuracy and Kappa coefficient were lower than the latter.

Table 2. Error matrix for image fusion classification: Landsat + SAR

Classes	**Reference**												**Total**
	# 1	**# 2**	**# 3**	**# 4**	**# 5**	**# 6**	**# 7**	**# 8**	**# 9**	**# 10**	**# 11**	**# 12**	
# 1	**299**	2									1		301
# 2	5	**205**	21										231
# 3			**187**	20									207
# 4			21	**190**	57								268
# 5				24	**186**								210
# 6	18		8			**233**	62		1				322
# 7	6		7			11	**242**		32	2	2		302
# 8								**129**	5		48	3	185
# 9							2	8	**233**	22		3	268
# 10									8	**192**	8		208
# 11								9			**283**	3	295
# 12								34				**228**	262
Total	328	207	244	234	243	244	306	180	279	216	342	237	3060

Overall accuracy = 85 % Kappa coefficient = 0.8384

Following the similar calculation of case 1, result for case 2 is presented in Table 3 that contrasts to the case 1. After fusion, the values are lower than that of prior to fusion. Possible conclusion for this comparison was that after fusion the accuracy was lower than that of prior to fusion. However, considering the disparity of accuracy between the two pairs on overall accuracy and Kappa, there is only a minor change of values and the conclusion is not rational.

To be more acceptable on comparison of results, more parameters of accuracy assessment such as Kappa variance, z-test and Margfit should be applied.

Table 3. Classification accuracy of SPOT and SPOT + SAR

	SPOT prior to fusion	SPOT + SAR after fusion
Overall accuracy (%)	90	88
Kappa coefficient	0.8901	0.8715

3.2 Kappa variance, Margfit and z-test analysis for comparison

There are two problems for comparison evaluation of classification results in cases of different data types, different areas and classification techniques. The first is that if only overall accuracy and Kappa coefficient are taken into account, it just presents values higher or lower than the other but the disparity between the two values in pair is omitted, the comparison will be bias. The reason is that magnitude of disparity is unknown and statistical confidence of comparison is insufficient to give a good conclusion. The second is that in the cross table 1 and 2 present dissimilarity on number of reference samples that there is a large difference between the two assessments for case 1 on number of reference pixels. The reason is when pixel size of Landsat image resampled from 30m up to 10m though position of reference samples kept unchanged making number of reference pixels increased. This does not arise in case 2, pixel sizes are not changed and number of reference pixels is retained.

Table 5. Comparison evaluation in pairs

	Overall accuracy (%)	K	K variance	Overall accuracy of Margfit (%)	K Dispa-rity	z-test
Landsat	83	0.81	0.0004963	70	0.03	0.9977
Landsat + SAR	85	0.84	0.0000491	85		
SPOT	90	0.89	0.0001841	81	0.02	0.9947
SPOT + SAR	88	0.87	0.0002118	80		

In order to have a reasonable evaluation and to solve the problems, it is required that accuracy should be paid more attention with statistic analysis. For the former, z-test has to be applied by calculation of Kappa variances and z-test to indicate the significance of difference in pairs. The latter require normalization of all value of both cross-tables by using Margfit analysis. Results of Margfit analysis and z-test are in the table 5 for comparison in pairs of both cases. The z-test shows that the values are lower than the absolute value $|1.96|$ of the 95% confidence level for both cases (0.9977 and 0.9974). Regarding to Kappa, the results of this pairwise test for significance between two error matrices indicates that these two results of Landsat and SPOT prior to and after fusion with SAR are not significantly different. At

other respect, in Landsat case, after normalizing by Margfit analysis, overall accuracy is considerable increased after fusion (70 % and 85 %) while in SPOT case it is just a minor change (81 % and 80 %).

4. CONCLUSIONS

Accuracy of classification result of case 2 is that different significance is low and the two kinds of data can be used interchanged while the case 1 has to be re-checked on reference sampling. In this study, fusion of SAR and optical images for land cover classification has come two views. In the pairwise comparison of Khat analysis and z-test presents a clear assessment and conclusion can get a high confidence. The other, with only analysis of overall accuracy with prior to and after Margfit analysis, for cases 1, it should be combined many respects of accuracy assessment to come a reasonable decision.

Accuracy assessment of classification should be highly paid attention when using different satellite images for comparison applications. It should be considered and analyzed at many respect to come a reasonable conclusion. Omission of any assessment parameter could lead to a bias conclusion on using satellite data like in this case should be avoided.

5. ACKNOWLEDGMENT

This report is part of the project RICEMAN, *Rice and Mangrove monitoring by using remote sensing technology in the South Viet Nam*, funded by the MONRE, Vietnam, implemented by the GIS and Remote Sensing Research Center - HCMC Institute of Resource Geography. Thanks to the U.S. Geological Survey, Department of the Interior/USGS with policy of open access of Landsat data and the METI (Japanese Ministry of Economy, Trade and Industry) and JAXA (Japan Aerospace Exploration Agency) retain ownership of the original ALOS PALSAR data used for this study, which were distributed by Earth Remote Sensing Data Analysis Center.

6. REFERENCES

Claudio Dupas, 2000. SAR and Landsat TM image fusion for land cover classification in the Brazilian Atlantic Forest domain. *International Archives of Photogrammetry and Remote Sensing*. Vol. XXXIII, Part B1. Amsterdam 2000.

Congalton, R., 1991. A review of assessing the accuracy of classifications of remotely sensed data, *Remnote Sensing Environment* 37-35-46.

Congalton, R. and Kass, G., 2008. *Assessing the accuracy of remotely sensed data: Principles and Practices*. CRC Press, Inc. 2nd Edition.

Hussin, Y.A., 2007. *Synthetic aperture radar and optical satellite images for detecting and monitoring tropical deforestation and forest degradation in Southeast Asia,* http://www.isprs2007ist.itu.edu.tr/plenary1.pdf, 15/05/2011.

Sheoran, A., Barry H. and Salim S., 2009. Land cover/use classification using optical and quad polarization radar imagery. *ASPRS 2009 Annual Conference*, Baltimore, Maryland, March 2009.