

Article

Land Cover Classification Using a KOMPSAT-3A Multi-Spectral Satellite Image

Tri Dev Acharya, In Tae Yang and Dong Ha Lee *

Department of Civil Engineering, Kangwon National University, Chuncheon, Gangwon 24341, Korea; tridevacharya@gmail.com (T.D.A.); intae@kangwon.ac.kr (I.T.Y.)

* Correspondence: geodesy@kangwon.ac.kr; Tel.: +82-33-250-6232

Academic Editor: Wen-Hsiang Hsieh

Received: 7 October 2016; Accepted: 16 November 2016; Published: 21 November 2016

Abstract: New sets of satellite sensors are frequently being added to the constellation of remote sensing satellites. These new sets offer improved specification to collect imagery on-demand over specific locations and for specific purposes. The Korea Multi-Purpose Satellite (KOMPSAT) series of satellites is a multi-purposed satellite system developed by Korea Aerospace Research Institute (KARI). The recent satellite of the KOMPSAT series, KOMPSAT-3A, demonstrates high resolution multi-spectral imagery with infrared and high resolution electro-optical bands for geographical information systems applications in environmental, agricultural and oceanographic sciences as well as natural disasters. In this study, land cover classification of multispectral data was performed using four supervised classification methods: Mahalanobis Distance (MahD), Minimum Distance (MinD), Maximum Likelihood (ML) and Support Vector Machine (SVM), using a KOMPSAT-3A multi-spectral imagery with 2.2 m spatial resolution. The study area for this study was selected from southwestern region of South Korea, around Buan city. The training data for supervised classification was carefully selected by visual interpretation of KOMPSAT-3A imagery and field investigation. After classification, the results were then analyzed for the validation of classification accuracy by comparison with those of field investigation. For the validation, we calculated the User's Accuracy (UA), Producer's Accuracy (PA), Overall Accuracy (OA) and Kappa statistics from the error matrix to check the classification accuracy for each class obtained individually from different methods. Finally, the comparative analysis was done for the study area for various results of land cover classification using a KOMPSAT-3A multi-spectral imagery.

Keywords: land use; land cover; classification; KOMPSAT-3A; multi-spectral imagery; Mahalanobis Distance; Minimum Distance; Maximum Likelihood; Support Vector Machine

1. Introduction

Advances in science and technology have enabled many applications for decision making and problem solving applications. Remotes Sensing is one such rapidly advancing space technology that has been providing significant solutions in the areas of natural resource management and environmental assessments. With the availability of remotely sensed data from different sensors of various platform with a wide range of spatiotemporal, radiometric and spectral resolutions, the technology has been applied in a variety of sectors such as agriculture, forestry, geology, hydrology, land use mapping, oceanography and urban development and planning [1].

In recent years, new sets of satellite sensors are being frequently added annually to the constellation of remote sensing satellites. They offer improved specification to collect imagery on-demand over specific locations and for specific purposes. Recent high resolution satellite images are now competing with images obtained through aerial photography particularly due to the highly enhanced spatial resolution which enables them to be applied to cartography. For further

information, Kim et al. described space-based earth observation activities in South Korea [2]. The Korea Multi-Purpose Satellite (KOMPSAT) series of satellites is a multi-purposed satellite system developed by Korea Aerospace Research Institute (KARI). The recent satellite of KOMPSAT series, KOMPSAT-3A is a sister spacecraft of KOMPSAT 3 (Arirang 3) and is Korea's first earth observation/infrared satellite with two imaging systems on board. It is equipped with two imaging payloads; the Advanced Earth Imaging Sensor System A (AEISS-A) and an Infrared Imaging payload. It is now demonstrating high resolution panchromatic and multi-spectral imagery in near infrared (NIR), Red (R), Blue (B) and Green (G) bands for Geographical Information Systems (GIS) applications in environmental, agricultural and oceanographic sciences as well as natural disasters [3].

Land cover is an essential variable that impacts and connects various aspects of human and physical environments, and is critical for the study of ecosystems, climate change, and biodiversity. Land cover mapping is one of the core applications of remote sensing technology [4], which provides a map-like representation of the Earth's surface that is spatially continuous and highly consistent [5]. Accurate and up-to-date land cover maps provide better visualization of the environment and help in planning, modeling and decision-making processes in natural resource management.

Traditionally, the land cover mapping was done by manual stereoscopic drawing, which has been overcome by digital methods in recent years. In the recent digital world, land cover mapping of multispectral data can also be done by manual digitizing on a computer screen. As manual digitizing takes much time and cost, the classification is usually carried out by implementing algorithms in computers. Various classification techniques that have been developed in recent decades can be categorized into three broad classes: unsupervised, supervised and object based classification. In unsupervised classification, methods such as K-Means and Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA) clusters are generated and classes were assigned later to those clusters [6]. In the case of supervised methods such as Mahalanobis Distance (MahD), Minimum Distance (MinD), Maximum Likelihood (ML), Decision Tree (DT), Neural Network (NN) and Support Vector Machine (SVM) training areas are selected based on expert knowledge and field work and classification is performed based on the generated signature file of the training data [7]. A little more advanced than supervised methods are object-based classification methods which first perform multiresolution segmentation, and then classification is performed based on conditional statistics of training segments. In order to effectively derive reliable information from satellite data, appropriate classification techniques are essential. A review of these algorithms can be found in Lu and Weng [8].

In this study, we perform the land cover classification of high resolution KOMPSAT-3A multispectral data using four supervised classification methods: MahD, MinD, ML and SVM around the Buan area of Jeollabuk province, Republic of Korea. The challenge with the KOMPSAT 3A data is that it has only four i.e. Red, Blue, Green and NIR (RGB-NIR) bands of very high resolution. A secondary objective, to examine the ability of the derived ratio and indices bands in improving classification of KOMPSAT-3A, was also investigated. The study will explore the potential of KOMPSAT multispectral data for classification of land cover and compare the performance of the results among the methods. The classification study is the first in the Buan area using KOMPSAT-3A data.

2. Materials and Methods

In this study, first of all, a study area was selected. Obtained data was georeferenced for geometric correction and band ratios and indices were derived and then stacked into one. Training and validation pixels from the same high resolution natural color composite and panchromatic image were extracted based on field visit and expert knowledge. These pixels were used for classification and accuracy assessment of the results. Finally, four common classifiers, i.e., MahD, MinD, ML and SVM were applied on original and stacked dataset to compare the accuracy results. The overall process adopted by the study is shown in Figure 1.

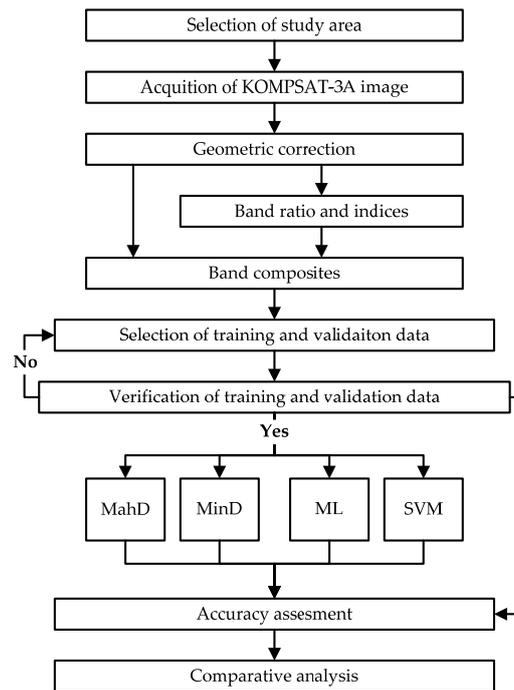


Figure 1. Overall process adopted in this study where land cover classification of Korea Multi-Purpose Satellite (KOMPSAT)-3A image was done using four supervised methods i.e. Mahalanobis Distance (MahD), Minimum Distance (MinD), Maximum Likelihood (ML) and Support Vector Machine (SVM).

2.1. Study Area

An approximate area of 25 sq. km. in central south western region of South Korea was selected for the study (Figure 2). The area lies in Jeollabuk province and is geographically bounded by 35°34'49.38'' N to 35°38'0.47'' N and 126°43'8.41'' E to 126°45'54.48'' E. Most of the area is agricultural land with barren and vinyl house farmlands. Few areas were covered by the forest or built-up structures like roads and residential areas. A river passes through the center which is dammed to preserve water for irrigation. The selection of the area was due to its availability of data and ease in identification of diverse land cover classes.

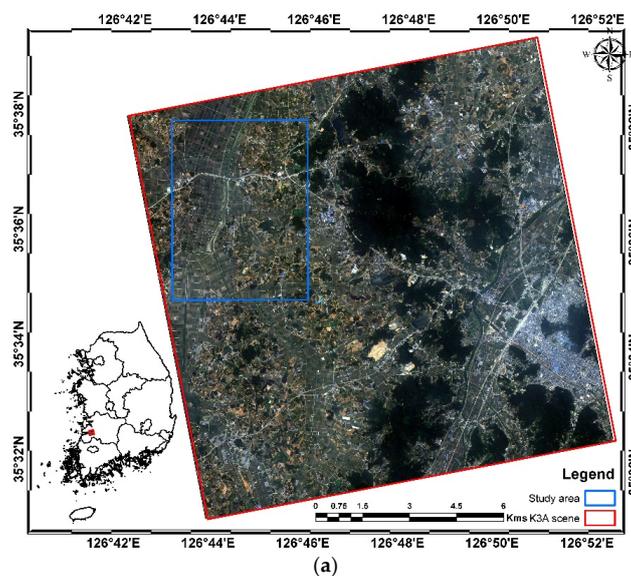


Figure 2. Cont.



Figure 2. Location of the study area in Korea: (a) Full KOMPSAT-3A scene in a natural color composite image taken on 17 June 2016; (b) False color composite of study area.

2.2. Data

The high resolution multispectral data used in the study was taken by KOMPSAT-3A on 17 June 2016. The data used in the study was purchased from KARI using the Arirang Satellite Image Search and Order System (ASISOS) portal. The provided data consisted of Blue (B), Green (G), Red (R), Near InfraRed (NIR) and Panchromatic images in GeoTIF format. The spectral wavelength ranges and the spatial resolution bands of KOMPSAT-3A are presented in Table 1. The image quality of the study area was good and cloud-free.

Table 1. Spectral Bands of Korea Multi-Purpose Satellite (KOMPSAT)-3A.

Band No.	Band Name	Spectral Ranges (nm)	Spatial Resolution (m)
1	Blue (B)	450–520	2.2
2	Green (G)	520–600	2.2
3	Red (R)	630–690	2.2
4	Near Infrared (NIR)	760–900	5.5
5	Panchromatic	450–900	0.55

Initially, the radiometrically corrected satellite image was composited to form a single multiband image and then was geometrically corrected in ArcGIS 10.3 software (Environmental Systems Research Institute, Redlands, CA, USA) through geo-referencing and rectifying. As the data consists of only four bands, for better classification inputs, all possible ratios (B/G, B/R, B/NIR, G/R, G/NIR and R/NIR) and extra Normalized Difference Vegetation Index (NDVI) [9] and Normalized Difference Water Index (NDWI) [10] were derived. Equations (1) and (2) shows the formulae for the calculation of NDVI and NDWI respectively. Again, all the original bands and derived ratio and index bands were composited to form a new multispectral image with 12 bands.

$$\text{Normalized Difference Vegetation Index (NDVI)} = (\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red}), \quad (1)$$

$$\text{Normalized Difference Water Index (NDWI)} = (\text{Green} - \text{NIR}) / (\text{Green} + \text{NIR}), \quad (2)$$

The high resolution georeferenced natural color image, panchromatic image and expert knowledge were used to derive training data for image classification and validation data for overall accuracy

assessment. As the high resolution images have many mixed pixels and it is not efficient in terms of time spent to assign them to a class one by one, we discarded the sampling of points. Hence, polygons were sampled randomly in the study area and classes were assigned to the polygons. All the pixels in the polygon were converted into points and the points were separated into 70% (76,700 pixels) and 30% (32,900 pixels) for training and validation respectively. For even better validation purposes, random pixels were added to the validation dataset outside the polygons that are not spatially influenced. Figure 3 shows all the sampled polygons in the study with zoomed pixels for better understanding of the sampling concept used in the study.

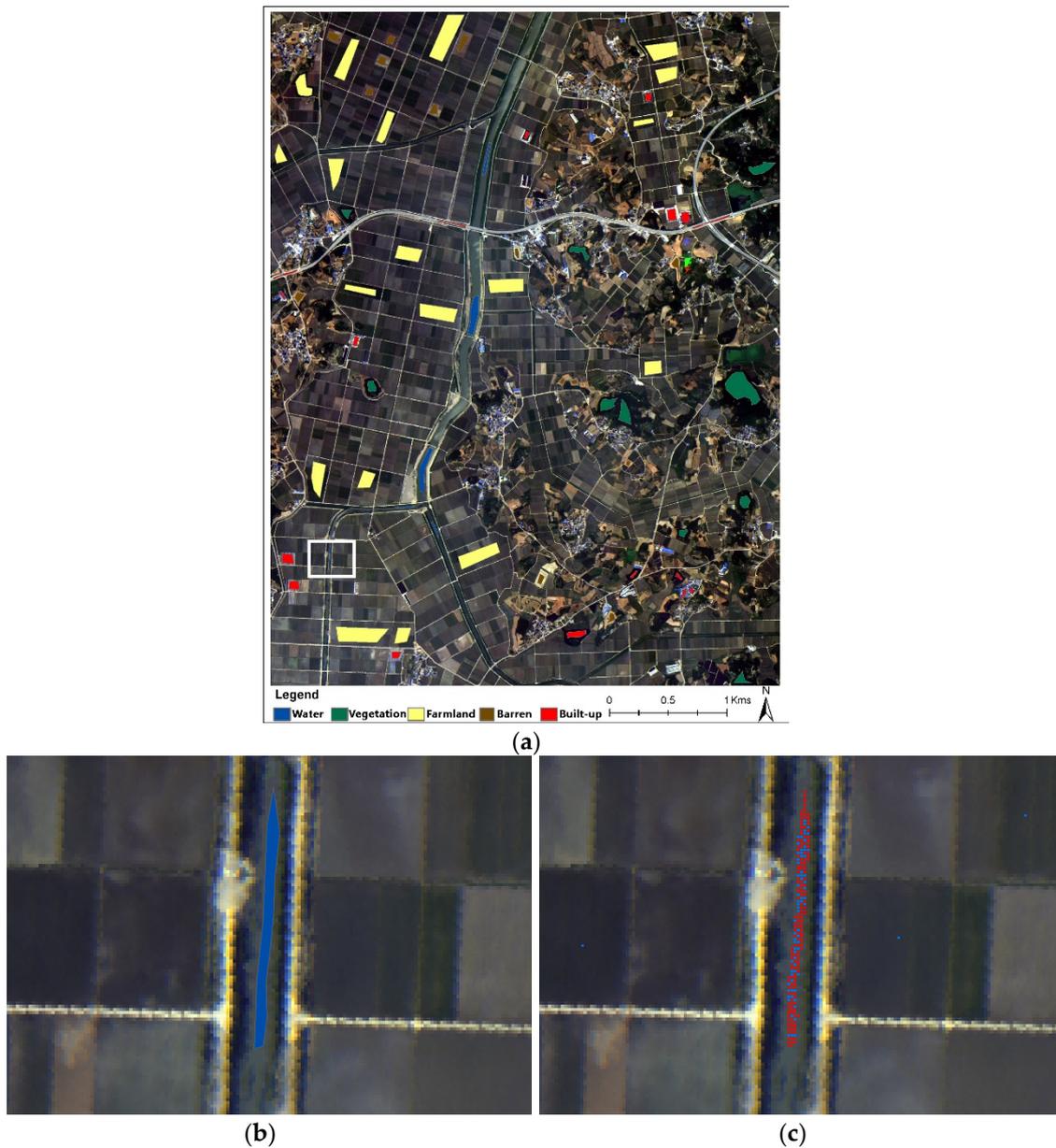


Figure 3. Sampling for training and validation: (a) Sampled polygons in the study area; (b) Zoomed view of a water polygon (white box on left-bottom of Figure 3a); (c) Zoomed area with training (red) and validation (blue) pixels.

2.3. Classification and Accuracy Assessment

The study area was classified into five land cover classes: agriculture, barren, built-up, forest and water (Table 2). These classes represent the most dominant and important land cover in the area. Selection of only five classes was due to ease of selection and an associated decrease in the error of misclassification. Well known classification methods MahD, MinD, ML and SVM were applied to the study area using Environment for Visualizing Images (ENVI) 5.1 software (Exelis Visual Information Solutions, Boulder, CO, USA). The training and validation data sets were common for each of the classification algorithms.

Table 2. Land cover classes classified in the study.

Class Name	Description
Barren	Land areas of exposed soil and barren areas
Built-up	Residential, industrial, roads, vinyl houses
Farmland	Crop fields and fallow lands
Vegetation	Mixed grasslands and forests
Water	River and reservoirs

In ENVI, all the classifications are made based on the probability or the Euclidian distance from pure pixels. ML is one of the most commonly supervised classifiers. It is based on statistics for each class in each band, which are normally distributed, and calculates the probability that a pixel or object belongs to each class and then assigns the pixel or object to the class with the highest probability [11]. The MinD uses the mean vectors of each endmember and calculates the Euclidean distance from each unknown pixel to the mean vector for each class. All pixels are classified to the nearest class unless a standard deviation or distance threshold is specified, in which case some pixels may be unclassified if they do not meet the selected criteria [11]. MahD classification is a direction-sensitive distance classifier that uses statistics for each class, and assumes all class covariances are equal and, therefore, it is a faster method [7]. All pixels are classified to the closest region of interest class unless a distance threshold is specified, in which case some pixels may be unclassified if they do not meet the threshold [11]. SVM is a non-parametric supervised classification algorithm that produces good classification results from complex and noisy data. The good accuracy is dependent on how well the process is trained.

Finally, the User's Accuracy (UA), Producer's Accuracy (PA), Overall Accuracy (OA) and Kappa statistics were calculated from the error matrix to compare the performance of classified maps.

3. Results and Discussion

The main challenge for classification of the KOMPSAT-3A data is that it has only four bands. The classification may not be as good as for other satellite data such as Landsat which provide more than four bands. Hence, to add more variables, the possible ratio and two indices were derived.

Figure 4 shows the example of two ratios and two indices examples where we can see some specific objects are clearly separable, even visually. Based on the assumption that derived bands may help in classification, two composite images were created for classification. The first one consisted of only four bands whereas the new derived composite consisted of 12 bands.

Figures 5 and 6 show the classified maps for the original and indices and ratio composite case studies for the study area. The results of the classification of the study area were based on the same 70% of the sampled training data set. Similarly, on the basis of the remaining 30% sampled validation data, the accuracy of classification were evaluated. Table 3 shows the PA, UA, OA and Kappa statistics for both composite cases using various classifiers.

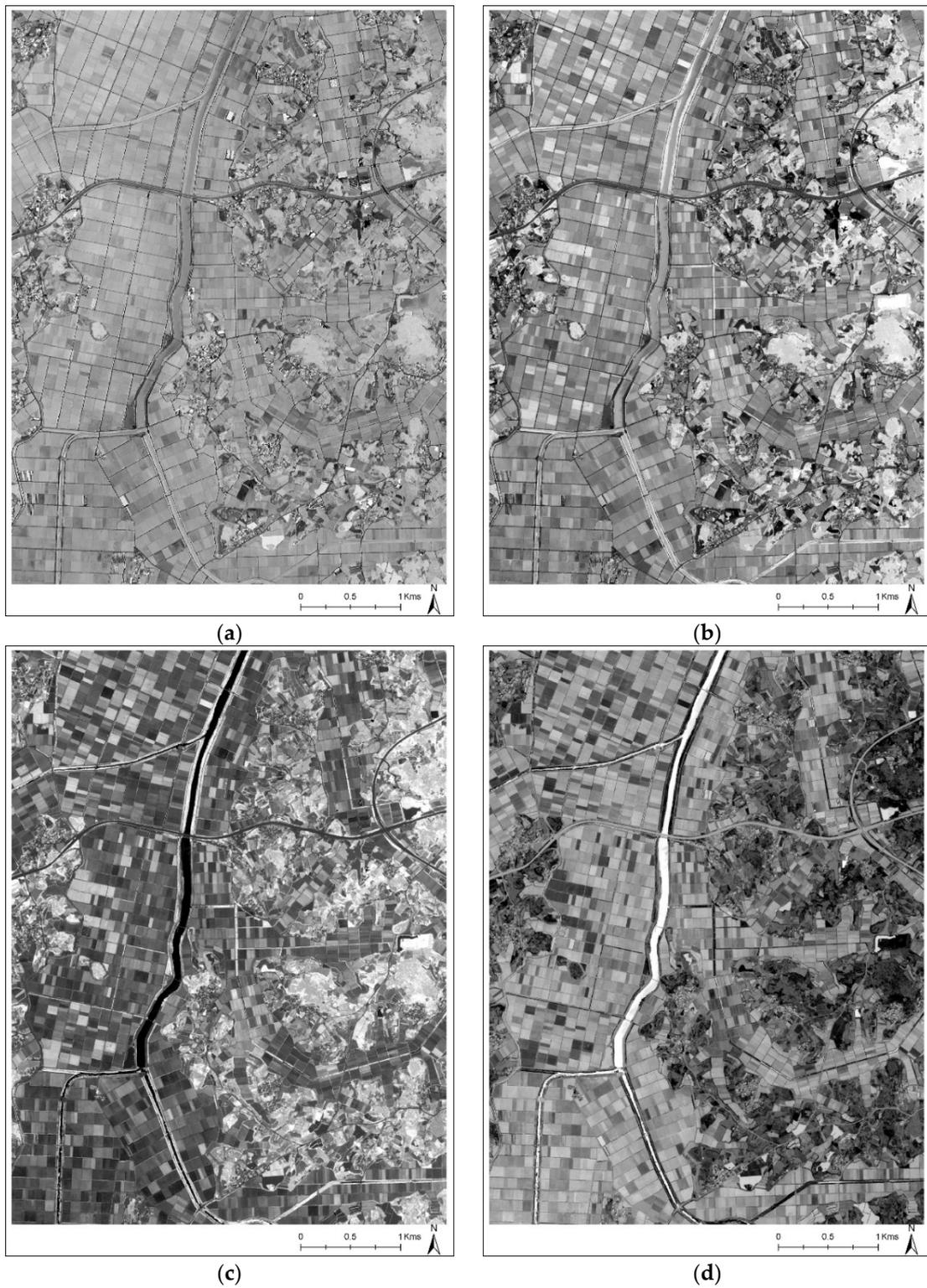


Figure 4. Example of derived ratio and indices from KOMPSAT-3A bands in the study area: (a) Blue/Green; (b) Green/Red; (c) Normalized Difference Vegetation Index (NDVI); and (d) Normalized Difference Water Index (NDWI). Compositing these bands could provide better classification information for original bands.

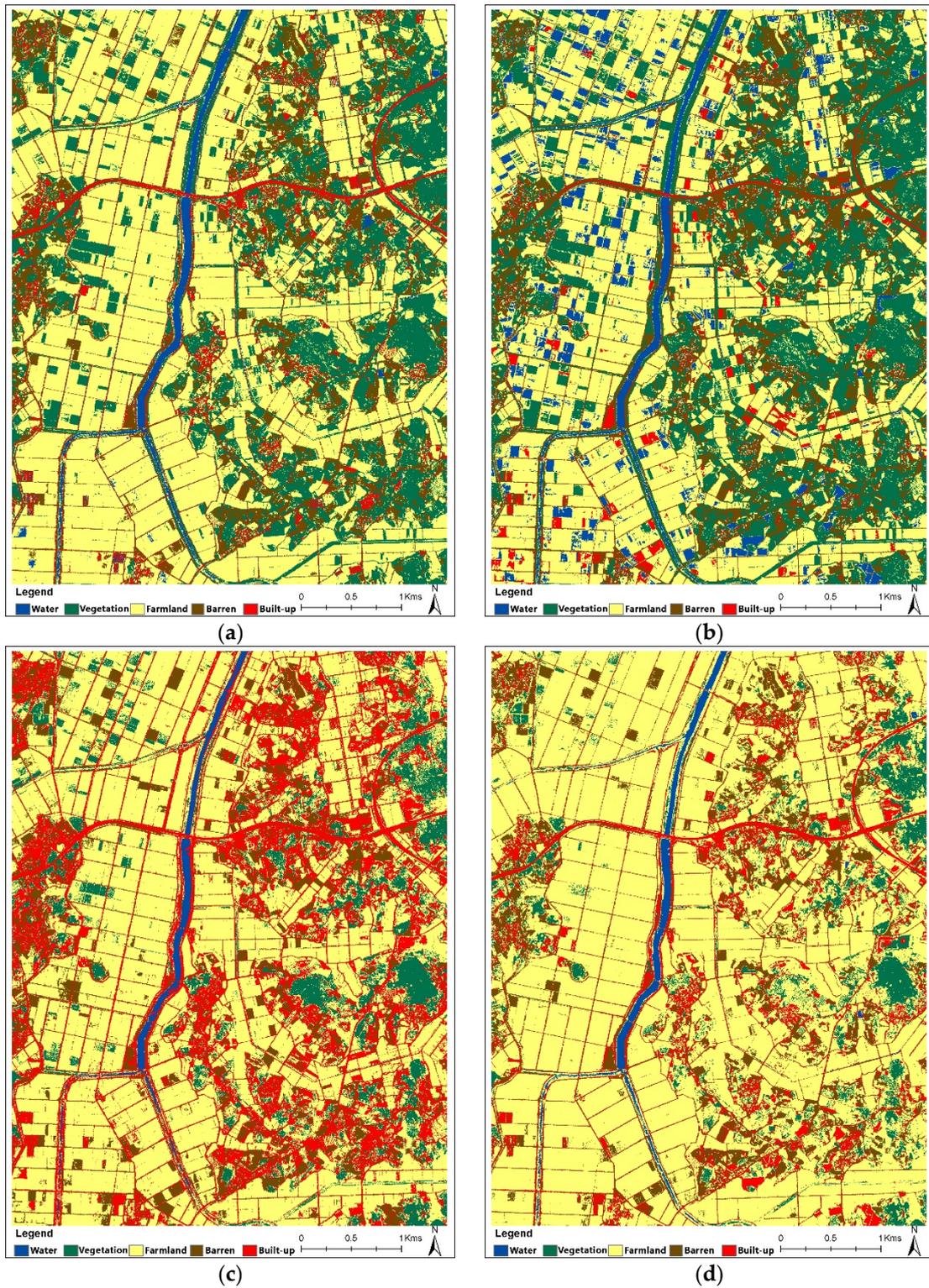


Figure 5. Land cover classification of the test site using original four bands composite: (a) Mahalanobis Distance (MahD); (b) Minimum Distance (MinD); (c) Maximum Likelihood (ML); and (d) Support Vector Machine (SVM).

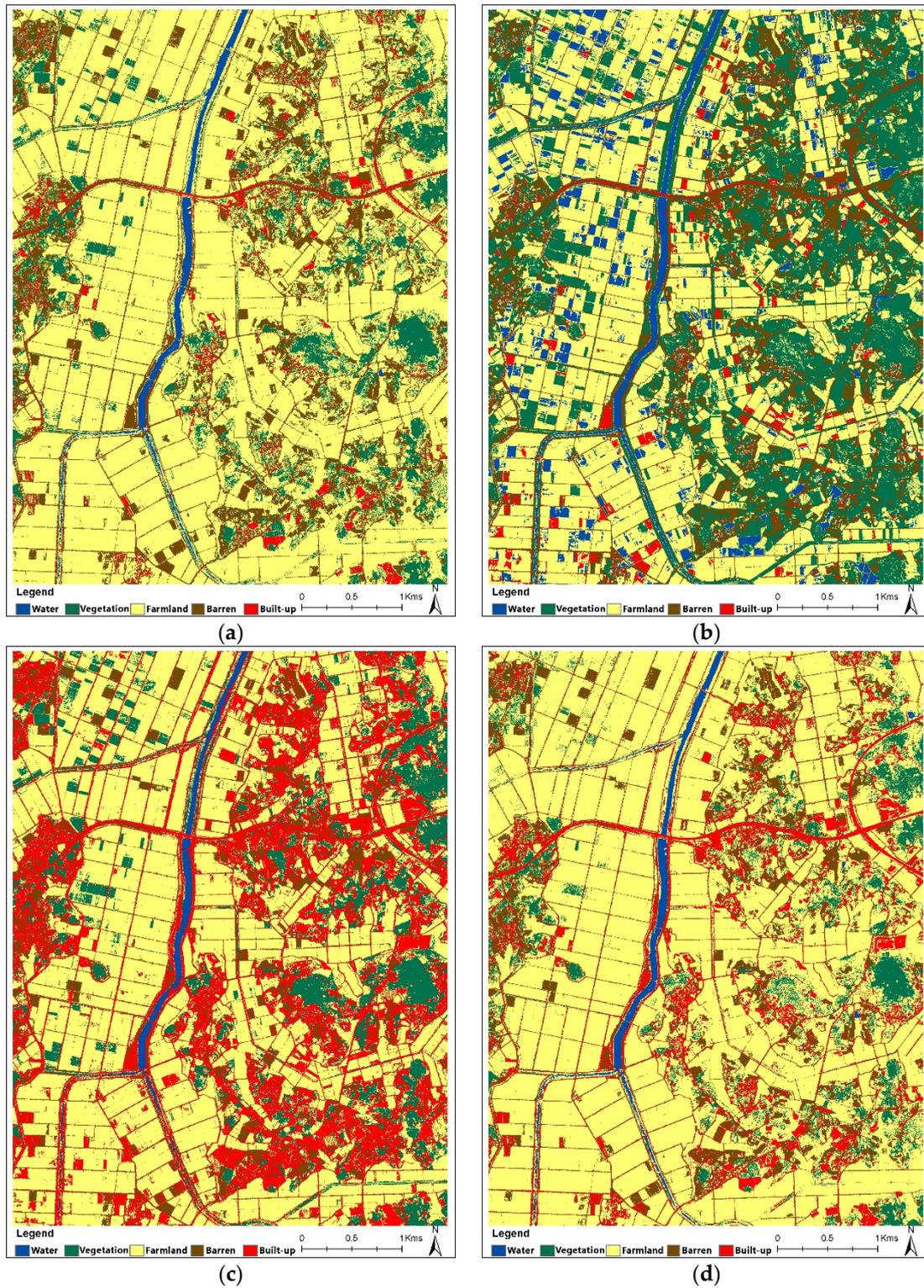


Figure 6. Land cover classification of the test site using ratio and indices bands composite: (a) MahD; (b) MinD; (c) ML; and (d) SVM.

Table 3. Accuracy measures of Mahalanobis Distance (MahD), Minimum Distance (MinD), Maximum Likelihood (ML) and Support Vector Machine (SVM) classifiers in the study area.

Case	Classifier	Water		Vegetation		Farmland		Barren		Built-up		OA	Kappa
		PA	UA	PA	UA	PA	UA	PA	UA	PA	UA		
Original	MahD	100	90.85	96.48	52.39	78.47	92.96	84.56	85.60	51.84	98.49	80.93	0.682
	MinD	100	38.46	91.09	48.74	65.84	90.87	47.69	53.77	24.51	60.83	67.33	0.493
	ML	99.33	100	97.36	68.57	87.11	99.00	98.49	71.80	69.77	85.76	88.90	0.812
	SVM	99.69	100	76.62	83.18	95.94	93.52	96.55	90.25	86.20	97.66	92.46	0.862
Ratio & index composite	MahD	99.69	100	96.32	70.84	90.34	97.07	97.37	82.86	68.98	99.89	90.68	0.838
	MinD	100	38.46	91.09	48.74	65.84	90.87	47.69	53.77	24.51	60.83	67.33	0.493
	ML	99.59	100	98.65	58.90	81.33	99.44	99.14	92.07	92.34	85.27	87.20	0.791
	SVM	99.74	100	83.36	85.98	96.47	95.24	97.41	91.42	87.11	97.25	93.98	0.890

PA, UA and OA are Producer’s Accuracy, User’s Accuracy and Overall Accuracy. All are represented in percentage (%).

From Table 3, the well classified land cover seems to be water except in the case of MinD where the UA is very low, i.e., 38.46%. Moreover, it is clear that the overall accuracy of MinD is very poor and is unchanged in the case of the derived composite. MahD showed much improvement of about 10% from 80.93% to 90.68% in overall classification accuracy with the addition of composite ratio and indices, SVM also showed an increase, but only of 1%.

Visually, the results of the classification of the study area seem to be very variant in each cases. The classification from MahD and SVM seems to have visually balanced classification of land cover. However, MahD seems to be covered with more vegetation than SVM which is a built-up area. Whereas, MinD and ML shows very poor classification, they were more favored in classification of vegetation and built-up areas, respectively. The vegetation and the farmlands were misclassified by the MahD classifier. Similarly, the ML classifier seems to be confused by barren open areas with built-up, roads and vinyl houses. The misclassifications shown by MinD classifier identified black mulching vinyl in farmland as water class, which was due to its dark nature. Additionally, with its better visual classification, SVM has a salt and pepper effect in areas where vegetation was confused with farmlands and vice versa.

SVM is a well-known classifier that produces highly accurate results as expected from a well-trained model. This was also seen in both cases of this study. ML showed higher accuracy compared to MahD and MinD in the first case and the results were more localized in the sampling polygon only. In the second case the results were also localized but with lower accuracy. The compositing derived bands had a negative effect on the probability of the pixel classifications. MahD is a direction-sensitive distance classifier that uses statistics for each class. Also, it is similar to the ML but assumes all class covariances are equal and therefore is a faster method [11]. The increase in classification accuracy shows that the MahD classifier seems to consider the derived band information for the classification.

4. Conclusions

Advancements in science and technology have improved the satellite sensors and new sets of sensors are being added annually with improved quality. These improvements in spatial, radiometric or temporal improvement have been found to be useful in various applications for the management of earth resources. KOMPSAT-3A is the newest family of KOMPSAT satellite constellation launched by KARI of Korea, which provides four high resolution RGB-NIR multispectral data sets along with very high resolution panchromatic images.

In this study, different classification techniques have been examined for the classification of land cover in a rural area of southwestern Korea. Various supervised classifiers were applied to examine the classification of KOMPSAT-3A data in two cases: original composite and derived ratio and indices composite. Due to minimum bands and high resolution, the classification was not as efficient as

expected for large areas. MahD and SVM showed better classification accuracy and also improved with the addition of derived bands, whereas the MinD showed lower overall accuracy and was unchanged. ML misinterpreted the classification of barren and urban areas and accuracy dropped with the derived ratio and indices composite. The classification does not seem to be uniform in separating farmland and vegetation, barren and built-up areas. The main reason behind that is the seasonal variability in the applied case.

In conclusion, the high resolution RGB-NIR bands are not very efficient at pixel-based automatic classification of land covers. Large areas require larger training sets and seasonal variability often affects the result very much in such datasets. The natural color composite or panchromatic are good sources for monitoring and examination of land cover and are best used as the verification or update data in mapping from mid-resolution multispectral data. Also, deriving ratio and indices seems to improve the accuracy in some classifiers, which is good for the KOMPSAT-3A dataset. In the future, different areas with different seasons and complex lands cover need to be explored to better examine the classification usability of the dataset provided by KOMPSAT-3A.

Acknowledgments: This study was supported by 2016 Research Grant from Kangwon National University (No. 520160528).

Author Contributions: The work presented in this paper is a collaborative development by all authors. Tri Dev Acharya and Dong Ha Lee processed the data and drafted the manuscript. In Tae Yang gave valuable suggestions and supervised the entire process. The manuscript was written, revised and corrected by all co-authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Melesse, A.M.; Weng, Q.; Thenkabail, P.S.; Senay, G.B. Remote sensing sensors and applications in environmental resources mapping and modelling. *Sensors* **2007**, *7*, 3209–3241. [CrossRef]
2. Kim, H.O.; Kim, H.S.; Lim, H.S.; Choi, H.J. Space-based earth observation activities in South Korea (space agencies). *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 34–39. [CrossRef]
3. KOMPSAT-3A. Available online: <https://directory.eoportal.org/web/eoportal/satellite-missions/k/kompsat-3a> (accessed on 11 September 2016).
4. Manandhar, R.; Odeh, I.O.; Ancev, T. Improving the accuracy of land use and land cover classification of landsat data using post-classification enhancement. *Remote Sens.* **2009**, *1*, 330–344. [CrossRef]
5. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201. [CrossRef]
6. Tou, J.T.; Gonzalez, R.C. *Pattern Recognition Principles*; Addison-Wesley Publishing Company: Reading, MA, USA, 1974.
7. Richards, J.; Jia, X. *Remote Sensing Digital Image Analysis: An Introduction*; Springer: New York, NY, USA, 1999.
8. Lu, D.; Weng, Q. A Survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870. [CrossRef]
9. Rouse, J., Jr.; Haas, R.; Schell, J.; Deering, D. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ.* **1974**, *351*, 309.
10. Gao, B.C. NDWI— A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [CrossRef]
11. Harris Geospatial Solutions. *Using ENVI*; Harris Geospatial Solutions: Broomfield, CO, USA, 2016; Available online: <http://www.harrisgeospatial.com/docs/routines-136.html> (accessed on 9 November 2016).



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).