A Study on the Land Cover Characteristics in Korea : Application of Hybrid Classifier and Topographic Normalization

Seong-Woo Jeon*, Hui-Cheul Jung*, Sung-Moon Chung*, and Sang-Ik Lee**

* GIS & RS Team, Korea Environment Institute, 1049-1 Sadang-Dong, Dongjak-Gu, Seoul, Korea.

E-mail: swjeon@keins.kei.re.kr, hchjung@keins.kei.re.kr, smchung@keins.kei.re.kr

**The Korean Ministry of Environment, Kwacheon Government Complex, Korea

E-mail: moers@me.go.kr

ABSTRACT

The topographical effect resulted from rugged terrains and inhomogeneous spectral characteristics due to the complexly mixed land cover condition of Korea substantially lower the remotely sensed land cover classification accuracy. In this study, a topographic correction method using digital elevation model to alleviate the topographic effects. To deal with inhomogeneous spectral characteristic, a hybrid classifier with inclusion of prior probabilities was introduced. This investigation concluded that the topographical normalization and hybrid classification with prior probabilities are effective on rugged landscape. The overall and average classification accuracies were improved by 0.92% and 1.016% respectively. The most substantial and noticeable accuracy improvement was observed in forest areas.

INTRODUCTION

Korean landscape consists of about 70% of mountainous topography. However, because of the rugged and indented topographic conditions, remotely sensed data have not been proven effective because the high relief in the mountainous topography encumbers the illumination and irradiance of feature; as a result, the remotely sensed data of high relief area do not accurately represent the true spectral responses of the features on the ground (Colby, 1991).

Many part of Korean landscape is minute and complexly mixed with various land cover types in relatively small units of land. This also present a crucial difficulty in using remotely sensed data in Korea because the complex and minute land cover structure frequently results a bi-modal spectral distribution, which is directly associated with higher range of errors with both parametric and non-parametric classifiers.

Therefore, to resolve or to reduce the addressed problems of using remotely sensed image data in Korea, it is expected to be that the correction of the topographical effect and substituting existing classifiers with more suitable and effective one.

In this paper, topographic correction using digital elevation model (DEM) is performed. Also, land cover classification was performed using the hybrid classifier. Accuracy analyses were performed to quantify the effectiveness of topographic normalization and hybrid classification.

Study Area and Satellite Images

1. Study Area

The study area is located in the center of *Kyoung Sang Nam Do* province of Korea. In Landsat Thematic Mapper satellite image, the study area lies on the left-upper portion of Path 114, Row 36 scene. *Nakdong* River flows through the study site. Most of the site is covered by forest and agricultural land. The topography of the site is mostly mountainous with steep relives. Figure 1 shows the location of the study site.

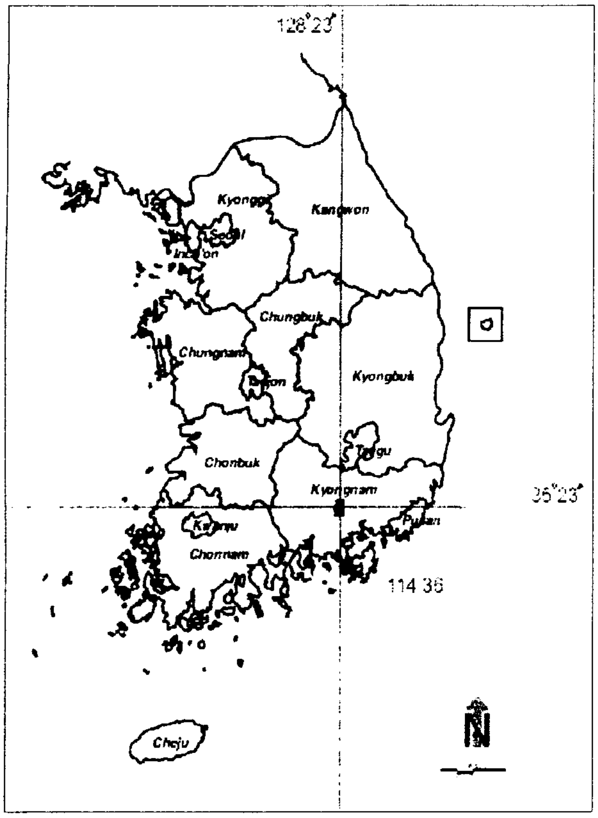


Figure 1. Study Area and Landsat TM Tracks

2. Satellite Image Data

- Landsat TM Path 114, Row 36 (sub-scene)
  - Acquired Date : May 17, 1997
  - Sun elevation angle : 61 °
  - Sun azimuth angle : 116 °
- 30-meter resolution digital elevation model (DEM)
  - Derived from 1:50000 scale national standard digital map (*Namji*, NI 52-2-17)
- Topological map used as a reference material

PROCEDURES AND METHODS

Intergraph Corporation's Image Analyst remote sensing image processing software was used to process the satellite image. The Modular GIS Environment System (MGE) was utilized for mapping. Procedures are as below:

- Artifacts removal and atmospheric correction (bulk correction)
- Geometric correction of the TM scene and DEM
- First stage topographic normalization
- Mask of water area
- Unsupervised competitive training classification
- Supervised classification with inclusion of prior probabilities
- Production of hybrid map
- Post Classification

1. Artifact Removal and Atmospheric Correction

Owing to occasional mechanical malfunctioning of the sensors, a number of anomalies in the raw data, such as striping or banding, were removed through the Fast Fourier Transform (FFT) process. Bulk correction was also performed because varying atmospheric conditions caused by meteorological and sun angle variations influence or change the spectral reflectance of materials on the ground. In this study, the illumination values of clear water areas on the near infrared channel (TM 4) was regarded as the amount of correction to compensate the atmospheric effect for each channels of TM.

2. Geometric Correction and Registration of the TM Scene and Digital Elevation Model (DEM)

The TM scene was geometrically corrected using ground control points (GCP) with a-half-a-pixel precision (RMSE = 0.5 pixel or 15m). The digital elevation model (DEM) and TM scene were registered to one another within a pixel. A linear transformation was performed using a low-degree affine linear interpolation algorithm. Pixels were resampled to 30 meters using the nearest neighborhood method in order to match the spatial resolution of the digital elevation model.

3. Topographic Normalization

For Landsat data, the topographic effect results in

Table 1. Mean and Standard Deviation of the TM scene

TM Band	Raw Data		Bulk Corrected data		1st Stage Normalized data	
	μ	σ	μ	σ	μ	σ
1	89.3	13.29	24.3	18.22	23.0	17.20
2	38.1	8.16	14.6	8.93	13.7	8.39
3	56.5	20.51	30.1	23.2	28.6	21.93
4	124.5	31.99	114.0	34.52	112.3	35.34
5	113.3	33.22	105.0	35.12	102.9	33.39
7	33.9	15.07	30.3	15.47	29.2	14.54

μ = overall scene mean (n=643,505 pixel)
σ = overall scene standard deviation

darker slopes facing away from the sun and brighter sun-facing slopes. In order to reduce much of these topographic effects from the TM image, it is required to create a shade relief model corresponding to the solar illumination conditions at the time of the Landsat overpass. Sun azimuth angle and sun elevation angle data provided by the satellite image distributor are used to produce the shade relief model. The topographical effect was corrected by the cosine law of spherical geometry under a Lambertian surface assumption (Civco, 1993):

$$\gamma = \cos\theta_o \cos\theta_n + \sin\theta_o \sin\theta_n \cos(\psi_n - \psi_o) \qquad (1)$$

where:

γ = cosine angle between incident angle and the local surface normal;

θ_o = solar zenith angle;

θ_n = zenith angle of the normal to the surface;

ψ_n = solar azimuth angle; and

ψ_o = topographic aspect angle.

A linear transformation of each of the original TM bands was performed to derive topographically normalized images through equation (2).

$$\delta DN_{\lambda ij} = DN_{\lambda ij} + \left(DN_{\lambda ij} \times \frac{(\mu_k - X_{ij})}{\mu_k}\right) \qquad (2)$$

Where:

$\delta DN_{\lambda ij}$ = the normalized radiance data for pixel$_{ij}$ in band λ

$DN_{\lambda ij}$ = the raw radiance data for pixel$_{ij}$ in band λ

μ_k = the mean value for the entire scaled (0, 255)

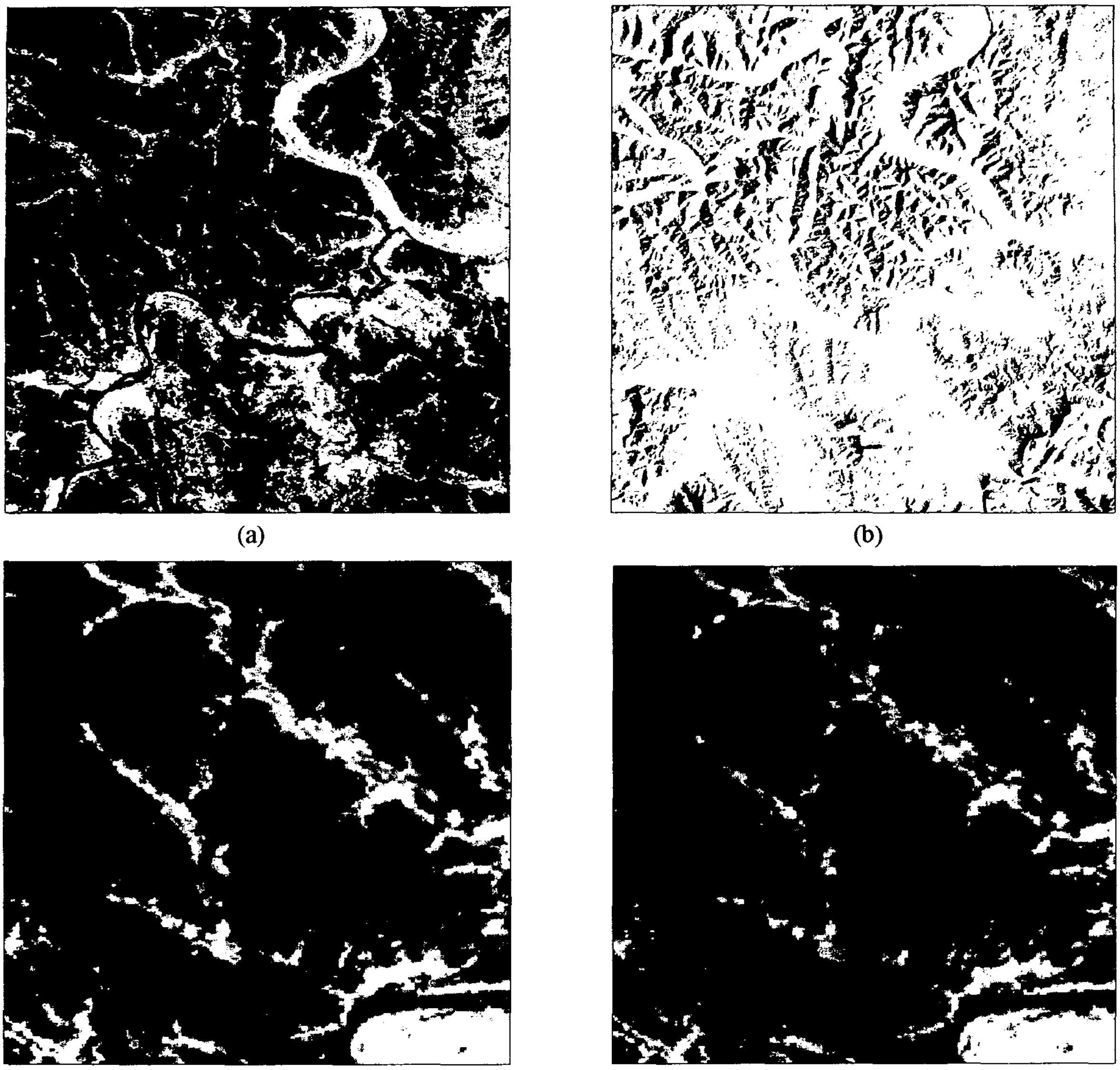
X_{ij} = the scaled (0. 255) illumination value for pixel$_{ij}$

4. Mask of Water Areas

Water is usually easily distinguished and classified by Landsat TM image thanks to the near-infrared band (TM 4). However, if a substantial area of water body is included in a scene or unit area, the classification gets more difficult because the histogram distribution tends to be bi-modal. Since most parametric classifiers, such as maximum -likelihood classifier, rely on a fundamental assumption that the spectral distribution is normal, the area of water needs to be masked off using the near infrared channel (TM 4) prior to classification.

5 Unsupervised Competitive Training Classification

Competitive training is Image Analyst's most sophisticated training method. Competitive training is an unsupervised training method that uses a type of artificial neural network called a simple competitive learning network. Pixels are placed in clusters through an iterative process of competitive and



(a) (b) (c) (d)

Figure 2. Topographically Normalized Image of Landsat TM bands:
(a) Study area; (b) Shaded relief model derived from a 30-meter DEM; (c) Original true-color composite image; (d) Normalized image (θ_o = 29 °, ϕ_n= 116 °)

spectral distance calculations. The primary input parameters are the number of clusters and the number of iterations. A mean is initialized for each band of each cluster with a random number between 0 and 1. The mean is updated by the following equation for all the pixel samples:

$$\text{mean}(t+1) = \text{mean}(t) + \text{delta} * (\text{sample} - \text{mean}(t)) \tag{3}$$

where delta is monotonically decreasing with each iteration.

6 Supervised Classification with Inclusion of Prior Probabilities

Supervised classification methods, both parametric and non-parametric, inherently involve with several limitations. Parametric procedures, such as the maximum-likelihood classifier, are statistically stable and robust but lack in flexibility and in the capability of making correct area estimate. On the other hand, non-parametric classifiers are generally

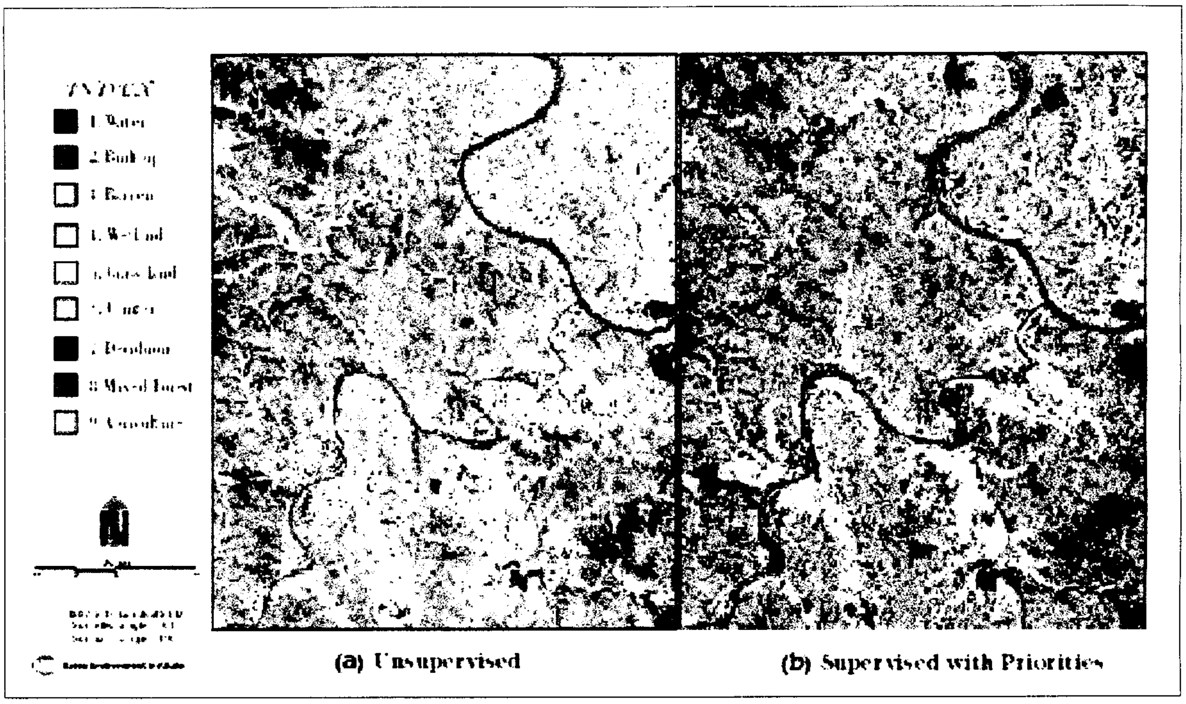


Figure 3. Land Cover Classification Results (a) Distribution of the nine color classes used Unsupervised classifier (b) Distribution of the nine color classes derived from the Maximum-likelihood classifier using prior probablities

too sensitive to distribution anomalies and are critically dependent on training sample size (Fabio Maselli, *et al*, 1992). In order to overcome these shortcomings of each method, an alternative method was chosen : Inclusion of prior probabilities derived from a non-parametric process into the maximum-likelihood classifier (Maselli, *et al*, 1992). Following is the equation provided by Maselli *et al.*.

$$F = (X-M)' C^{-1} (X-M) + \mathrm{Ln}|C| - 2\,\mathrm{Ln}(P) \quad (4)$$

X = Pixel vector

M = Mean vector of the class under examination

C = Variance-covariance matrix of class under examination

In this study, the competitive training unsupervised classification was performed to obtain the prior probabilities.

The inclusion of prior probabilities to supervised classification influenced the area of each class (Figure 3). The prior probabilities are calculated by equation (4) with training areas possessing homogeneous spectral characteristics.

Table 2 shows the prior probability results of seven classes.

$$P = Fr / Frt \quad (5)$$

Where:

Fr = Number of pixels under examination in pixel vector

Frt = Sum of all pixels in all classes in pixel vector

Figure 3(b) shows a result of supervised classification with inclusion of prior probabilities.

Figure 3(a), which is a result of supervised

Table 2. Prior Probabilities of Training Areas

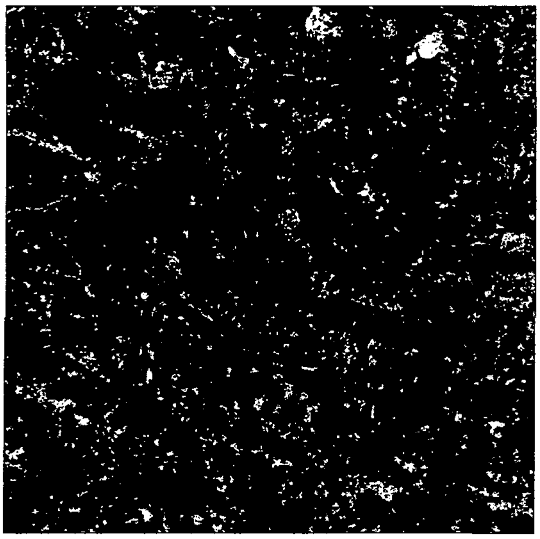
	Water	Wetland	Built-Up	Barren	Agri.	Coniferous Forest	Deciduous Forest	Mixed Forest	Total
Correctly Classified Pixels	1714	144	169	31	376	467	906	94	3901
Total Number of Pixels	1906	156	182	35	413	585	1209	121	4607
Prior Probabilities (P)	0.372	0.031	0.037	0.007	0.082	0.101	0.197	0.020	0.847

classification with inclusion of prior probabilities, demonstrates much distinct and vivid borderlines between built-up areas and agricultural areas. The number of misclassified pixels was decreased; however, the number of unclassified pixels were increased. Therefore, much correction for the unclassified pixels was inevitably required when including prior probabilities.

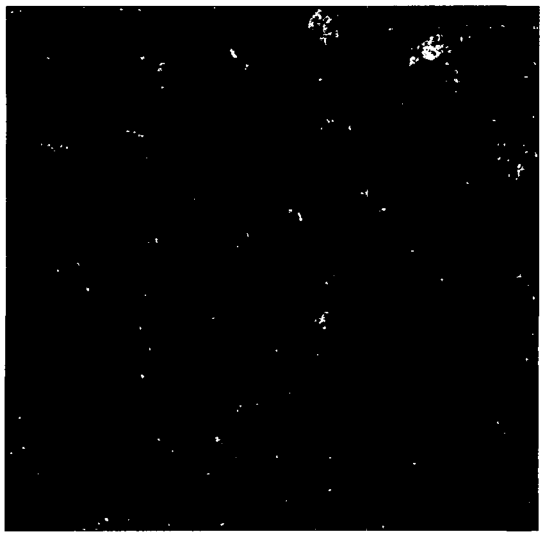
7 Production of Hybrid Classification Map

The pixels and regions left unclassified in supervised classification can be unsupervisedly reclassified. In other words, the outcome image of supervised classification is used as a mask in unsupervised training restricting the unsupervised algorithm to pixels that were not classified previously. The outcome image of this unsupervised classification is then overlaied onto the supervised classification outcome image to produce a hybrid classification map. A validation process using reference data was performed on the outcome image of unsupervised classification.

Figure 4(a) shows the outcome image of supervised classification to be used as a mask. Figure 4(b) is



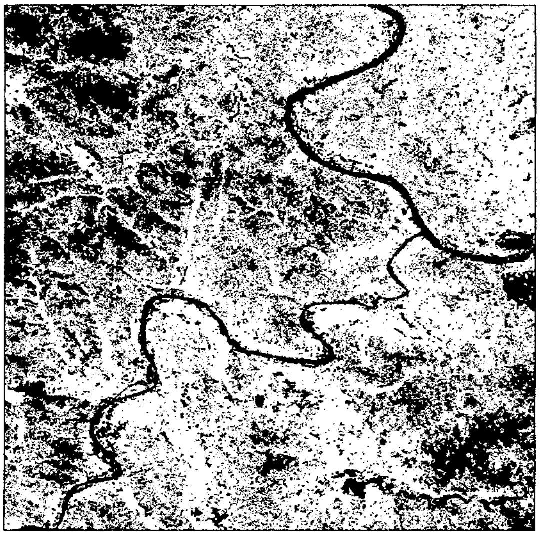
(a)



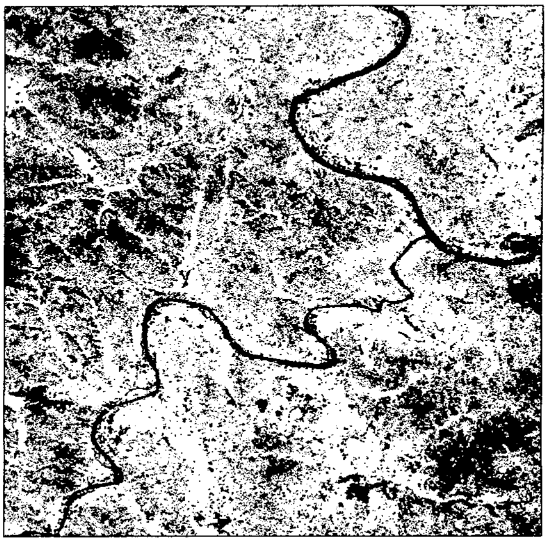
(b)

Figure 4. (a) Unclassified areas from supervised classification

(b) Unclassified areas from competitive unsupervised classification with Para-ML clustering



(a)



(b)

FIgure 5. (a) Hybrid Classification Image (b) Final Classification Image After Post Classification

the outcome results of unsupervised classification using competitive training and Para-ML clsutering methods. Figure 4(b) shows more distinct clusters than Figure 4(a) in general. Some areas in the clusters classified into 'barren' (sky blue clusters) in 4(b) represent green house (so-called "vinyl house") agricultural lands that need to be corrected to agricultural land class.

8 Post Classification

Post classification is a process to correct and remove remaining misclassified and unclassified pixels or region from the outcome image of hybrid classification. Through this process, higher overall accuracy can be achieved. Large clusters of unclassified pixels can be manually corrected using reference data. Small clusters of unclassified pixels can be removed by using filters, such as low pass.

Misclassified areas was corrected after a validation process through random sampling or ground truth to maintain higher accuracy (Sohl, 1999).

Figure 5(a) demonstrates the produced hybrid map and Figure 5(b) shows the final land cover classification image after post classification. A large scale green house agricultural areas on the riparian areas of *Nakdong* River which has been misclassified into 'barren' was corrected. Some misclassification of 'deciduous forest' was observed in 'grass' and thus manually corrected.

RESULTS

In order to quantify the effectiveness of topographic correction to the accuracy of classification, error analyses were performed on the topographically normalized classification image and not-normalized bulk-corrected classification image (Lillesand and Kiefer, 1994). Table 3 and 4 are error matrices for non-topographically normalized image and for topographically normalized classification image respectively.

As Table 3 demonstrates, the average and overall accuracy of non-normalized classification image were 85.77% and 87.35% respectively. On the other hand, as Table 4 shows, the average and overall accuracy of normalized classification image were turned out to be 87.18% and 88.27% respectively. The accuracy of

Table 3. Error Matrix for Non-Normalized Image

Class	Water	Builtup	Barren	Wetland	Grass	Coniferous	Deciduos	Mixed	Agri.	Total	%
1	1495	0	0	14	0	2	0	1	0	1512	31.2
2	0	801	2	0	0	0	0	0	7	810	16.7
3	0	7	301	0	0	0	0	0	0	308	6.4
4	0	0	0	34	0	0	0	0	0	34	0.7
5	0	0	0	0	186	0	0	2	5	193	4.0
6	0	0	0	0	0	910	1	76	5	992	20.5
7	0	0	0	0	0	0	129	23	0	152	3.1
8	0	0	0	0	3	59	45	199	0	306	6.3
9	0	0	0	0	31	0	0	0	509	540	11.1
Total	1495	808	303	48	220	971	175	301	526	4847	-
%	30.8	16.7	6.3	1.0	4.5	20.0	3.6	6.2	10.9	-	100
Accuracy (%)	100.0	99.13	99.34	70.83	84.55	93.72	73.71	66.11	96.77	-	-
Average Accuracy = 85.77 %, Overall Accuracy = 87.35 %											

Table 4. Error Matrix for Normalized Images

Class	Water	Builtup	Barren	Wetland	Grass	Coniferous	Deciduos	Mixed	Agri.	Total	%
1	1499	0	0	15	0	2	0	0	1	1517	31.3
2	0	675	3	0	0	0	0	0	8	686	14.1
3	0	5	330	0	0	0	0	0	0	335	6.9
4	0	0	0	34	0	0	0	0	0	34	0.7
5	0	0	0	0	182	0	0	0	3	185	3.8
6	0	0	0	0	0	917	0	75	3	995	20.5
7	0	0	0	0	1	0	123	25	0	149	3.1
8	0	0	0	0	0	32	22	258	2	314	6.5
9	0	5	2	0	27	0	0	0	604	638	13.1
Total	1499	685	335	49	210	951	145	358	621	4853	
%	30.9	14.1	6.9	1.0	4.3	19.6	3.0	7.4	12.8	-	100
Accuracy (%)	100.0	98.54	98.51	69.39	86.67	96.42	84.83	72.07	97.26	-	-
Average Accuracy = 87.18%, Overall Accuracy = 88.27%											

topographically normalized image was turned out to be 1.016% higher in average than non-normalized data and 0.92% higher overall. To be more specific, most noticeable improvement in accuracies of coniferous, deciduous, and mixed forests classes was observed. The accuracy of 'coniferous forest' was substantially gained by 2.7% from 93.72% to 96.42%, 'deciduos forest' by 11.12% from 73.71% to 84.83%, and 'mixed forest' by 5.96% from 66.11% to 72.07%. The accuracies on 'grassland' and 'agricultural land' were also improved by 2.12% and 0.49% respectively. In 'built-up,' 'barren,' and 'wetland' classes, the accuracies were slightly decreased by 0.59%, 0.83%, and 1.44% respeictvely.

CONCLUSIONS AND DISCUSSION

About 70% of Korean landscape consists of mountainous topography. Thus, there has been a lot of problems regarding maintaining the accuracy of land cover classification and overcoming the topographic effects in Korea. Also, in Korean landscape, a numerous cover types are complexly mixed together in relatively small areas; therefore, precise and accurate land cover classification has been very difficult to achieve neither by existing parametric nor non-parametric classifiers.

The goal of this study was to develop a suitable method to classify the land cover on the areas of relief by applying the most effective classifier and minimizing the topographical effect on multi spectral satellite image data. In this study, a combination of topographic normalization using digital elevation model and hybrid classification with inclusion of prior probability was proven to be very effective. Although the overall classification accuracy was improved by 0.92%, the topographic normalization substantially enhanced the accuracy especially in the forest classes (coniferous, deciduous, and mixed forests) as much as 11.12% (deciduous).

Through out this investigation, several important lessons were found.

First, the most influencing factor to the quality of classification is the quality of satellite data. The satellite data used in this investigation was in very high quality in terms of the amount of haze present and the definition of spectral signature. Thanks to the good quality of the scene, the classified outcome image was able to maintain good quality with precision.

Second, topographic normalization carries a great significance to the quality of classification, especially in downland areas with relief. This study has proven that the topographic normalization enhanced the classification accuracy of all forest classes (deciduos, coniferous, and mixed forest), which mostly located on mountainous topography, was improved substantially while relatively insignificant levels of effects was observed on the flat surfaces, such as built-up, barren, and agricultural areas.

Lastly, Korea's landscape structure and land cover types are relatively tightly and complexly mixed in general which results inhomogeneous spectral characteristics. In many cases, the histogram shows bi-modal distribution curves which inherently result a wider range of errors. To resolve the problem, the hybrid classification method with inclusion of prior probabilities was proven to be effective.

This study was conducted to develop a classification method that is suitable and effective for the Korean landscape conditions. This study was only conducted with minimal data and reference materials as a pilot project. Further studies with multi-temporal scenes in the same classification method and topographic normalization present in this study are strongly suggested to achieve higher level of classification accuracy.

ACKNOWLEDGEMENTS

The authors would like to thank to professor Hong-Sik Yoon, Mr. Yong-Dae Kwon, and Mr. Tae-Young Kim of *Sung Koun Kwan* University's Civil Engineering Department. Also give gratitudes to Mr. Yong-Duk Yang, Mr. Woo-Song Jeon, Mr. Yong-Min Shin, Mr. Seung-Hoon Lee, Ms. Mi-Jung Kang, Ms. Eun-Ah Lee, and Ms. Mi-Hee Kwon of Myil Engineering, Inc..

REFERENCES

Consese, Claudio *et al.*, Feb. 1992. Inclusion of Prior Probabilities Derived from a Nonparametric Process into the Maximum-Likelihood Classifier, Photgrammetric Engineering and Remote Sensing, Vol.58, No.2, pp. 201-207

Civco, Daniel L., Sep. 1989, Topographic Normalization of Landsat Thematic Mapper Digital Imagery, Photogrammetric Engineering and Remote Sensing, Vol.55, No9, pp. 1303-1309

Sohl, Terry L., Apr. 1999. Change Analysis in the United Arab Emirates: An Investigation of Techniques, Photogrammetric Engineering and Remote Sensing, Vol.65, No.4, pp. 475-484

Colby, Jeffery D., May 1991, Topographic Normalization in Rugged Terrain, Photogrammetric Engineering and Remote Sensing, Vol.57, No.5, pp. 531-537

Conese, S. *et al.*, Dec. 1993, Topographic Normalization of TM Scenes through the Use of an Atmospheric Correction Method and Digital Terrain Models, Photogrammetric Engineering and Remote Sensing, Vol.59, No.12, pp. 1745-1753

Richards, John A., 1993, *Remote Sensing Digital Image Analysis: An Introduction*, Springer-Verlag, pp. 265-291

Lillesand, Thomas M. and Keifer, Ralph W., 1994. *Remote Sensing and Image Interpretation*, John Wiley and Sons, pp. 612