Effect of Prior Probabilities on the Classification Accuracy under the Condition of Poor Separability

Kim, Chang Jae[1) · Eo, Yang Dam[2) · Lee, Byoung Kil[3)

Abstract

This paper shows that the use of prior probabilities of the involved classes improve the accuracy of classification in case of poor separability between classes. Three cases of experiments are designed with two LiDAR datasets while considering three different classes (building, tree, and flat grass area). Moreover, random sampling method with human interpretation is used to achieve the approximate prior probabilities in this research. Based on the experimental results, Bayesian classification with the appropriate prior probability makes the improved classification results comparing with the case of non-prior probability when the ratio of prior probability of one class to that of the other is significantly different to 1.0.

Keywords : prior probability, separability, LiDAR, Baysian Classification

1. Introduction

Classification is a common method of geospatial attribute information extraction from images in the field of remote sensing (Choi et al., 2003). Classification with LiDAR (Light Detection And Ranging) has been attempted by several researchers since the 1990s. Axelsson (1999) classified laser point clouds into buildings, vegetations, and electrical power lines using minimum description length criterion. Antonarakis et al. (2008) extracted forests and ground types from LiDAR points by using supervised object-oriented approach. Among various kinds of classification methods, this research focuses on Maximum Likelihood approach, which is based on Bayes' theorem and is the most widely used classification method (Eo et al., 1999). Even though the prior probabilities of the classes are significant, equal probabilities of them have been used in many previous approaches due to the lack of prior probabilities of the involved classes (Ahmet, 2004; Alesheikh, 2003). Hence, this research focuses on the improvement of the classification results by Maximum Likelihood Classification using prior information. In detail, Bayes' theorem is briefly reviewed to investigate the effect of the prior probabilities on the classification results in section 2. The reasonable methods to achieve the appropriate prior probabilities and separability are mentioned in section 3. The relationship among class separability, prior probability, and classification accuracy are investigated through the designed experiments in section 4. Three different cases of experiments are designed using two LiDAR datasets. Finally, conclusive remarks are commented in the last section.

2. Bayes' Theorem

Let the spectral classes for an image be represented by ω_i, $i = 1,...,M$, where M is the total number of classes. In trying to determine the class to which a pixel at a location x belongs, it is strictly the conditional probabilities, $P(\omega_i \mid \mathbf{x}), i = 1,\ldots M$.

The position vector $\mathbf{x}$ is a column vector of brightness values for the pixel. The probability $P(\omega_i|\mathbf{x})$ gives the

1) Department of Geomatics Engineering, University of Calgary, Calgary, Alberta, T2N 1N4, Canada (E-mail: cjkim@ucalgary.ca)
2) Corresponding author · Department of Advanced Technology Fusion, Konkuk University, Seoul 143-701, Korea (E-mail: eoandrew@konkuk.ac.kr)
3) Department of Civil Engineering, Kyonggi University, Suwon 443-760. Korea (E-mail: basil@kgu.ac.kr)

likelihood that the correct class is ω_i for a pixel at position $\mathbf{x}$. Classification is performed as follows;

$$\mathbf{x} \in \omega_i \text{ if } P(\omega_i \mid \mathbf{x}) > P(\omega_j \mid \mathbf{x}) \text{ for all } j \neq i \quad (1)$$

The pixel at $\mathbf{x}$ belongs to class ω_i if $P(\omega_i|\mathbf{x})$ is the largest. The equation (1) can be expressed with $P(\mathbf{x}|\omega_i)$ which is estimated from training data, and $P(\omega_i)$.

$$\mathbf{x} \in \omega_i \text{ if } P(\mathbf{x} \mid \omega_i)P(\omega_i) > P(\mathbf{x} \mid \omega_j)P(\omega_j) \text{ for all } j \neq i \quad (2)$$

where $P(\mathbf{x})$ has been removed as a common factor. For mathematical convenience equation (2) is restated as

$$\mathbf{x} \in \omega_i \text{ if } g_i(\mathbf{x}) > g_j(\mathbf{x}) \text{ for all } j \neq i \quad (3)$$

where $g_i(\mathbf{x}) = \ln\{P(\mathbf{x} \mid \omega_i)P(\omega_i)\} = \ln P(\mathbf{x} \mid \omega_i) + \ln P(\omega_i)$ and ln is the natural logarithm. It is assumed that the probability distributions for the classes follow Gaussian normal distributions. After simplifying and removing common factors the final form of the decision function for Maximum Likelihood Classification, based upon the assumption of normal distribution is equation (4) (John, 1993; Lillesand et al., 1994).

$$g_i(\mathbf{x}) = \ln P(\omega_i) - \frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(\mathbf{x} - \mathbf{m_i})^t \Sigma_i^{-1}(\mathbf{x} - \mathbf{m_i}) \quad (4)$$

where $\mathbf{m}_i$ and Σ_i are the mean vector and covariance matrix of the data in class ω_i, respectively.

It is important to understand how this decision rule behaves with different prior probabilities. If the prior probability $P(\omega_i)$ is very small, then in equation (4) the first term, ln $P(\omega_i)$, will be a large negative number. Therefore, setting a very small prior probability will effectively remove a class from the output classification. Thus it is entirely possible that the observation will be classified into a different class. As the prior probability becomes large and approaches to 1, its logarithm will go to zero. Sine sum of all the possibilities must be one, the prior probabilities of the remaining classes will be small numbers and their decision function, equation (4) will be greatly small (Alesheikh, 2003). After understanding the significance of the effects of the prior probabilities, the reasonable approaches to achieve the appropriate prior probabilities are mentioned in the following section.

3. Prior probability and Separability

Once the class categories are decided, two factors can be considered to improve the classification accuracy. The first one is prior probability and the other is class separability. Prior probability $P(\omega_i)$ is probability that class ω_i occurs in the involved image. In other words, it is the probability mentioning that how many pixels belong to class ω_i in the dataset. This probability is also known or can be estimated from the analyst's knowledge of the image. Generally, three approaches to achieve the prior probability of each class are reviewed as follows.

Supervised classification results using non-prior probability

Prior probability can not be acquired easily, so we may use supervised classification results with non-prior probability. However, when we use this prior probability for new classification without sufficient inspection, this might lead to the biased classification results.

Unsupervised clustering results

This method is more reasonable than the first approach abovementioned. In this approach, neither training data nor information about prior probability is needed. However, it is hard to determine the appropriate number of clusters. Moreover, the decided clusters through the unsupervised approach may not match class categories which the researcher wants to acquire finally.

Random sampling with human interpretation or existing information

Existing information such as census or land plan is not always available for all the datasets. Due to the limited achievement of the existing information, random sampling method is usually adopted to acquire the prior probabilities of the involved classes. The more detailed explanation of the applied method will be mentioned in section 4.

The other factor, separability, is critical to determine how well one class is separable from the other class. If the separability value between two classes is close to 2.0, it implies that these two classes will be classified clearly. Usually, Bhattacharrya Distance and Transformed Divergence are popular separability measurements. (Swain,

1978) Both of them are real values between 0 and 2, where 0 indicates complete overlap between the signatures of two classes and 2 indicates a complete separation between the two classes. Both measures are monotonically related to the classification accuracies. The larger the separability values are, the better the final classification results will be. The value between 0.0 and 1.0 indicates very poor separability. The value between 1.0 and 1.9 indicates poor separability. The value between 1.9 and 2.0 indicates good separability (John, 1993). In this research, the experiments are designed while considering these factors and their effects on the classification results in the next section.

4. Experiments and analysis

Figure 1 illustrates the proposed methodology for the comparison between the Bayesian classification results with prior probability and non-prior probability. Four bands, which are produced from LiDAR points, are utilized for the classification. *n*DSM (Normalized Digital Surface Model) is generated from the height difference between ground and features. *n*DSM is decided to be used to consider the height differences of the features without being affected by the height of the ground. The second band is Laplacian Image, which is modified *n*DSM by Laplacian filter. This image has large values in tree areas and almost zero values in building and flat areas. The third band, Sobel image, is also modified from *n*DSM by Sobel filter. Digital numbers in tree areas are very large, but small and homogeneous in building and flat areas. The last band is Intensity data, which is proportional to ratio of signal power between emission and response. The intensity changes according to the characteristics of the features' materials. Using these four different bands, Bayesian classifications with prior probability and non-prior probability are performed. Finally, the classification results are compared with each other.

The image processing for preparing four bands and the random sampling for computing prior probability are implemented in MATLAB. Image classification is accomplished by using PCI S/W. The proposed methodology for the comparison between the classification results is applied to three different experimental cases. Detailed explanation of the designed experiments is as follows. First, three different categories of classes (building, tree, and flat grass areas) are determined in this research. Two LiDAR datasets are selected to construct the experiments. The first dataset (Dataset 1) covers relatively large areas. The other dataset (Dataset 2) is a small subset of the first dataset (in Figure 2). Figure 3 shows the produced four different bands. The prior probabilities of building and tree areas in Dataset 1 are similar each other. Differently, the prior probabilities of them in Dataset 2 are significantly different each other. Training data is shown over DSM in Figure 4. Building, tree, and flat areas are represented with red, green, and blue color, respectively.

Before implementing the Bayesian classification, separability between different classes is investigated. For this purpose, Bhattacharrya distance and Transformed divergence are calculated using the training data of the classes. As shown in Table 1, the separability between flat grass and building areas is high. Also, the value between tree and

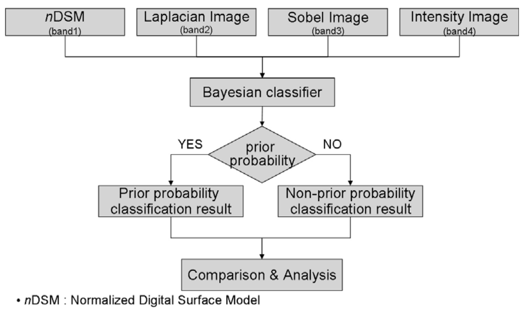

Fig. 1. Methodology used for the comparison between the Bayesian classification results with prior probability and non-prior probability.

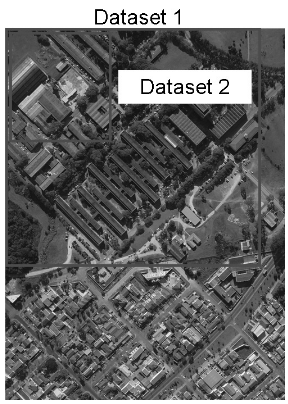

Fig. 2. Aerial photo over study area.

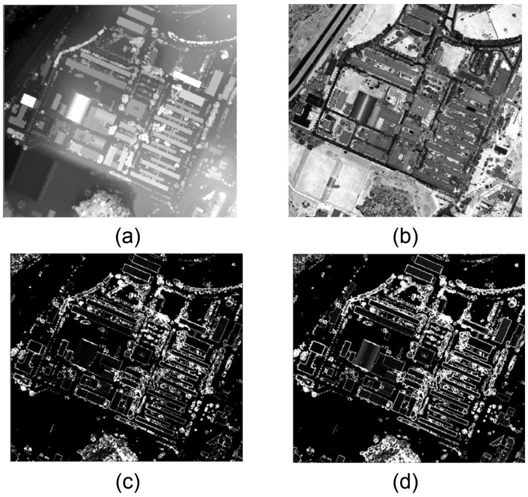

Fig. 3. Produced four different bands in Dataset 1; (a) nDSM, (b) Intensity image, (c) Laplacian image, and (d) Sobel image.

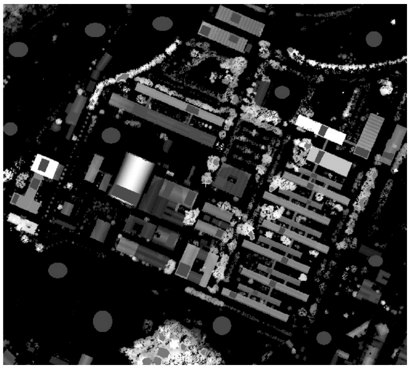

Fig. 4. Training data over DSM.

flat grass areas is relatively high. However, the separability between building and tree areas is low.

The other important factor, prior probability, is computed through random sampling method. To achieve prior probability with more than 95% accuracy, sample size should be larger than 180 (60 X 3 classes) points in each dataset (Van Genderen et al., 1978). 100, 300, 500, 700, and 999 random points are sampled to check the trend of prior probability of each class. Figure 5 shows randomly sampled 500 points on DSM.

After investigating the trend, the appropriate prior probabilities of the involved classes in Dataset 1 are determined. Approximate prior probabilities acquired from the random sampling method are 0.17 for building, 0.15 for tree, and 0.68 for flat areas; see Table 2.

Same approach is applied to acquire the probabilities of the classes in Dataset 2. Approximate prior probabilities acquired from the random sampling method are 0.38 for building, 0.6 for tree, and 0.56 for flat areas. To investigate the effect of the separability and prior probability on the classification results, three different cases of the experiments are designed in this research as shown in Table 3.

Case A and B are accomplished with dataset 1. In case A, the appropriately estimated prior probabilities are utilized; however, inappropriately estimated prior probabilities are utilized in case B. One should note that the there are classes with similar prior probabilities in case A and B. Case C is using dataset 2. The calculated prior

Table 1. Bhattacharrya distance and Transformed divergence results

Band	Building & Trees	Trees & Flat area	Flat area & Building
1 2 3	1.300918	1.999401	1.772855
1 2 4	1.202758	1.999967	1.983950
1 3 4	1.217916	1.999990	1.987323
2 3 4	1.300274	1.974750	1.816416
1 2 3 4	1.459926	1.999995	1.988726

Bhattacharrya Distance

Band	Building & Trees	Trees & Flat area	Flat area & Building
1 2 3	1.837902	2.000000	2.000000
1 2 4	1.759470	2.000000	2.000000
1 3 4	1.630240	2.000000	2.000000
2 3 4	1.821728	2.000000	2.000000
1 2 3 4	1.870926	2.000000	2.000000

Transformed Divergence

probabilities are significantly different and appropriately estimated ones are utilized in case C. Accuracy assessments of the classifications are implemented through an error (or confusion) matrix and computed accuracy statistics. One example of error matrix and accuracy statistics are shown in Table 6 and 7. The experiment, which has the highest kappa coefficient in case C, is selected and shown in these tables.

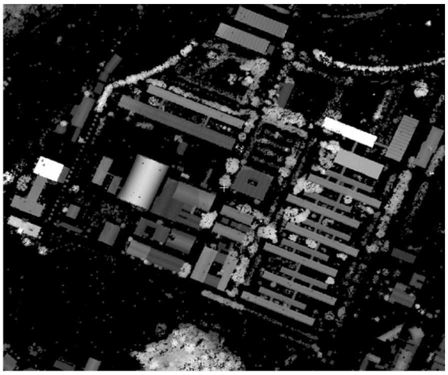

Fig. 5. Random Sampling for prior probability.

In all the experiments of case A, the differences between the classification results with prior and non-prior probabilities are almost ignorable; see Table 4. As shown in Figure 6, the boundary between trees and buildings are not changed after applying the estimated prior probabilities. In case of the classification with non-prior probabilities, the probability of trees and buildings are equally 0.33. And the estimated prior probabilities of trees and building are 0.15 and 0.17, respectively. Hence, the ratio of prior probability of trees to that of building is close to one. In this case, the boundary between classes

Table 2. Prior probabilities of classes according to number of random samples

Classes / # of random samples	Building	Trees	Flat Area
100	0.180	0.130	0.690
300	0.173	0.143	0.684
500	0.178	0.146	0.676
700	0.176	0.149	0.675
999	0.171	0.152	0.677
Approximate Prior probability	0.170	0.150	0.680

Table 3. Three experimental cases

	Separability	Prior probability	Estimated prior probability
Case A	building & trees: low building & flat area: relatively high trees & flat area: high	building & trees: similar building & flat area: different trees & flat area: different	appropriate
Case B	building & trees: low building & flat area: relatively high trees & flat area: high	building & trees: similar building & flat area: different trees & flat area: different	inappropriate
Case C	building & trees: low building & flat area: relatively high trees & flat area: high	All are significanitly different	appropriate

Table 4. Classification accuracy results of case A – Kappa coefficient

Band composition	Non-prior probability	Prior probability	Difference
1 2 3	0.761	0.761	0
1 2 4	0.798	0.794	-0.004
1 3 4	0.768	0.768	0
2 3 4	0.736	0.736	0
1 2 3 4	0.769	0.769	0

does not move significantly even though the appropriately estimated probabilities are applied; see Figure 6.

To investigate the effect of the inappropriately estimated probabilities on the classification results, experimental case B is implemented. The comparison between two results with inappropriately estimated prior probability and non-prior probability is shown in Table 5. The classification results after applying inappropriately estimated prior probability become worse compared to the results with non-prior probability. The wrong estimation of prior probabilities leads the classification results to the worse accuracy. The boundary between classes moves to the wrong direction due to the inappropriately estimates prior probabilities; see Figure 7.

Through the case A and B, the effects of the prior probabilities which are estimated either appropriately or inappropriately are investigated. In these two cases, the ratio of prior probability of trees to that of building is

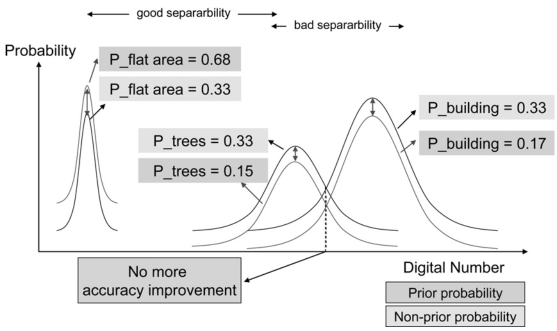

Fig. 6. Probability distribution of case A.

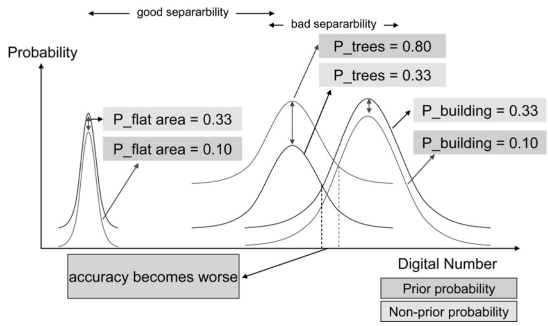

Fig. 7. Probability distribution of case B.

Table 5. Classification accuracy results of case B – Kappa coefficient

Band composition	Non-prior probability	Wrong prior probability	Difference
1 2 3	0.761	0.757	-0.004
1 2 4	0.798	0.767	-0.031
1 3 4	0.768	0.747	-0.021
2 3 4	0.736	0.714	-0.022
1 2 3 4	0.769	0.752	-0.017

Table 6. Error (confusion) matrix for the classification results in case C with band 124, 1-sigma, and prior probability

		Reference Data			
		Building	Forest	Flat grass	Totals
Classified Data	Building	99	2	2	103
	Forest	14	14	12	40
	Flat grass	2	1	154	157
	Totals	115	17	168	300

Table 7. Accuracy statistics for the classification results in case C with band 124, 1-sigma, and prior probability

	Producer's Accuracy	User's Accuracy	Kappa Statistics
Building	86.09%	96.12%	0.937
Forest	82.25%	35.00%	0.311
Flat grass	91.67%	98.09%	0.957
Overall Kappa		0.806	

Table 8. Classification accuracy results of case C & 1 sigma – Kappa coefficient

Band composition	Non-prior probability	Prior probability	Difference
1 2 3	0.745	0.775	+0.030
1 2 4	0.796	0.806	+0.010
1 3 4	0.779	0.783	+0.004
2 3 4	0.740	0.771	+0.031
1 2 3 4	0.785	0.784	-0.001

Table 9. Classification accuracy results of case C & 2 sigma – Kappa coefficient

Band composition	Non-prior probability	Prior probability	Difference
1 2 3	0.740	0.749	+0.009
1 2 4	0.796	0.812	+0.016
1 3 4	0.779	0.789	+0.010
2 3 4	0.719	0.749	+0.030
1 2 3 4	0.785	0.791	+0.006

Table 10. Classification accuracy results of case C & 3 sigma – Kappa coefficient

Band composition	Non-prior probability	Prior probability	Difference
1 2 3	0.745	0.763	+0.018
1 2 4	0.796	0.800	+0.004
1 3 4	0.779	0.783	+0.004
2 3 4	0.744	0.765	+0.021
1 2 3 4	0.785	0.778	-0.007

close to one.

On the other side, the prior probabilities of the involved classes are all significantly different in case C. The prior probability of building is 0.38, trees have 0.06, and flat area has 0.56. Actually, the thresholds of 1, 2 and 3 sigma are used in the classification procedures. The differences between classification results with prior probability and non-prior probability are shown in Table 8, 9, and 10. Thirteen out of fifteen classification results have the improved accuracies. Specific improvements of the accuracies with 1, 2, and 3 sigma are 3.1%, 3.0%, and 2.1%, respectively. These are significant improvements in classification accuracy. The ratio of prior probability of trees to that of building is far from 1.0. Therefore, the boundary between trees and buildings is moved significantly. This movement causes accuracy improvement; see Figure 8.

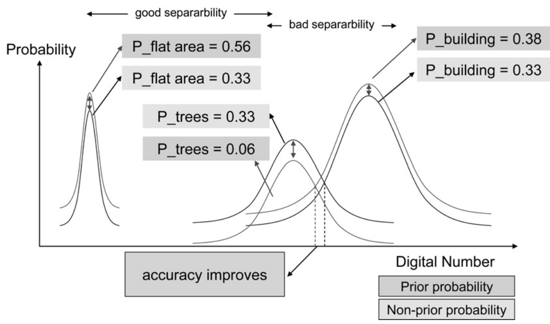

Fig. 8. Probability distribution of case C.

5. Conclusion

Three different experiments have been performed and analyzed in this research. When the ratio of the prior probabilities of the involved classes is close to 1.0, the classification results with the appropriately estimated prior probabilities are not significantly changed compared to the results with non-prior probabilities. Using the same dataset, in another experiment, the inappropriately estimated prior probabilities lead to the deteriorated classification results. When the prior probabilities of the involved classes are significantly different, the appropriately estimated prior probabilities move the boundaries of the classes to the correct direction. Finally, the correct movement of the class boundaries makes the improved classification accuracies. The maximum improvement of the classification accuracy is 3.1%. The improvement happens for most of the band combinations in case C. Conclusively, while comparing the classification results with non-prior probabilities, the classification accuracy significantly improves when the involved classes have relatively low separability, the ratio of the probabilities is far from 1.0, and the prior probabilities are appropriately estimated.

Acknowledgement

This study was partially supported by Chung-Ang Aerosurvey Co. Ltd. (Research Title: A Study on the Establishment of Military Geospatial Attribute Information using Satellite Imagery).

References

Ahmet B. Orun (2004), Automated Identification of Man-Made Textural Features on Satellite Imagery by Bayesian Networks, Photogrammetric Engineering and Remote Sensing, ASPRS, Vol. 70, No. 2, pp. 211-216

Alesheikh, Ali A. (2003), Improving Classification Accuracy Using Knowledge Based Approach, Map India Conference 2003, unpaginated CD-ROM.

Antonarakis, A., Richards, K., Brasington, J.(2008), Object-based land cover classification using airborne LiDAR, Remote Sensing of Environment, Volume 112, Issue 6, pp. 2988-2998.

Axelsson, P. (1999). Processing of laser scanner data - algorithms and applications. ISPRS Journal of Photogrammetry and Remote Sensing, 54, 138-147.

Choi, S. P. and Yang, I. T. (2003), Extraction of Land Surface Change Information by Using Landsat TM Images, Korean Journal of Geomatics, Korean Society of Surveying, Geodesy, photogrammetry, and cartography, Vol. 21, No. 3, pp. 261-261-262.

Eo, Y. D., Kim, C. J. and Kim. Y. I. (1999), Correlation Analysis between the Multivariate Conditional Probability and the Classification Accuracy in Satellite Image Classification, Journal of Civil Engineering, Korean Society of Civil Engineers, Vol. 19, No. III-5, pp. 935-947.

John A. Richards (1993), Remote Sensing Digital Image Analysis: An Introduction, Second Edition, Springer, 340 p.

Lillesand T. M. and Kiefer R. W. (1994), Remote Sensing and Image Interpretation, Third edition, John Wiley & Sons, INC, 750 p.

Swain, P. H. and Davis, S. M.(1978), Remote Sensing : The Quantitative Approach, McGrow-Hill, pp. 170-173.

Van Genderen, J.L. (1978), Remote sensing: statistical testing of thematic map accuracy. In: Remote sensing of environment: an interdisciplinary journal, Vol. 7, Issue 1, pp. 3-14.

(접수일 2008. 6. 4, 심사일 2008. 7. 12, 심사완료일 2008. 7. 23)