DEVELOPMENT OF GROWTH STAGE CLASSIFICATION METHOD FOR PADDY FIELDS BY USING SPARSE LINEAR DISCRIMINANT ANALYSIS

Atsushi Uchida[1], Hozuma Sekine[1], Kotaro Fukuhara[1], Keigo Yoshida[1], Taichi Takayama[1],
Chiaki Kobayashi[2], Osamu Kashimura[2],
Sidik Muljono[3], Arief D.[3], M. Evri[3], Muhamad. Sadly[3]

[1] Mitsubishi Research Institute, Inc., 2-10-3, Nagata-cho, Chiyoda-ku, Tokyo 100-8141, JAPAN
Tel: (+81) 3-6705-6039, Fax: (+81) 3-5157-2145, E-mail: a-uchida@mri.co.jp
[2] Earth Remote Sensing Data Analysis Center, JAPAN
[3] Agency for the Assessment and Application of Technology (BPPT), INDONESIA

KEY WORDS: Hyperspectral, HyMAP, paddy, Sparse Linear Discriminant Analysis

ABSTRACT: The purpose of this study is to develop the growth stage classification method for hyperspectral data of paddy field Indonesia. Rice is one of agricultural crops and constitutes major staple food for Asian countries, and it is difficult to distinguish the growth stages in detail using multispectral data. In this study, we show the result of growth stage classification for estimating the harvest time by applying a machine learning technique, Sparse Linear Discriminant Analysis (SLDA) to hyperspectral (HyMap) data. With regards to the accuracy of classification, SLDA is much better than Spectral Angle Mapper (SAM), a famous and traditional method. Using SAM method to HyMap data, the classification accuracy is about 70% (Uchida *et al*., 2010). Applying SLDA method, the classification accuracy is increased to 97%.
This result indicates the potential ability of sparse linear discriminant function for analyzing the hyperspectral data.

1. INTRODUCTION

The purpose of this study is to develop the hyperspectral data utilization technology for paddy field in Indonesia, in collaboration with the Agency for the Assessment and Application Technology (BPPT) in Indonesia. Rice is one of agricultural crops and constitutes major staple food for Asian countries. The Indonesian government recognizes that the low cost and effective approach for monitoring and managing paddy fields is the key strategy for ensuring national food security, and remote sensing technology and data, especially using a hyperspectral data is expected to be a highly effective solution. By using this growth stage classification information, Indonesia government can know the harvest time for their Area of Interest (AOI), so this information is quite useful for the national food security in Indonesia.
Multispectral data can classify land covers into small categories, and it is difficult to distinguish the growth stages of vegetations in detail. On the other hand, because hyperspectral data has more than 100 bands and ability to extract surface information more in detail, it is required to develop more effective and accurate method.

We have already classified the paddy field growth stage in same area by applying Spectral Angle Mapper (SAM) (Kruse *et al*., 1993) for airborne hyperspectral data, and achieved about 70% accuracy (Uchida *et al*., 2010). But, at the same time, we found the high potential ability of hyperspectral data and if we use more powerful method, we can extract more information and achieve more accurate classification.

This study shows the result of growth stage classification for estimating the harvest time by applying a machine learning technique, Sparse Linear Discriminant Analysis (SLDA) to hyperspectral data. The SLDA, famous machine learning technique, is said to be faster than traditional feature selection methods and the results are quite better with regards to classification rates and sparseness. With regards to the accuracy of classification, we find the result of SLDA is much better than SAM.

2. AREA AND DATA SET

The study areas are located in Indramayu, Subang and Karawang which are well known as a major granary in West Java area. Dual and triple cropping of rice is major trend in these areas. Additionally, because the mountain dam supplies water to paddy fields though irrigation network from the mountain side (south) to the sea side (north), there is a time difference of water supply among the paddy field, so the mix of growth stage is general in these area.

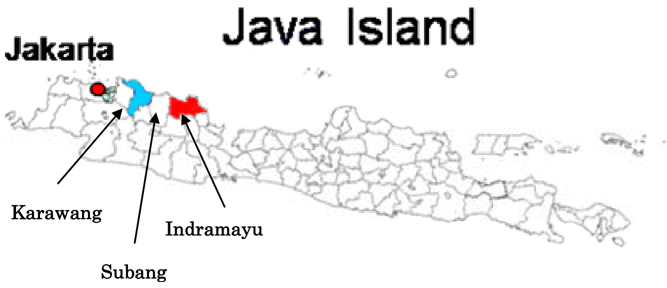


Figure 1 Location of study area

We conducted airborne observation and simultaneous ground measurement at Subang and Indramayu in 2008 and at Karawang and Indramayu in 2011. We selected the HyMap sensor with 126 bands (450nm - 2480nm) as an airborne hyperspectral sensor, and acquired images by 5m spatial resolution. Figures 2 and 3 show the HyMap images of each area. The airborne observational days are 30th June (Indramayu) and 1st July (Subang) in 2008, and 13th July (Karawang) and 14th July (Indramayu) in 2011.

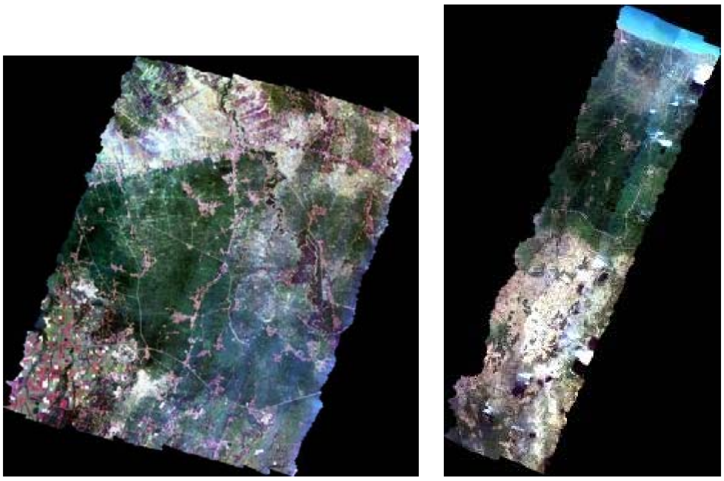


Figure 2 HyMap image acquired in 2008 (left: Subang, right: Indramayu)

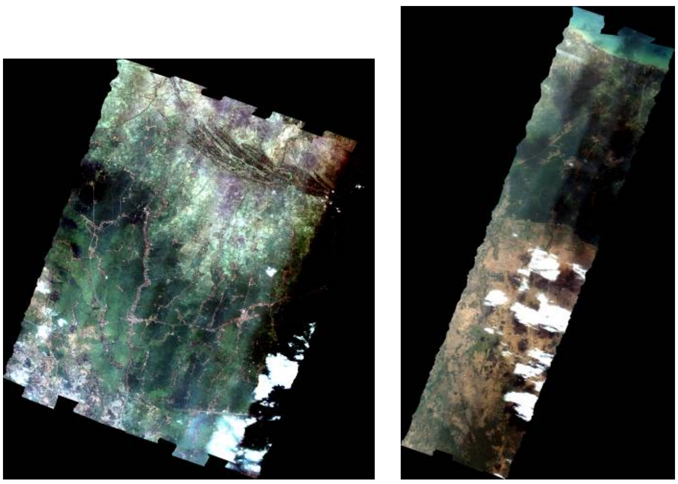


Figure 3 HyMap image acquired in 2011 (left: Karawang, right: Indramayu)

We also recorded the growth stage information at about 100 quadrats which we set in airborne observed area. We recorded them by using IRRI (International Rice Research Institute) definition. IRRI defined 3 growth phases (Vegetative, Reproductive and Ripening) and 9 growth stages (Seedling, Tillering, Stem elongation, Panicle initiation to booting, Heading, Flowering, Milk grain, Dough grain and Mature grain). Table 1 shows the comparison table between IRRI definition and our definition and Figure 4 shows the mix growth stage in observed area.

Table 1 Growth stage definition (Definition by IRRI)

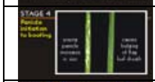
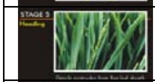
	Growth phase name	Growth stage name	Class Name
	IRRI		Our definition
	Vegetative	Seedling	Vegetative early (Veg_early)
		Tillering	Vegetative middle (Veg_mid)
		Stem elongation	Vegetative late (Veg_late)
	Reproductive	Panicle initiation to booting	Reproductive early (Rep_early)
		Heading	Reproductive middle (Rep_mid)
		Flowering	Reproductive late (Rep_late)
	Ripening	Milk grain	Ripening early (Rip_early)
		Dough grain	Ripening middle (Rip_mid)
		Mature grain	Ripening late (Rip_late)

Figure 4 Mix growth stage in observed area
(Left-to-Right: Vegetative late, Reproductive phase, Ripening early, Ripening late)

Regarding the 2008 measurement data, we set up 101 samples (data sets) with 7 growth stages and 116 bands. Ten bands are deleted from original observed HyMap data (126 bands) because of the influence of water absorption in the atmosphere. Unfortunately, we could not measure the data of Rep_mid and Rip_mid, so we classify these data sets into 7 growth stages except for Rep_mid and Rip_mid.

Table 2 Number of samples

Veg_early	Veg_mid	Veg_late	Rep_early	Rep_mid	Rep_late	Rip_early	Rip_mid	Rip_late
20	21	23	22	0	4	5	0	6

3. METHOD

The large number of spectral bands acquired by the hyperspectral sensor gives us much more spectral information for observed objects. But, as is often the case, compared to the number of spectral bands, those of ground measurement samples are relatively small. It often leads overfitting problems. Regularization is one of the most promising solutions for this problem, and there are some successful studies of the hyperspectral application regularized discriminant analysis (RDA) (Bandos *et al.*, 2009), Sparse Regularization (Yoshida *et al.*, 2011).

In this study, we selected the SLDA (Clemmensen *et al*., 2008), famous machine learning technique and one of the supervised methods for classification. Generally, linear model such as Linear Discriminant Analysis (LDA) is easy to interpret and shows us the effective bands information. Additionally, it is simple implementation for Multi-class classification. But, when we make a model with small number of samples by using LDA, we sometimes face overfitting problems, too. On the other hands, "Sparse" means a model with a low number of non-zero parameters and SLDA is a simple model with smaller number of bands which causes less overfitting problems and achieves better accuracy. Also, effective bands are selected automatically by optimization calculation.

Because of the large number of spectral information and relatively small number of samples, effectual band selection methods from machine learning technique are particularly useful. We applied SLDA, one of machine learning methods, for growth stage classification of paddy and validate the advantage of this approach.

4. RESULT AND DISCUSSION

4.1 Result

We set up the 101 data sets of reflectance data which has already deleted ten atmospheric absorption bands from original observed data (126 bands) and pre-processed geometric, atmospheric and BRDF correction. Using these data sets, we developed a growth stage classification model by SLDA and applied this classification model to HyMap image.
The results of the growth stage classification by using SLDA are shown in Figure 5. Both images show the large growth stage trend from mountain side (south) to the sea side (north). It also shows not only the big trend but also the detail of it such as aqua "Rip_early" in the center of left image. The image of Indramayu (Fig.5 left) shows yellow-green "Reproductive early" and blue "Rip_late" is mainly spread over center and bottom of it. On the other hand, in the image of Subang (Fig. 5 right), orange "Vegetative_mid" and red "Vegetative_late" are mainly spread. This means growth stage of paddy in Subang is delayed compared with Indramayu. These results are consistent with the record of ground measurement.

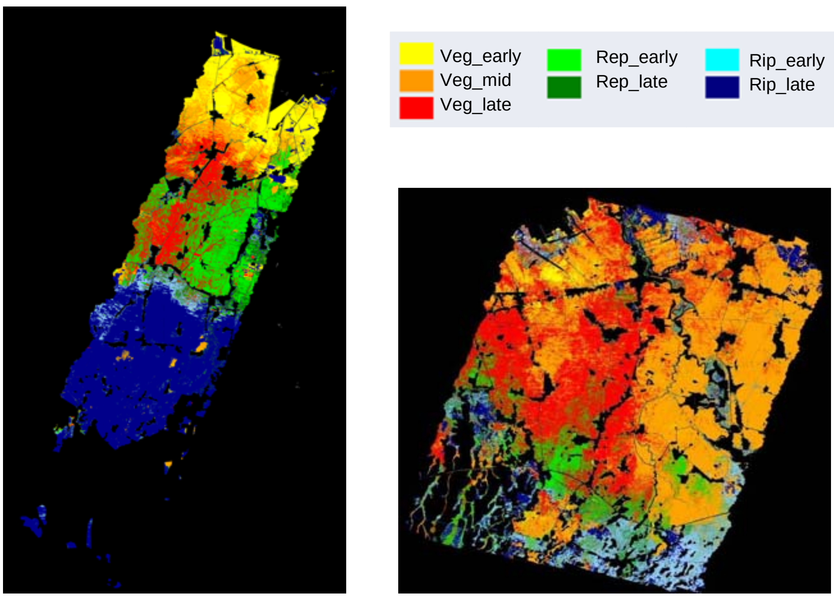


Figure 5 Classified HyMap image (Left: Indramayu, Right: Subang)

4.2 Evaluation

We checked the accuracy by comparing the result of HyMap data classification to ground measurement record data in each quadrat during field campaign. The result of growth stage classification by SLDA is summarized in Table 3. As shown in Table 3, the overall classification accuracy is 97.0% (98/101). Table 3 show the corrigenda of classification when we use all sample data and 2 of 3 misclassification data is around correct growth stages. It shows SLDA has the strong power for growth stage classification. We also conducted 4-fold cross validation (CV) 1000 times in order to evaluate the generalization capability of models and a mean classification accuracy of 4-fold CV is 85.3%.
We will continue this study by using 2011 HyMap data to validate the robustness of this method for time difference and area difference for future operational use to contribute to the national food security in Indonesia.

Table 3 Comparison table predicted data with measured data

		Predicted Class						
		veg_early	veg_mid	veg_late	rep_early	rep_late	rip_early	rip_late
measured Class	veg_early	20	0	0	0	0	0	0
	veg_mid	0	21	0	0	0	0	0
	veg_late	0	0	22	1	0	0	0
	rep_early	0	0	1	21	0	1	0
	rep_late	0	0	0	0	4	0	0
	rip_early	0	0	0	0	0	4	0
	rip_late	0	0	0	0	0	0	6

5. CONCLUSION

In this study, we applied SLDA, famous machine learning technique and one of the supervised classification methods, for growth stage classification of paddy and found SLDA can classify the growth stage by quite good accuracy (97.0%) without overfitting problems. This indicates that comparing with the traditional method such as SAM, etc., the SLDA has much potential ability for analyzing the hyperspectral data. The information provided by this method can contribute to the national food security in Indonesia.

Acknowledgements

This work was performed as part of a collaborative research project between ERSDAC (Japan) and BPPT (Indonesia).

References

Bandos et al., 2009. Classification of hyperspectral images with regularized linear discriminant analysis, IEEE Trans. Geosci. Remote Sens., vol. 47, 862 - 873
Clemmensen et al., 2008. Sparse discriminant analysis, Technical report, IMM, Tech. Univ. Denmark, 2008.
Kobayashi et al., 2009. Development of Hyperspectral Data Utilization Technology for Paddy Field in Indonesia, ACRS 2009
Kruse et al., 1993. The Spectral Image Processing System (SIPS) - Interactive Visualization and Analysis of Imaging spectrometer Data. Remote Sensing of Environment vol.44: 145 - 163.
Uchida et al., 2010 Development of Hyperspectral Data Utilization Technology by Using Data Mining Method for Paddy, West Java, Indonesia, ACRS 2010
Yoshida et al., 2011 A Methodology of Forest Monitoring From Hyperspectral Images With Sparse Regularization, IGARSS 2011