J. Trop. Agric. and Fd. Sc. 36(2)(2008): 000–000

Estimation of rice plant population using digital image processing technique

(Penganggaran populasi tumbuhan padi menggunakan teknik pemprosesan imej digit)

C.C. Teoh*, A.R. Mohd Syaifudin*, O. Muhamad Isa* and B.H. Abu Bakar*

Key words: rice plant population, estimation, image classification, statistical analysis

Abstract

Monitoring of rice plants population density is important for crop setting and fertilizer management to achieve high target yield. Currently, the population density is determined by manually counting the tiller number of total rice plants in a 25 cm x 25 cm square frame. Generally, several random sampling locations of a paddy plot are selected to perform tiller counting. This is time consuming, labour intensive and costly. An automatic counting tiller number method using digital image processing technique was introduced to overcome the problem. PCI image processing software was used to process 113 rice plant digital images obtained from MADA, Bukit Besar, Kedah paddy plots. The captured images were classified into plant and non-plant regions by image processing technique. The area of plant region in frame of each classified image was calculated by the software and used to correlate with tiller number using linear regression analysis. The relationship analysis result showed that area of plant has high relationship with tiller number with correlation coefficient value of 0.8328. A linear model was developed to estimate tiller number in the analysis. The model was verified by 100 rice plant digital images obtained from FELCRA, Seberang Perak, Perak in terms of tiller number estimation accuracy. The verification result showed that, the model capable to estimate tiller number with 92.17% average accuracy. As a result, the image processing technique is practical, feasible and effective in estimating tiller number for monitoring of rice plant population density.

Introduction

Rice is an essential plant for a large part of the world's population especially in Asia, because it is the principal staple food. Production of rice also is an important source of employment and income for rural people in most of Asia. Asia's hot and humid climate during the long and heavy monsoon season and the fertile land along river basins of the major deltas that are regularly flooded, provide most favourable agro-ecological environment for rice cultivation (Hossain 2003). According to Department of Agriculture Malaysia, rice plantations in Peninsular Malaysia covered an area of 670,000 ha in 2007.

Malaysian rice fields have a high degree of variability in soil fertility, soil topography, soil moisture, weed, pest and diseases infestation that affect crop yield. Field operations such as the mechanical seed broadcasting is one of the rice field

*Mechanization and Automation Research Centre, MARDI Headquarters, Serdang, P. O. Box 12301, 50774 Kuala Lumpur, Malaysia
Authors' full names: Teoh Chin Chuang, Mohd Syaifudin Abdul Rahman, Muhamad Isa Othman and Badril Hisham Abu Bakar
E-mail: cchin@mardi.gov.my

variabilities where patterns of seeds distribution in field are non-uniform (Chan and Rukunuddin 2003).

Hence, monitoring rice plants population at the different growing stages in the paddy field is essential for estimating crop yield and rice production. This information is important for the local authorities to manage and adjust the rice distribution within the nation.

Remote sensing technology has been widely used to provide timely and accurate information about crop distribution, area planted and potential production for decision makers. Studies specifically focus on rice monitoring using Synthetic Aperture Radar (SAR) technologies have been reported by Shao et al. (2001). For example, monitoring of rice crop growth from space using the ERS-1 C-band SAR (Kurosu et al. 1995); identification of rice fields using multi-temporal ERS-1 C-band SAR (Kurosu et al. 1997); rice crop mapping and monitoring using ERS-1 data based on experiment and modeling results (Le et al. 1997); multi-temporal RADARSAT data for rice classification in Zhaoqing, Guangdong Province (Liu et al. 1997); estimation rice growth stage using RADARSAT data (Shao et al. 1997a); and evaluation of SAR image for rice monitoring and land cover mapping (Shao et al. 1997b).

The reports showed that the SAR technologies produced very promising and encouraging results for rice area calculation and estimation of the life span of the rice, which are important pieces of information in order to produce a production forecast. A number of rice yield estimation models mostly rely on meteorological data were also reported in the literature. However, the meteorologically based yield models are restricted locally as they are based on empirical observations.

In Malaysia, a decision support system called CREST-FERTO for crop setting and fertilizer management to achieve sustainable target yield performance in precision rice farming has been developed by Othman et al. (2005). According to the authors, the CREST-FERTO package was developed to enhance high yield performance in rice production system to achieve 10.0 t/ha yield target. The package involved development of specifications for ideal crop setting and fertilizer management option recommendations to achieve specific target. Seedling management is one of the important steps in the package to ensure achievement of target population at critical growth stages. Currently, monitoring of rice plant population in the package is performed by counting the tillers of the rice plant in 25 cm x 25 cm sampling frame, manually. The manual counting of the tillers is time consuming, requires high labour input and very tedious.

This paper proposes a proper counting tiller system based on image processing technique for estimating rice plant population density.

Materials and methods

Image sampling

A total of 113 and 100 random rice plant image samples at 30 days growth stage were obtained from MADA at Bukit Besar, Kedah (26 October 2004) and FELCRA at Seberang Perak, Perak (14 April 2005), respectively. The images were captured by digital colour camera with a proper control height of 1 m from camera to ground. The images from MADA were separated into 19 groups based on tiller number and used to develop a model for estimating rice plant population. Whereas, the other images from FELCRA were used to verify the model.

Image processing

The PCI software was used to calibrate all the image samples into 5.84 mm/pixel resolution. Each calibrated image was classified into plant and non-plant regions. The plant area (cm^2) in the sampling frame was calculated by the software.

The procedure of the classification basically consists of training sample collection, band selection and classification steps. Plant and non-plant classes of training samples were collected and red (R), green (G) and blue (B) bands were selected to perform classification using minimum distance to means classifier. The minimum distance to means algorithm shown in the equation below is commonly used for classifying data because it was based on the simple computations to classify an unknown pixel.

$$d_c = \sqrt{(BV_{ij1} - \mu_{c1})^2 + (BV_{ij2} - \mu_{c2})^2 + \ldots\ldots + (BV_{ijk} - \mu_{ck})^2}$$
$$= \sqrt{\sum_{k=1}^{n} (BV_{ijk} - \mu_{ck})^2}$$

where

BV_{ijk} = Brightness value in a row i, column j, of band k
μ_{ck} = Mean vectors for class c measured in bands k
c = 1, 2, 3,…, m number of classes
k = 1, 2, 3,…, n number of bands

It uses mean vectors for each class in each band μ_{ck}, from the training data to perform classification. A pixel of unknown identity can be classified by computing the Euclidian distance, d between the value of each unknown pixel (BV_{ijk}), and each of the μ_{ck} in their n bands, respectively (Jahne 1991; Lo and Yeung 2002). The unknown pixel was assigned to the class 1, if the calculated d_1 is smallest compared to $d_2, d_3, \ldots, d_m$. It should be obvious that any unknown pixel in each image sample will definitely be assigned to plant or non-plant using this algorithm.

Statistical analysis
A linear regression analysis of SAS 8.0 software was used to analyse the relationship between plant area and tiller number in the frame. A linear model was developed to predict the number of tillers (dependent variable) from the calculated plant area (independent variable). A 95% confidence interval graph was used to prove the linear model adequately fits the data. The confidence region was obtained using the formula: (Estimate) ± t * $S(D)^{1/2}$, while the prediction region was calculated by the formula: (Estimate) ± t * $S(1+D)^{1/2}$, where S is the standard deviation (estimated by mean square error), t-value from Student's t-distribution depending on the level of confidence and degree of freedom, and D depending on each value of the regression line.

In statistical analysis, the validity of the results requires fulfilment of certain assumptions about data. Especially when the linear model is assumed as appropriate, the residuals must be independent and normally distributed with the same variance everywhere. A plot of residuals versus independent values was used to detect the violation of assumptions. The residual was determined by the difference between actual and estimated tiller number. The plot also used to observe outliers or erroneous problem.

Results and discussion
Plate 1 shows one of the random rice plant image samples obtained from MADA. The image was classified into plant and non-plant regions (*Plate 2*) using collected training samples, RGB bands and minimum distance to means classifier.

Area (cm^2) of the plant region in the frame of the classified image was calculated by the software and used to analyse relationship with tiller number. *Table 1* presents the average areas of the

Plate 1. Rice plant image

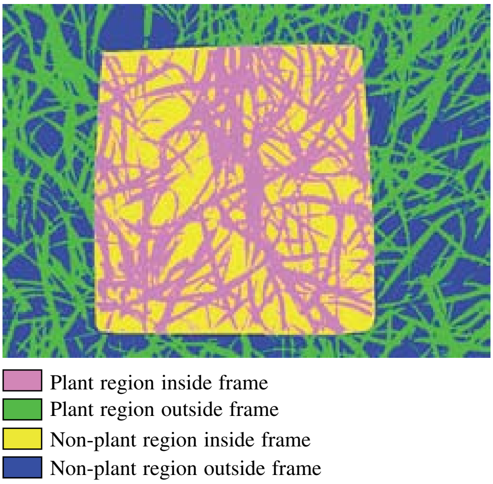

Plate 2. Classified image

Table 1. Plant region area of classified image and tillers number of rice plant in 25 cm x 25 cm sampling frame

Group	Plant region area (cm^2)	Tillers (no.)
1	373.06	38
2	409.28	40
3	391.44	41
4	418.21	42
5	445.47	43
6	436.9	44
7	447.15	45
8	432.2	46
9	443.33	47
10	439.92	48
11	447.03	49
12	450.48	50
13	466.23	51
14	474.88	52
15	449.22	53
16	493.55	54
17	463.28	55
18	472.06	56
19	515.28	63

plant region and the tiller number for each individual group obtained from MADA. Mean (μ) and standard deviation (*s*) for the average areas were 445.73 and 33.37, respectively while, for tiller number were 48.26 and 6.37, respectively. The calculated 95% confidence interval for μ of the average areas and the tiller number were ranged from 429.65 to 461.82 and 45.19 to 51.33, respectively.

The best fit line that described the relationship between the average areas of plant and tiller number is shown in *Figure 1*. The plot indicated a good dependency exits between area of plant and tiller number with coefficient of correlation (r^2) of 0.8328. The regression model describing this dependency is:

Tiller number = 0.1743 * Plant area – 29.445

The 95% confidence and prediction intervals result gives a good confidence for prediction of tiller number. The prediction bands are wider than the corresponding confidence bands to allow for the fact that the linear model predicting the value of a random

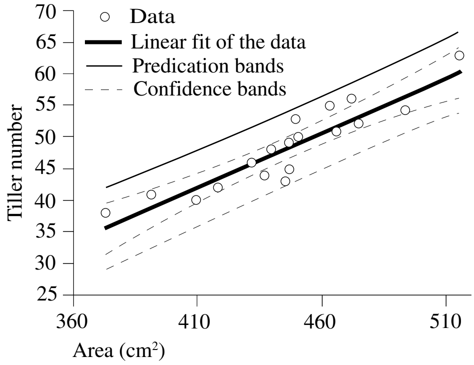

Figure 1. The dependency between plant area and tiller number in 25 cm x 25 cm sampling frame

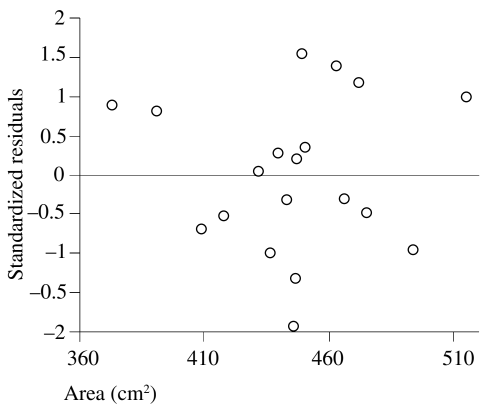

Figure 2. The residual (error) distribution

Table 2. Accuracies of the linear model to estimate number of tillers in 25 cm x 25 cm sampling frame for 100 image samples

Samples	Actual tiller number	Estimated tiller number	Accuracy (%)	Samples	Actual tiller number	Estimated tiller number	Accuracy (%)
1	60	53.87	89.78	51	49	48.69	99.37
2	62	50.44	81.36	52	46	56.63	76.88
3	53	46.92	88.52	53	43	43.21	99.51
4	51	53.23	95.63	54	45	47.74	93.92
5	57	56.35	98.85	55	47	51.44	90.55
6	64	53.09	82.96	56	45	45.55	98.78
7	41	48.35	82.08	57	50	48.05	96.11
8	45	48.98	91.14	58	49	51.88	94.12
9	47	47.57	98.78	59	48	52.88	89.84
10	50	59.47	81.06	60	48	56.39	82.52
11	52	55.25	93.76	61	45	47.12	95.28
12	49	56.58	84.52	62	47	50.32	92.93
13	47	43.04	91.56	63	51	52.74	96.6
14	45	43.99	97.76	64	50	44.82	89.63
15	50	43.3	86.6	65	52	49.35	94.9
16	48	49.12	97.66	66	48	49.49	96.89
17	49	59.31	78.95	67	49	43.21	88.19
18	44	48.29	90.25	68	50	49.55	99.09
19	40	44.51	88.73	69	40	44.16	89.6
20	42	45.22	92.34	70	46	42.94	93.36
21	52	44.13	84.86	71	48	43.34	90.29
22	42	44.96	92.96	72	41	43.01	95.1
23	52	50.83	97.75	73	43	38.76	90.15
24	49	45.29	92.43	74	40	44.76	88.09
25	44	42.53	96.66	75	50	50.54	98.93
26	47	42.99	91.46	76	46	46.57	98.76
27	48	50.45	94.89	77	44	41.34	93.96
28	51	46.63	91.44	78	48	47.04	98
29	61	50.98	83.58	79	49	49.43	99.13
30	45	42.06	93.46	80	44	42.13	95.76
31	46	51.49	88.07	81	44	44.41	99.07
32	64	56.16	87.76	82	50	42.18	84.35
33	43	48.61	86.96	83	42	40.77	97.07
34	54	52.93	98.02	84	42	43.49	96.45
35	53	50.46	95.21	85	42	48.64	84.18
36	44	46.2	95	86	48	54.18	87.13
37	53	48.4	91.32	87	43	44.25	97.1
38	51	48.78	95.65	88	47	43.78	93.15
39	51	52.54	96.98	89	44	44.92	97.9
40	53	48.43	91.37	90	47	48.49	96.84
41	55	54.43	98.97	91	48	55.22	84.96
42	56	52.91	94.47	92	51	46.38	90.95
43	50	59.41	81.18	93	45	44.68	99.28
44	46	50.33	90.58	94	45	44.4	98.67
45	48	52.39	90.86	95	44	45.86	95.77
46	44	49.2	88.18	96	41	44.59	91.23
47	54	44.53	82.46	97	44	41.49	94.3
48	43	43.32	99.25	98	49	46.74	95.4
49	50	54.75	90.5	99	43	47.59	89.32
50	54	55.76	96.74	100	44	50.14	86.05

variable rather than estimating a parameter. The 95% confident interval for intercept and slope of linear model were – 47.31 to – 11.58 and 0.134 to 0.214, respectively. *Figure 2* illustrates a plot of residuals versus independent values. It can be seen that the plot shows homogeneous error variances and does not show any pattern. This gave confidence regarding the normal distribution of the data. The result from the plot also showed that there are no extremely large residuals (and hence no apparent outliers) and that there is no trend in the residuals to indicate the linear model is inappropriate.

The developed linear model was verified by a set of image samples taken from FELCRA. The images were processed by following training sample collection, band selection and classification steps. The area of plant region in the frame of classified images was calculated and inputted to the model for estimating tiller number. *Table 2* summarizes actual tiller number, estimated tiller number and accuracy estimation results of the image set. Mean (μ) and standard deviation (s) for actual tiller number were 48.03 and 5.05, respectively. The calculated of 95% confidence interval for μ of tiller number were ranged from 47.03 to 49.03. After comparing the actual and estimated tiller number the average accuracy of the model to estimate tiller number was 92.17%. Thus, tiller number can be estimated with a high degree of accuracy by the image processing technique.

Conclusion

This paper describes an image processing technique for estimating the tiller number to predict rice plant population. Processing rice plant image provided a simple and effective means for counting the plant area in the square frame. The plant area has high relationship with tiller number with r^2 value of 0.8328. In the statistical analysis a linear model has been developed using image samples of MADA for estimating tiller number based on counted plant area. The model was verified by image samples of FELCRA in terms of tiller number estimation accuracy. Results showed that the model was capable to estimate tiller number with average accuracy of 92.17%. As a result, image processing technique can be used to replace the manual counting of tiller number to determine rice plant population for Malaysia paddy fields. This new technique gives benefit by increasing speed, accuracy and consistency of tiller number counting for crop setting and fertilizer management to achieve high target yield.

References

Chan, C.W. and Rukunuddin, I. H. (2003). Developments in the use of precision agriculture technology in rice production. *Proc. of International rice conference*, 13–17 Oct. 2003, Alor Star, Kedah, p.133–142. Serdang: MARDI

Hossain. M. (2003). Rice production and trade liberalization: Global perspective. *Proc. of International rice conference*, 13–17 Oct. 2003, Alor Star, Kedah. p. 7–19. Serdang: MARDI

Jahne. B. (1991). *Digital image processing*, p. 219–230. New York: Springer-Verlag

Kurosu, T., Fujita, M. and Chiba, K. (1995). Monitoring of rice crop growth from space using the ERS-1 C-band SAR. *IEEE Transactions on Geoscience and Remote Sensing 33(4):* 1092–1096

—— (1997). The identification of rice fields using multi-temporal ERS-1 C-band SAR data. *International Journal of Remote Sensing 18(14):* 2953–2965

Le Toan, T., Ribbes, F., Floury, N., Wang, L., Kong, J., Kurosu, T. and Fujita, M. (1997). Rice crop mapping and monitoring using ERS-1 data based on experimental and modeling results. *IEEE Transactions on Geoscience and Remote Sensing 35(1):* 41–56

Liu, H., Shao, Y. and Wang, C. (1997). Multitemporal RADARSAT data for rice classification in Zhaoqing, Guangdong Province. *Land Resource Remote Sensing 4:* 1–6

Lo, C.P. and Yeung, A.K. (2002). *Concepts and techniques of geographic information system.* Upper Saddle River, NJ: Prentice-Hall

Othman, S., Abd Razak, H. and Rukunudin, I.H. (2005). CREST-FERTO: A decision support system for crop setting and fertilizer management to achieve sustainable target yield performance in precision rice farming.

Proc. of the National conference on AgrICT 2005, 27–28 Sept. 2005, Kuala Lumpur. Serdang: MARDI

Shao, Y., Fan, X., Liu, H., Xiao, J., Ross, S., Brisco, B., Brown, R. and Staples, G. (2001). Rice monitoring and production estimation using multitemporal RADARSAT. *Remote Sensing of Environment 76:* 310–325

Shao, Y., Fan, X., Wang, C. and Liu, H. (1997a). Estimation rice growth stage using RADARSAT data. *Proc. of Geoscience and remote sensing,* Vol 4, Aug. 1997, Singapore. NY: IEEE

—— (1997b). Evaluation of SAR image for rice monitoring and land cover mapping. Paper Presented at Geomatics in Era of RADARSAT, May 1997, Ottawa, Canada. Organiser: Canadian Aeronautics and Space Institute

Abstrak

Pemantauan kepadatan populasi tumbuhan padi penting untuk penetapan tanaman dan pengurusan baja supaya mencapai hasil padi yang tinggi. Pada masa kini kerja menentukan kepadatan populasi dijalankan secara manual dengan membilang jumlah anak pokok di dalam 25 cm x 25 cm rangka segi empat sama. Biasanya, pengiraan anak pokok dijalankan di beberapa kawasan tertentu yang dipilih secara rawak di dalam satu petak padi. Ini akan memakan masa, memerlukan tenaga kerja yang banyak dan kos yang tinggi. Satu kaedah automatik pengiraan bilangan anak pokok dengan menggunakan teknik pemprosesan imej diperkenalkan untuk mengatasi masalah tersebut. Pengisian pemprosesan imej (PCI) digunakan untuk memproses 113 keping imej digit tumbuhan padi yang diperoleh dari petak padi MADA, Bukit Besar, Kedah. Imej yang dirakam itu dikelaskan kepada bahagian tumbuhan dan bukan tumbuhan oleh teknik pemprosesan imej. Keluasan bahagian tumbuhan di dalam rangka bagi setiap pengelasan imej dikira oleh perisian dan dihubung kait dengan bilangan anak pokok menggunakan analisis regresi linear. Keputusan analisis perhubungan menunjukkan bahawa keluasan tumbuhan mempunyai hubungan yang amat rapat dengan bilangan anak pokok dengan nilai pekali korelasi 0.8328. Satu model linear juga dibangunkan daripada analisis tersebut untuk menganggarkan bilangan anak pokok. Model itu disahkan oleh 100 keping imej digit tumbuhan padi yang diperoleh dari FELCRA, Seberang Perak, Perak berdasarkan kejituan penganggaran bilangan anak pokok. Keputusan pengesahan menunjukkan bahawa model itu mampu menganggarkan bilangan anak pokok dengan purata kejituan 92.17%. Oleh itu, teknik pemprosesan imej didapati praktikal, boleh dipercayai dan berkesan dalam penganggaran bilangan anak pokok untuk pemantauan kepadatan populasi tumbuhan padi.

Accepted for publication on 12 August 2008