Longan Yield Estimation Using Image Processing Technique and Multiple Regression Model for Decision Support Systems

Vorraveerukorn Veerachitt[1*] and Chanchai Sangchyoswat[2]

[1]Multiple Cropping Centre, Faculty of Agriculture, Chiang Mai University, Chiang Mai 50200, Thailand

[2]Department of Plant Science and Natural Resources, Faculty of Agriculture, Chiang Mai University, Chiang Mai 50200, Thailand

**Corresponding author. E-mail: vorraveerukorn.v@cmu.ac.th*

ABSTRACT

Longan production system requires a quick and accurate estimation technique for planning. This paper presents the development of longan yield estimation method. This research identified types of land mapping units (LMUs) for longan production in Phrao district, Chiang Mai Province using spatial data, consisted of age of longan trees, irrigation systems and land gradient. The LMUs types suggested the determination of sampling areas for collecting data and taking photos of longan bushes. The visible wavelengths (RGB), and a range of relationship barrier patterns were determined for the conditions of image processing. The proportion of fruits to bushes was one factor among others gathered from farmer interviews. This information was applied to develop the longan yield model using a multiple regression equation. The equation could explain the relationship of various factors with the accuracy of 89% and the R2 of 0.79. The longan yield in each LMU was calculated by this equation also to generate an estimated yield map for describing the distribution of longan production at the crop scale. This information is crucial for the planning of longan production management prior to the harvesting. The planning can benefit to longan farmers for improving the efficiency of resources allocation and utilization.

Key words: Longan, Image processing, Longan yield estimation, Decision support systems

INTRODUCTION

Estimation of longan yields can benefit to stakeholders to handle their management and planning efficiently. Recently, the Office of Agricultual Economics has estimated longan yield based on field survey sampling and statistics. Random sampling method was implemented to conduct crop cutting, which required skills and time consuming (OAE, 2009). Therefore, this study aims to develop an accurate and efficient approach of longan yield estimation. The approach implemented spatial information for planning and managing the production systems comprehensively. Digital photos of longan fruits on the bushes were taken, then, the images were analyzed for estimating the amount of fruits to be harvested. This approach is currently applied to estimate various crop productions. For example, the image processing of apple fruits were conducted by Stajnko et al. (2004) in order to count the number of apple fruits. Moreover, Okamoto and Lee (2009) and Bulanon et al. (2009) also reported that to the image processing was also applied for identifying a number of citrus fruits. The result of this research was satisfied.

Another approach implemented popularly to estimate crop productions is the regression model, representing the relationship between independent factors and dependent variable as the yield based on statistical data. The model indicates a number of factors involved and the impacts of these factors on the predicted yields (Jaikla et al., 2008; Merdun et al., 2006; Schaap and Leij, 1998). This approach has been widely applied in agronomy, such as the rice production model (Jaikla et al., 2008), corn and soybeans production models (Li et al., 2007; Prasad et al., 2006) and wheat production model (Dadhwal and Sridhar, 1997).

Consequently, in this study, integration of the two approaches has been implemented for estimating longan yields. The output model can explain a number of factors in longan production system. This tool can also contribute to resources allocation and utilization related to longan production.

MATERIALS AND METHODS

Site of study

The study area is located on Phrao district, north of Chiang Mai province, with the area of 2,021 km^2. The area is significantly the most longan productive areas in Chiang Mai. The longan planting areas have increased rapidly from 884 ha in 1999 to 5,464 ha in 2008 (DOAE, 2009).

Process of study

This study implements three technical approaches including Geographical Information Systems (GIS), image processing analysis and multiple regression model as the tool for estimating longan yield in different areas. Land Mapping Units (LMUs) are created for classifying farmer samples into groups. Moreover, image processing technique, was employed to analyze longan sampling photos for estimating the proportion of fruits to the bushes. Then multiple regression model was achieved using the factors from the previous process.

Land Mapping Units (LMUs) for data samplings

Land Mapping Units (LMUs) of longan production system were generated from the 3 spatial data layers. Firstly, the layer of longan tree ages was acquired from aerial Ortho-photos (with the resolution of 0.75 meter) associated with the land use map in 2007 from Land Development Department. The second layer was the water use system in particular the type of water used for longan production. Lastly, the layer of topographic slope generated from Digital Elevation model (DEM) was also employed in the system. The slope levels were classified into 3 classes covering 0-5%, 5-20% and greater than 20%. The spatial data were analyzed by spatial overlaying technique using ArcMap 9.3. The analysis procedure is shown in Figure 1.

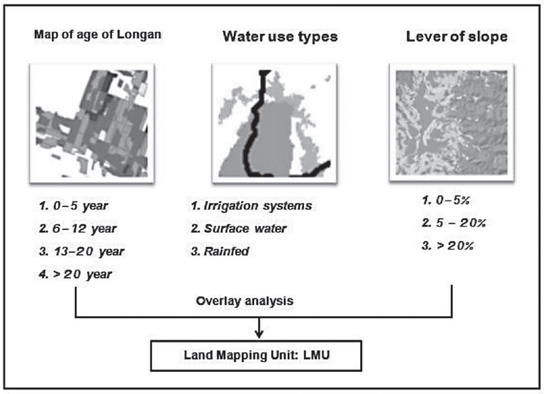


Figure 1. LMUs from overlaying spatial data.

Classification of longan fruits in the bushes

Thirty sampling orchards were selected from the LMU types and two fruit trees were randomly chosen in each orchard. Then, digital photos of longan bushes were taken to RGB colors in JPG format within 1x1 m^2 of the region of interest (ROI). Regarding a longan tree, 6 photos of the partial bush; lower, middle and upper were taken on each tree depending on the tree shape. Then, the 360 images were classified into groups according to the lighting conditions, and then a new data layer was created using ERDAS imagine program for proportion of fruits to a bush. The procedure is illustrated in Figure 2.

To classify separately longan fruits from the bush, the segmentation method was implemented. This method can define between fruits and the background distinguished by the difference of characteristics of objects in image (Cheng et al., 2001). In addition, the image classification analysis was operated. Wavelengths data in terms of Red (R), Green (G) and Blue (B) was analyzed along with OHTA color space form which is more detailed values in the grayscale (Yu-ichi and Takeo, 1980). Inclusion of the color space in the HIS (Hue, Intensity, and Saturation), was converted. The informative color spaces were analyzed for calculating the threshold between fruits and other parts using statistics of 6 training areas consisted of fruits, leave, branches/stem, sky, and ground. Then, the classification was operated automatically using "Model maker" in ERDAS program, resulted in proportion of fruits to the bush. The image processing procedure is shown in Figure 3.

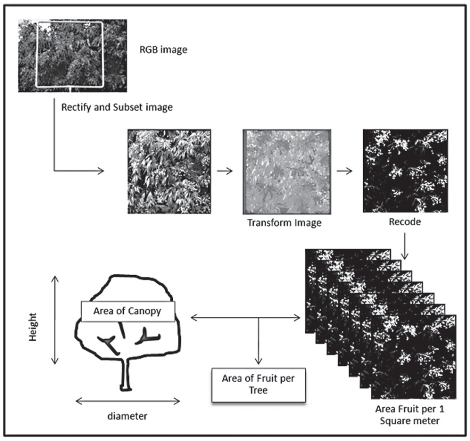


Figure 2. Proportional of fruit to a bush procedure.

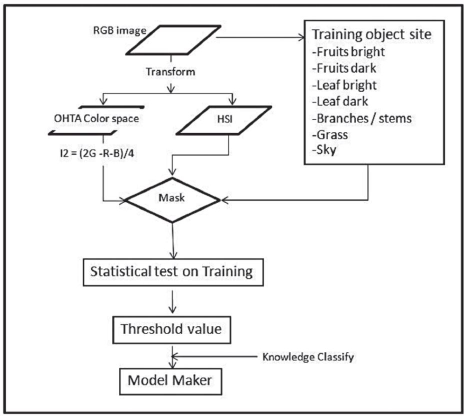


Figure 3. Image processing procedure.

Multiple regression model

A multiple regression model was applied to estimate longan yields. This model was developed to study the relationship between a set of proposed determining variables (X_i) and the quantity of longan yields (Y). In the model, the estimated longan yield is dependent variable ($\hat{y}$), while various factors are independent variables represented by x_1, x_2, ..., x_k. The general form of the regression equation is:

$$\hat{y} = \beta_0 + \beta_1 X_1 + \beta_1 X_1 + ... + \beta_k X_k \quad (1)$$

In this study, the equation of longan yields is applied as equation 2.

$$YieldPerM^2 = \beta_0 + \beta_1 (Canopy) + \beta_2 (FruitArea) + \beta_3 (No\_treePerM^2) + \beta_4 (No\_CultivarTreePerM^2) + \beta_5 (Age) + \beta_6 (DayLength) + \beta_7 (WaterSystem) + \beta_8 (SoilType) + \beta_9 (Manage) + \beta_{10} (No\_Land) + \beta_{11} (Cost) \quad (2)$$

Where *YieldPerM²* is a longan yield (kg/m²) as in equation 3,
Canopy is bush area (cm²),
FruitArea is the area of fruit per tree (m²) as equation 4,
No_TreePerM² is the number of trees in orchard area (trees/m²) as equation 5,
DayLength is duration date from taken photo to harvest (days),
No_CultivarTreePerM² is the number of the harvested tree per orchard area (trees/m²) as equation 6,
Age is the age of tree (years),
WaterSystem is the water use system type (irrigation system, surface water, and rainfed)
SoilType is a suitability level of cultivation soil (high suitability, medium suitability, and low suitability)
Manage is the level of management (high, moderate, and low)
No_Land is amount of land holdings (land), and
Cost is the cost using to production (Baht)

And the equations as:

$$Longan\ yield\ (kg/m^2) = (yield\ (kg)) / (spac\ X\ (m)\ space\ Y\ (m)) \quad (3)$$

$$Proportion\ fruit\ in\ bush = (fruit\ part\ area\ (m^2)\ x\ area\ of\ bush\ (m^2)) / 6 \quad (4)$$

$$Number\ of\ trees\ per\ square\ meter\ (tree/m^2) = (number\ of\ trees) / (planted\ area\ (m^2)) \quad (5)$$

$$Harvested\ trees\ per\ square\ meter\ (tree/m^2) = (Harvested\ trees) / (planted\ area\ (m^2)) \quad (6)$$

RESULTS

Land mapping units (LMUs)

According to a spatial overlay, the result shows 27 LMUs as Table 1.

Table 1. LMUs from spatial data overlaying.

LMU	Area (ha)	LMU	Area (ha)	LMU	Area (ha)
0-5 years, surface water, 0-5%	269	6-12 years, surface water, > 20%	116	13-20 years, surface water, >20%	16
0-5 years, surface water, 5-20%	254	6-12 years, groundwater, 0-5%	108	13-20 years, groundwater, 0-5%	26
0-5 years, surface water, > 20%	40	6-12 years, groundwater, 5-20%	109	13-20 years, groundwater, 5-20%	39
0-5 years, rainfed, 0-5%	105	6-12 years, groundwater, > 20%	27	13-20 years, rainfed, 0-5%	426
0-5 years, rainfed, 5-20%	133	6-12 years, rainfed, 0-5%	831	13-20 years, rainfed, 5-20%	440
0-5 years, rainfed, > 20%	147	6-12 years, rainfed, 5-20%	858	13-20 years, rainfed, > 20%	84
6-12 years, irrigation systems, 0-5%	54	6-12 years, rainfed, > 20%	272	21-30 years, surface water, 0-5%	22
6-12 years, surface water, 0-5%	802	13-20 years, surface water, 0-5%	336	31-30 years, rainfed, 0-5%	42
6-12 years, surface water, 5-20%	585	13-20 years, surface water, 5-20%	185	31-30 years, rainfed, 5-20%	26

Classification of longan fruits in the bush

Lighting conditions

Generally, the quality of image depends on condition of the natural light. The lighting conditions can be classified into 5 groups, including sun/less bright, sun/brightness, sun/very bright, no sun/less bright, and no sun/brightness. The examples of lighting conditions were demonstrated in Figure 4.


a) sun/less bright b) sun/brightness c) sun/very bright d) no sun/less bright e) no sun/brightness

Figure 4. Lighting conditions of images.

Data transformation

All images were obtained in RGB color. These can be transformed into various forms such as HIS. The forms rely on intensity of light that determines the object color. According to Ohta Yu-ichi's test, RGB format can be converted into another form called OHTA color space (Feng et al., 2008). In this study, the forms of (2 * R-G-B) /4 and (2 * G-R-B)/4 were used. In addition, the redness form: 3 * R-G+B was utilized in the process.

Classification threshold

The categories of RGB format were identified in each image with the descriptive statistics by average value. They were shown in Table 2.

Table 2. Summarizes the average values of 7 categories

Type	Red	Green	Blue	Hue	(2G-R-B)/4	(2R-G-B)/4	3R-G+B	3G-R+B
Fruits bright	169	164	139	119	5	8	204	183
Fruits dark	93	91	71	125	4	6	119	107
Leaf bright	166	197	136	144	23	0	165	289
Leaf dark	41	61	50	186	7	-7	13	90
Branches/stems	104	109	97	146	4	0	106	125
Grass	76	84	66	160	6	1	78	110
Sky	254	255	254	41	0	0	252	255

Furthermore, the color model which converted the RGB data into a related pattern was employed in the classification of longan fruits separated from the background. The model facilitated the clarity of longan fruits and enhanced the image related to the analysis. Regarding the process, 7 models of the color relating patterns were found as shown in Table 3.

Table 3. Conditions for the separation of longan fruit from other areas.

Condition of light	(R-G) - (G-R)	(R-B) - (G-R)	(2R-G-B)/4 - (2G-R-B)/4	R-G	R-B	HUE	3R-G+B
condition 1	> -2	> 15	> 0	-	-	-	-
condition 2	> 0	> 15	> 0	-	-	-	-
condition 3	> 15	> 40	> 5	-	-	-	-
condition 4	> 20	> 20	> 5	-	-	-	-
condition 5	> 0	-	-	> 0	> 30	-	-
condition 6	-	-	-	-	> 30	< 145	> 140

Image classification with threshold value

Based on the threshold value, the algorithm of classification was set up, and the tool of "model maker" in ERDAS program was implemented in this step. Firstly, threshold value was assigned appropriately for 'input image', resulted in the 'output image' which consisted of the longan fruits separated from the backgrounds. On the output image, longan fruit areas were calculated in a square centimeter using a table form. Then, all tables of fruits part area were combined to calculate a proportion of fruits to the bush. The procedure was operated automatically and readily with a high accuracy of the results. Regarding the accuracy assessment, the best accuracy could achieve to 92.3%, with the Kappa of 0.935. Furthermore, the average accuracy was higher than 80% and the average Kappa of 0.765. Some examples of the classification by threshold value were demonstrated in Figure 5.

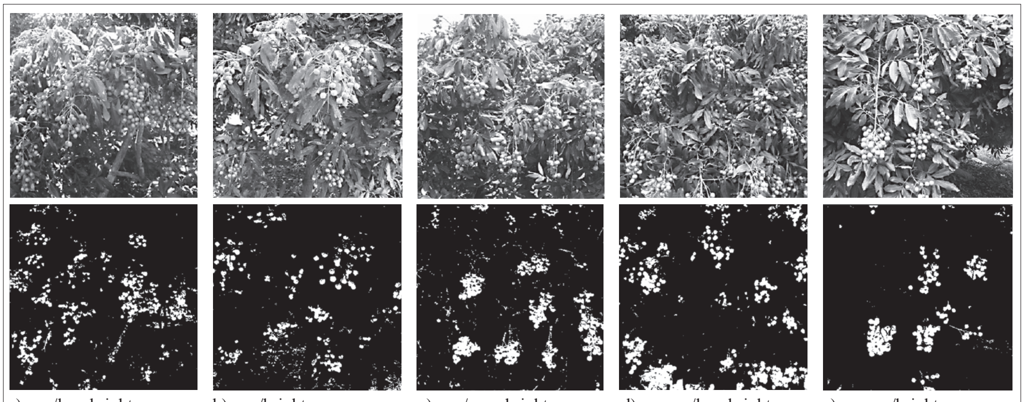


Figure 5. Results of classification by threshold value.

Multiple regression model

The regression model was carried out using SPSS statistic program version 17.0 (IBM Corporation 2010). This program facilitated for the factors into or out of the equation. The suitability was individually determined by partial correlation value (significant), which was lower than 0.05 to select some appropriate variables in the equation until no more factors could affect on the equation. This step represented by the value of R-square (R^2) and the *t*-Statistic at significance level 0.05. The estimating model is expressed in normal logarithm form as follows;

$$Ln\_YieldPerM^2 = 0.649\,(Ln\_FruitArea) - 1.095\,(Ln\_No\_TreePerM^2) + 1.152\,(Ln\_No\_CultivarTreePerM^2) - 0.859\,(Ln\_No\_Land) + 0.326\,(Ln\_WaterSystem) + 0.463\,(ln\_SoilType) \quad (7)$$

According to the regression model, there were 6 factors determining the quantity of longan yields. These factors were statistically acceptable ($P < 0.05$). The model represented relationship between independent and dependent variables with the R value of 89%. It represented all 6 factors affected to the longan yields of 79% (adjusted $R^2 = 0.79$), while the remains of 21% were affected from other factors. In addition, the error of this model was 0.54% when the model was implemented to estimate the yields. Durbin Watson's statistics was 1.37 which was considered the statistical test. This equation was applied to estimate yield validated with an observed data. The result closeness between the estimated and observe yields was illustrated in Figure 6.

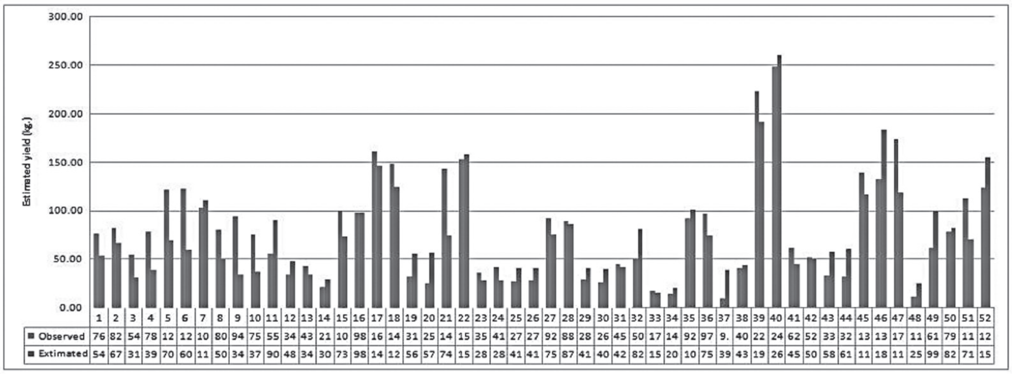


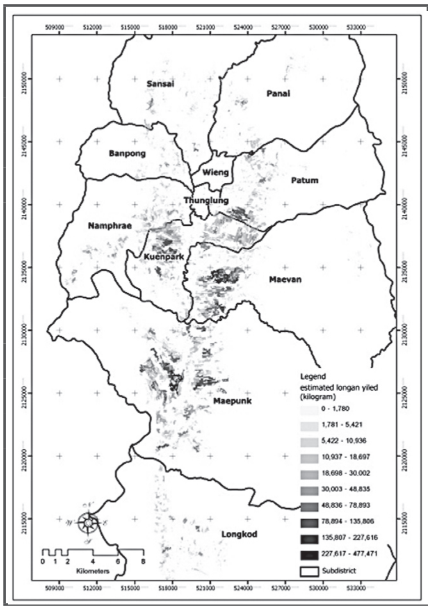
Figure 6. Comparing with estimated and observed yield.

The estimated longan yield was totally 3,912.5 kg (Figure 6), which was 4.93% lower than the observed yield (4,125.6 kg). The average percentage error of the equation is 12.73%.

Longan yield estimation

Practically, the model was applied to estimate longan yields in Phrao district. Average longan yields per hectare of each LMU types were estimated by this equation.

In general, longan yields fluctuate from year to year and possibly vary in each tree. Natural disasters, such as drought during growing and fruits setting periods can also affect yield stability. Therefore, it is possible that longan trees might not be able to provide yields at a full capacity in every year, particularly in the low productive year. Chiangchen (2005) revealed that in the low productive year, the yields dropped approximately to 65 percent from the average yields in normal situation. Considering the yield stability, this study assigned to test the yield tolerance. It was assumed that each orchard could produce the yields of 50 to 100%. We found that the average estimated yields of 60 to 80% were 899.64 to 1,049.58 kg/rai, respectively. These estimated yields were not significantly different from the observed yields. Interestingly, the estimated yields of 65 percent is closest to the observed yields . Therefore, the estimated yields of 65% was applied to create of the longan yield maps for the district. Then, the estimated longan yields of the district was compared to the official longan yield records authorized by the OAE (2009). The results were shown in Figure 7.



Subdistrict	Planted area[1]	Yield of model[2]	Planted area[3]	Estimated yield 09/10[4]	Estimated yield[5]
Thungluang	151	131,217	504	403,200	504,000
Namphrae	3,203	1,964,226	2,012	1,609,600	2,012,000
Banpong	891	719,370	1,178	942,400	1,178,000
Patum	3,573	1,905,752	2,034	1,627,200	2,034,000
Panai	1,055	835,457	1,379	1,103,200	1,379,000
Sansai	2,318	1,813,288	1,560	1,248,000	1,560,000
Kuenpark	7,969	4,979,807	6,087	4,869,600	6,087,000
Wieng	43	34,855	586	468,800	586,000
Maepunk	12,161	10,855,302	11,259	9,007,200	11,259,000
Maevan	6,096	5,511,310	5,540	4,432,000	5,540,000
Longkod	2,939	2,493,929	2,806	2,244,800	2,806,000
Total	40,400	31,244,516	34,945	27,956,000	34,945,000

[1]Planted area from GIS data
[2]calculate from equation
[3]Planted area report from office of agriculture of Phrao
[4]The estimated yield of government staff on 2009/2010
[5]Estimated yield base on good year

Figure 7. Map of estimated longan yiled in Phrao district.

DISCUSSION AND CONCLUSION

This paper presents the development of longan yield estimation and the procedure using Geo-informatics system. Image processing analysis was applied collaboratively with the multiple linear regression model to create the LMUs for grouping types of longan production and collecting data samplings (interviews and taking bush images). Threshold value was acquired to generate color format (RGB and a form of RGB color conversion) and applied for the image classification. It was enabling to calculate a proportion of fruits to the bush easily. The proportion value was one of important factors for the longan yields estimation. It was applied with the regression model using the factors assigned from interviewing, resulted in an equation to estimate longan yields and generate a map of yield estimation at the district and crop scales of the study area. The model can provide good accuracy and reliability for the longan yields estimation. It can be applied for the longan yields management and planning. It can also improve efficiency of the resources allocation and utilization for longan farmers.

This study has established as comprehensive guidelines for further advanced researches, such as comparison of this classification method to the other factors using multiple regression model for getting a better result.

ACKNOWLEDGEMENT

The authors would like to thank Thailand Research Fund (TRF) and Multiple Cropping Centre (MCC) for the financial support.

REFERENCES

Bulanon, D.M., T.F. Burks, and V. Alchanatis. 2009. Image fusion of visible and thermal images for fruit detection. Biosystems Engineering 103: 12-22.

Cheng, H.D., X.H. Jiang, Y. Sun, and J. Wang. 2001. Color image segmentation: advances and prospects. Pattern Recognition 34: 2259-2281.

Chiangchen, U. 2005. The decision support system for agricultural products over the Internet. Independent Study. Chiang Mai University, Chiang Mai, Thailand.

Dadhwal, V.K., and V.N. Sridhar. 1997. A non-linear regression form for vegetation index-crop yield relation incorporating acquisition date normalization. International Journal of Remote Sensing 18: 1403-1408.

DOAE. 2009. Crops in Chiang Mai in year 2008. (Online). Available: http://chiangmai.doae.go.th/reports/stat_plan/stat_plantproduction51-52.pdf (January 3, 2010).

Feng, G., C. Qixin, and N. Masateru. 2008. Fruit detachment and classification method for strawberry harvesting robot. International Journal of Advanced Robotic Systems 5: 41-48.

Jaikla, R., S. Auephanwiriyakul, and A. Jintrawet. 2008. Rice yield prediction using a support vector regression method. p. 29-32. In Proceedings of IEEE ECTI-CON 2008, Krabi, Thailand.

Li, A., S. Liang, A.Wang, and J. Qin. 2007. Estimating crop yield from multi-temporal satellite data using multivariate regression and neural network techniques. Photogrammetric Engineering & Remote Sensing 73: 1149-1157.

Merdun, H., C. Ozer, R. Meral, and M. Apan. 2006. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. Soil and Tillage Research 90: 108-116.

OAE. 2009. The major agricultural crops. (Online). Available: http://www2.oae.go.th/Prcai/area.php (january 3, 2010).

Okamoto, H., and W.S. Lee. 2009. Green citrus detection using hyperspectral imaging. Computers and Electronics in Agriculture 66: 201-208.

Prasad, A.K., L. Chai, R.P Singh, and M. Kafatos. 2006. Crop yield estimation model for Iowa using remote sensing and surface parameters. International Journal of Applied Earth Observation and Geoinformation 8: 26-33.

Schaap, M.G., and F. J. Leij. 1998. Using neural networks to predict soil water retention and soil hydraulic conductivity. Soil and Tillage Research 47: 37-42.

Stajnko, D., M. Lakota, and M. Hocevar. 2004. Estimation of number and diameter of apple fruits in an orchard during the growing season by thermal imaging. Computers and Electronics in Agriculture 42: 31-42.

Yu-ichi, O., and K. Takeo. 1980. Color information for region segmentation. Computer Graphics and Image Procsssing 13: 222-241.

none