Rice Cultivation and Harvest Date Estimation Using MODIS NDVI Time-series Data

Saran Suwannachatkul[1*], Teerasit Kasetkasem[1**], Kitti Chumkesornkulkit[1],
Preesan Rakwatin[2], Thitiporn Chanwimaluang[3], Itsuo Kumazawa[4]
[1]Faculty of Engineering, Kasetsart University, Bangkok 10900, Thailand
*g5514553442@ku.ac.th, **fengtsk@ku.ac.th, g5414552797@ku.ac.th
[2]Geo-Informatics and Space Technology Development Agency (GISTDA)
120 (Building B) Chaeng Wattana Road, Laksi District, Bangkok 10210, Thailand
preesan@gistda.or.th
[3]Image Technology Laboratory National Electronics and Computer
Technology CenterPathumthani, Thailand, 12120
thitiporn.chanwimaluang@nectec.or.th
[4]Imaging Science and Engineering Laboratory
R2-59, 4266 Nagatsuta-cho, Midori-ku, Yokohama, 226-8503, Japan
kumazawa@isl.titech.ac.jp

Abstract— Rice cultivation and harvest dates are very useful information since they are the key factors in accessing damaged rice field from various disasters. However, the actual field survey of these important dates over a large area wastes huge amount of resources. To solve this problem, this paper proposes a new approach to estimate rice cultivation and harvest dates by using 8-day composite normalized difference vegetation index (NDVI) derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data. In this work, we divide the rice growth states into 4 states, namely, nothing, growing, mature, and harvest states and applied them to the hidden Markov model (HMM). Then, we assign the state to the NDVI time-series data by using the Viterbi algorithm. By using these derived states, we are able to estimate the rice cultivation and harvest dates. The date estimation results are compared with the ground truth data to access the accuracy and we found the average cultivation dates and harvest dates to have errors of 15.48 days and 6.525 days, respectively.

Keywords—*MODIS, NDVI, Hidden Markov Model, Viterbi Algorithm*

I. INTRODUCTION

The rice phenology estimation from remote sensing data can provide useful information in many ways. For example, the severe flooding throughout the provinces of Northern, North-eastern and Central Thailand in 2011, Thai government issued a policy to subsidize the damaged rice fields, depending on the development stage of the rice. The rice phenology estimation can be used to assess the damage of the flood by acquiring the rice's growth in the days before flooding. Since the flooding covered a large area, field surveys would require huge resources.

By using satellite image such as the moderate resolution imaging spectroradiometer (MODIS), the manpower and time required can be reduced. The MODIS is a key instrument aboard the Terra and Aqua satellites. The Terra MODIS and Aqua MODIS scan the Earth's entire surface every 1 to 2 days and acquire data in 36 spectral bands. One MODIS product is the normalized difference vegetation index (NDVI). The NDVI time-series data are widely used in vegetative studies. It has been used to estimate crop yields[3][4], crop intensity[2] etc. However, most of the studies encountered a problem with the data qualities which were affected by cloud contamination and atmospheric variability.

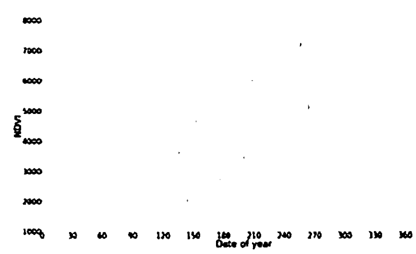


Fig.1 Raw NDVI time-series data of rain-fed rice in 2009 from Chaiyaphum province, Thailand. The NDVI value that drops irregularly is the error from various factors mostly by cloud contamination.

The MODIS satellite data have been widely used to monitor crops. The MODIS data include quality control flags to account for various image artefacts, including clouds and aerosols but the time-series normalized difference vegetation index (NDVI) data still contains some errors due to cloud contamination and atmospheric variability as in Fig.1. In order to estimate the crop phenology change from the NDVI, a continuous and smooth NDVI curve is required.

The study by Chumkesornkulkit et al.[1] proposed to use MODIS data for the cultivation date estimation on rain-fed and irrigated rice. They modelled the MODIS NDVI data as a triply modulated cosine function. Then, they used the extended Kalman filter (EKF) to smooth the data and estimate the mean, amplitude and phase parameters of the cosine function. The cultivation dates are estimated as the dates where the seasonal variation derived from the EKF is greater than a threshold after its minimum. Their model defines the noises as the additive parameter to their model and assumes that the NDVI data contains the same amount of noises in every period of time. However, according to Fig.1, the NDVI data in Thailand can be affected by cloud mostly in rainy season which is the cultivation period of the rice, but in other periods the effect of cloud is a lot less. This problem could influence the accuracy in cultivation date estimation.

The study by Zhao *et al.* [4] used the Savitzky-Golay filter to smooth the NDVI time-series data and compared them with a method using a double logistic function. After that, they used smoothed data to detect the vegetative to reproductive period, the emerged dates and harvest dates using the following characteristics, in which the vegetative to reproductive period are characterized by: (1) NDVI reaching to high-value range and (2) the curve reaching a near-saturation range and the first derivative of the NDVI staying near zero over these periods. For the emerged date, three facts have been considered: (1) the date is before the silking date / setting pods; (2) the second derivative of NDVI is approximately zero and (3) the NDVI value before this point is around 0.3 (i.e. NDVI value for bare soil). For judging the harvest date three conditions have been used: (1) the date is later than the silking date / setting pods; (2) the second derivative is almost zero and (3) the NDVI is close to the bare soil, where the dough stage can be recognized as the joint point of saturation and senescence.

In the study of Chen *et al.* [2], their aim was to monitor rice cropping systems and classify rice crops in the upper Mekong Delta, Vietnam, from 2001 to 2007 using time-series MODIS data. The empirical mode decomposition (EMD) was used to filter out noise from the time-series NDVI data prior to rice crop classification. They employed the linear mixture model (LMM) applied to quantify the abundance faction of land-use classes in a pixel then they used these classes with the filtered NDVI data to map the rice cropping systems.

Apart from the filter approachs, there are a number of methods to derive crop phenology that can also be used. The study by Chandola *et al.*[3] used the data mining technique called the time-series segmentation methods to derive phenology indices from NDVI data, and compare them with the phenology indices derived from the AmeriFlux data. The time-series segmentation methods employed in [3] were the piecewise linear regression and bottom-up segmentation. The time-series segmentation based methods break the given time series into various segments and used each segment match with phenological characteristics. The derived phenological characteristics were compared with data from the AmeriFlux to identify the performance.

The purpose of this study is to propose new approach to estimate the rice cultivation and harvest date from the MODIS NDVI time-series data. From our study, we discovered that the NDVI time-series data of rain-fed rice have almost the same characteristic every year. If there are not any abnormally events happened such as flooded and drought, the NDVI time-series data of rain-fed rice will have the similar characteristic. Due to this peculiarity, we define the ideal NDVI time-series data of rain-fed rice in a year and divide it into 4 stages. By following the experiment NDVI profile, we applied Hidden Markov Model (HMM) and Viterbi algorithm to the real NDVI data to define the state of the rice by its own NDVI data.

II. DATA AND PRE-PROCESSING

A. Study Area

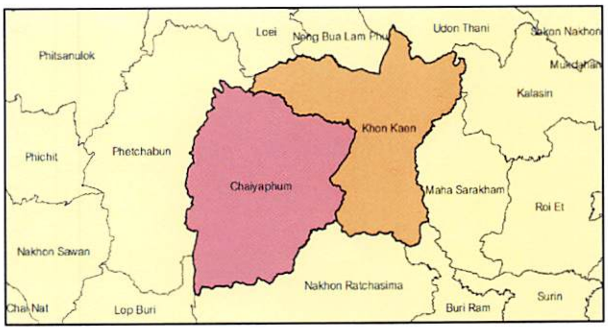


Fig.3 The two study areas, Khonkean and Chaiyaphum provinces, Thailand

In this study, the data collected around Khonkean and Chaiyaphum provinces are used in this experiment. The data include the rice cultivation and harvest information obtained from Thailand's rice department. These two provinces are located in Northeastern region of Thailand which is the rice major production area. Most of the rice cultivated in this area is rain-fed rice which is high quality rice and can be grow only one crop per year. According to rice information from the Thailand's rice department, the rice in this area are planted from May to October which is during the rainy season and takes about six months before ready to harvest.

B. Ground truth data

This ground truth data provided by Thailand's rice department consist of location information (UTM coordinates and province), cultivation date and harvest date of rice. The obtained data provide 40 areas of rice information which consist of 25 areas which have cultivation and harvest dates and another 15 areas which only have harvest dates. This data will be used to access the accuracy of the date estimation by our algorithm.

C. Data and Pre-processing

The 8-day MODIS/Terra composite data (MOD09Q1) acquired from NASA from 2000 to 2012 are used as the experiment data. This image data have 250 meter resolution, and can be obtained every 8 day in the red and near infrared (NIR) spectral bands (0.6μm -0.9μm). This data can also download freely via the Earth Observing Data Gateway. The data were re-projected to the Geographic coordinate over Thailand. The NDVI time-series data is derived from red and NIR bands of MOD09Q1 datasets as (1).

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad (1)$$

Thailand is located in the south eastern region of the continent of Asia. The latitude of Thailand is around 15°, i.e., Thailand is in the topical region. It is commonly known that there are a lot of cloud contaminations in satellite images in the topical area. The raw NDVI time-series data of Thailand are very noisy and hard to apply in study. In order to reduce noise problem in raw NDVI data, we smooth the time-series data by using the Savitzky-Golay filter[5] which can be implemented easily. In addition, before smoothing the NDVI data by Savitzky-Golay filter, the extremely noisy samples with a random NDVI increase or decrease greater than 0.4 were rejected then replaced by a linearly interpolated value using the adjacent points. The example of smoothed NDVI time-series data are shown in Fig.4.

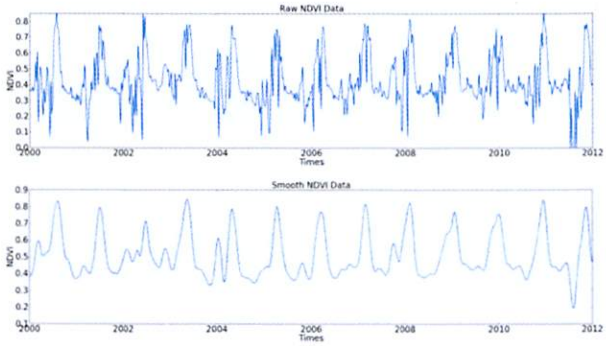


Fig.4 The raw NDVI time series data and the NDVI time-series date after pre-processing and smooth by Savitzky-Golay filter

III. METHODOLOGY

A. Method description

Rice usually has simple characteristic of growth. It starts from the initial state or growing state which NDVI value starts to increase from NDVI value of bare soil (around 0.3-0.4). When rice grows to mature state, the NDVI value will be saturated. After that, the rice starts to wither and in this state farmer will harvest their rice. As a result, the NDVI value in this state dramatically decreases back to NDVI value of bare soil. Because of these characteristics, we declare the ideal NDVI profile of rain-fed rice in a year as shown in Fig.4 which includes the nothing state where NDVI value is the value of bare soils (0.3), the growing state where NDVI value is increasing, the mature state where NDVI is saturation, and the harvest state where NDVI value is decreasing.

For the rain-fed rice (cultivated only one time per year), the simplified NDVI profile within a year (Fig.5) is employed, and the hidden Markov model (HMM) is used to characterize the transition from one state to another.

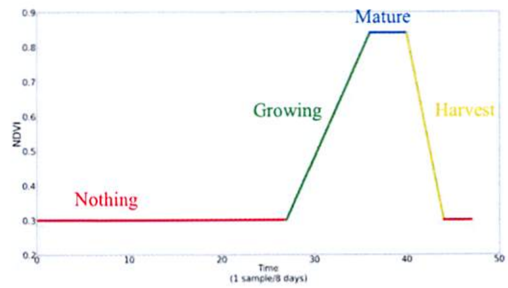


Fig.5 Ideal NDVI profile of rain-fed rice in a year and its states

Let x_k be the state of rain-fed rice at the k-th sample (Fig. 5.) It is clear that the possible states of x_k are nothing, growing, mature and harvest which are denoted as $n, g, m,$ and h, respectively. Furthermore let $P_{i \to j}$ be the transition probability from the state i to the state j. In this paper, we assume that the rice states can only change from nothing to growing, growing to mature, mature to harvest, and harvest to nothing. Fig.6 depicts the state transition considered in this paper.

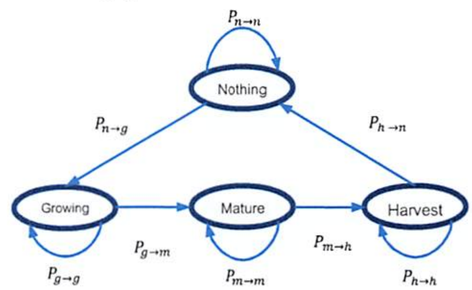


Fig.6 Hidden Markov Model of 4 states in NDVI data

Next, we estimate the transition probabilities from the actual observed NDVI data and found that, on average, the length of nothing, growing, mature, and harvest states are at around 30, 9, 3, and 4 samples, respectively. Since the average number of samples that rice remains in a given state i is equal to $1 + \frac{P_{i \to i}}{1-P_{i \to i}}$, the transition probabilities are equal to 29/30, 1/30, 8/9, 1/9, 2/3, 1/3, 3/4, and 1/4 for $P_{n\to n}$,$P_{n\to g}$, $P_{g\to g}$, $P_{g\to m}$, $P_{m\to m}$,$P_{m\to h}$, $P_{h\to h}$ and $P_{h\to n}$, respectively.

Furthermore, let y_k be the observed NDVI value derived from the 8-day composite NDVI MODIS data. When the current state is nothing, the conditional probability for the observed NDVI value given all the previous rice growth states is assumed to have Gaussian distribution with the mean of μ_n and the variance of σ_n^2. If the current state is growing, the observed NDVI is assumed to Gaussianly distributed with the mean value of $\mu_n + g_k\alpha$ and the value of σ_g^2 where g_k is the number of consecutive samples including the current one that rice is in growing state, and α is the increase rate of NDVI value in growing state. Similar to the nothing state, rice in the mature state is assumed have the NDVI value that is Gaussianly distributed with the mean of μ_m and the variance of σ_m^2. Since the NDVI value in the harvest state drastically decreases, we assume that the NDVI values in the harvest state also have a Gaussian distribution with the mean value of $\mu_m - h_k\beta$ and the variance of σ_h^2 where h_k is the number of consecutive samples including the current one that rice remain in the harvest state, and β is the decreasing rate of the NDVI value in the harvest. Table I summarizes the conditional probabilities of the observed NDVI given all the previous rice states. In this work, the unknown parameters ($\alpha, \beta, \mu_n, \mu_m, \sigma_n^2, \sigma_g^2, \sigma_m^2, \sigma_h^2$) are estimated directly from each yearly data (46 samples) since data from different year may have different amount of variations and offset. These parameters are the simple parameters and can be estimated easily. The parameter μ_n is the mean of NDVI value that less than 0.4. μ_m is the mean of 3 highest values. α and β are the increasing rate and decreasing rate of the data. σ_n^2 is the variance of the data. Due to the more fluctuated of the data in state growing and less in state mature and harvest, the parameter σ_g^2 is set to 2 times higher than σ_n^2 and σ_m^2, σ_h^2 is set to half of σ_n^2.

TABLE I
THE CONDITIONAL PROBABILITY OF THE OBSERVED NDVI VALUE GIVEN ALL THE PREVIOUS STATE

The k-th state	$Pr(y_k \mid x_1, \dots, x_k)$
Nothing	$N(\mu_n, \sigma_n^2)$
Growing	$N(\mu_n + g_k\alpha, \sigma_g^2)$
Mature	$N(\mu_m, \sigma_m^2)$
Harvest	$N(\mu_m - h_k\beta, \sigma_h^2)$

To determine the rice growth state, we employ the maximum *a posteriori* (MAP) criteria, i.e.,

$$X_{1:N}{}^{opt} = \arg\max_{X_{1:N}} \Pr(X_{1:N} \mid Y_{1:N}) \tag{2}$$

where $X_{1:N} = \{x_1, \dots, x_N\}$ and $Y_{1:N} = \{y_1, \dots, y_N\}$. By using the Bayes' rule, and the fact that $\Pr(Y_{1:N})$ is independent of the choice of $X_{1:N}$, Eq. (2) can be modified to

$$X_{1:N}{}^{opt} = \arg\max_{X_{1:N}} [\Pr(Y_{1:N} \mid X_{1:N}) \Pr(X_{1:N})]. \tag{3}$$

Furthermore, we assume that the observed NDVI from any distinct samples are statistically independent when X_N are given, and depends only on the previous and current state of rice. By applying the chain rule, the MAP criteria becomes

$$X_{1:N}{}^{opt} = \arg\max_{X_{1:N}} \left[P(x_1) \prod_{k=1}^{N} \Pr(y_k \mid X_{1:k}) \Pr(x_k \mid X_{1:k-1}) \right]. \tag{4}$$

Since log(.) is monotonically increasing function, Eq. (4) can be rewritten as

$$X_{1:N}{}^{opt} = \arg\min_{X_{1:N}} \left[\sum_{k}^{N} E_k(y_k, X_{1:k}) \right] \tag{5}$$

where

$$E_k(y_k, X_{1:k}) = -\begin{cases} \log \Pr(x_1) + \log \Pr(y_1 \mid x_1) & k = 1 \\ \log \Pr(x_k \mid X_{1:k-1}) + \log \Pr(y_k \mid X_{1:k}) & k \neq 1. \end{cases} \tag{6}$$

To find the solution of Eq. (5), we create the trellis diagram (Fig.6) that collects all the possible paths of the rice growth in time-series data. Each path has its own weight. The weight of each path is calculated by the sum of the all branches the path passing through. The weight computation depends on the observed NDVI value and the next state. For example, if the NDVI value is about 0.3, the weight of paths that are in state nothing will be low and the weight of paths that are in state mature will be high. If a particular node is reached by two or more path the Viterbi algorithm will keep the path with lowest cost and ignore the rest. Then, weight of all final possible paths are compared to find the path that has the lowest weight and use the state that this path go though as the state of NDVI time-series data. In fact, we assign the weight of each path equal to $E_k(y_k, X_{1:k})$.

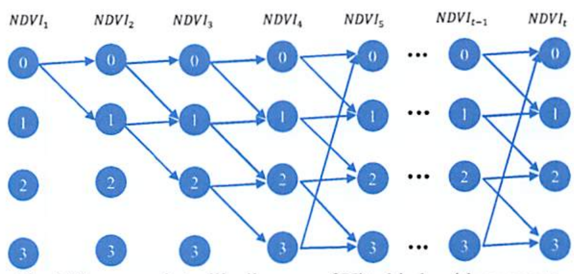
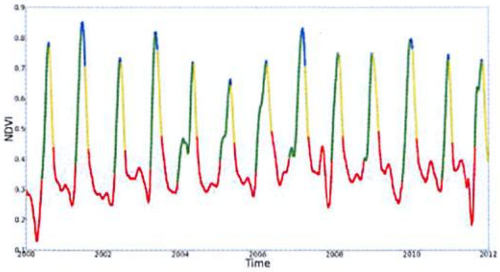


Fig.6 The example trellis diagram of Viterbi algorithm uses to define state of each NDVI value ($NDVI_1, NDVI_2, \dots, NDVI_t$). 0 = nothing state, 1 = growing state, 2 = mature state, 3 = harvest state

IV. Experimental Results

The defined states are plot into NDVI time-series and divided into 4 colours, red stands for nothing state, green stands for growing state, blue stands for mature state, and yellow stands for harvest state. Some example result applied algorithm with NDVI date of rain-fed rice from Khonkean and Chaiyaphum province are shown below.

a) 

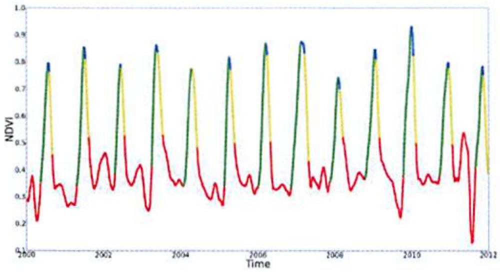
b) 

Fig.7 The state defining result from Chaiyaphum province (a) and Khonkean province (b).The red, green, blue, and yellow lines represent nothing state, growing state, mature state, and harvest state respectively.

The rice cultivation date and harvest dates are estimated by these defined states. The cultivation dates are indicated by the sample that is the end of nothing state and the next sample of this sample changing to state growing. The time of this sample is indicated as cultivation date. The time of the sample that is the end of harvest state and the next sample of this sample changing to state nothing is indicated as the harvest date.

For accessing accuracy, we use 25 areas of cultivation dates and 40 areas of harvest dates comparing with estimate dates from the algorithm. We have two ways to find error. First, we find error in term of differences in days between the ground data and the estimated dates from the NDVI time-series. We can also measure the accuracy in terms of samples. The averaged sample error is at 1.818 samples in estimating the cultivation dates and 0.75 samples for the estimate of harvest dates. The averaged errors in days are 15.48 and 6.525 days for cultivation and harvest dates, respectively. The errors are reasonable given the fact that NDVI observations are 8 day composites.

The harvest date estimation has much more accuracy than the cultivation date due to the fact that the growing season in rain-fed rice usually start in rainy season which have lots of cloud contaminations. This makes the NDVI data in this period are very noisy. These noises might affect the state defining process. On the other hand, there are very few noises in harvest period and the Savitzky-Golay filter can handle most of them.

V. Conclusion

In this study, we proposed a new approach to identify the rice cultivation and harvest dates by dividing the rice's NDVI time-series data into 4 states (nothing, growing, mature, harvest). We use the hidden Markov model to define state transition and its probability. After that, we use Viterbi algorithm for applying the state into the NDVI time-series data. Then, we find the rice cultivation and harvest date by using these defined state. The cultivation and harvest date estimation result have acceptable accuracy. The average error of harvest date estimation is less than 8 days or less than 1 sample which is quite impressive. In cultivation date estimation, the accuracy drops because of noise effect from cloud in growing season but the result still has average error below 16 days which is still satisfactory accuracy. Apart from the rice cultivation and harvest date estimation, the result in this study also has the state of the rice in the NDVI time-series data which can be useful in future study.

Acknowledgment

This research is financially supported by Thailand Advanced Institute of Science and Technology (TAIST), Kasetsart University (KU).

References

[1] Chumkesornkulkit, K.,Kasetkasem, T.,Rakwatin, P.; Eiumnoh, A.; Kumazawa, I.; Buddhaboon, C., "Estimated rice cultivation date using an extended Kalman filter on MODIS NDVI time-series data," Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2013 10th International Conference on , pp.1,6, 15-17 May 2013

[2] C.F. Chen, N.T. Son, L.Y. Chang, "Monitoring of rice cropping intensity in the upper Mekong Delta", Vietnam using time-series MODIS data, Advances in Space Research, Volume 49, Issue 2, 15, Pages 292-301, January 2012

[3] Chandola, V.; DafengHui; LianhongGu; Bhaduri, B.; Vatsavai, R.R., "Using Time Series Segmentation for Deriving Vegetation Phenology Indices from MODIS NDVI Data," Data Mining Workshops (ICDMW), 2010 IEEE International Conference on , pp.202,208, 13-13 Dec. 2010

[4] Hu Zhao; Zhengwei Yang; Liping Di; Lin Li; Haihong Zhu, "Crop phenology date estimation based on NDVI derived from the reconstructed MODIS daily surface reflectance data," Geoinformatics, 2009 17th International Conference on , pp.1,6, 12-14 Aug. 2009

[5] Jin Chen, Per. Jönsson, Masayuki Tamura, ZhihuiGu, Bunkei Matsushita, Lars Eklundh, "A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter", Remote Sensing of Environment, Volume 91, Issues 3–4, 30 June 2004, Pages 332-344, ISSN 0034-4257