A new MODIS forest fire detection algorithm for early small fires with regression analysis between surface temperature and NDVI

Yong Huh[1] and Sanggu Kang[1]
[1]Spatial Information Research Institute, Korea Land and Geospatial Informatrix Corp., Rep. of Korea,
Email: huhyong78 @lx.or.kr

KEY WORDS: MODIS contextual algorithm**,** Prediction interval, Surface temperature, NDVI

ABSTRACT: In this paper, we developed a forest fire detection algorithm using a regression function between NDVI and land surface temperature. Previous detection algorithms use the land surface temperature as a main factor to discriminate fire pixels from non-fire pixels. These algorithms assume that the surface temperatures of non-fire pixels are intrinsically analogous and obey Gaussian normal distribution, regardless of land surface types and conditions. And the temperature thresholds for detecting fire pixels are derived from the statistical distribution of non-fire pixels' temperature using heuristic methods. This assumption makes the temperature distribution of non-fire pixels very diverse and sometimes slightly overlapped with that of fire pixel. So, sometimes there occur omission errors in the cases of small fires. To ease such problem somewhat, we separated non-fire pixels into each land cover type by clustering algorithm and calculated the residuals between the temperature of a pixel under examination whether fire pixel or not and estimated temperature of the pixel using the linear regression between surface temperature and NDVI. As a result, this algorithm could modify the temperature threshold considering land types and conditions and showed improved detection accuracy.

1. Introduction

Forest fires make an abrupt increase of land surface temperature comparing to its neighborhood. So, forest fire detection algorithms in the field of remote sensing use temperature thresholds to discriminate fire pixels from non-fire pixels. The contextual algorithm is known as the most flexible method to determine these temperature thresholds because it introduces a background characterization that statistically compares the temperature of a pixel being considered with those of neighbouring background pixels to determine thresholds (Giglio et al., 2003; Manyangadze 2009). This algorithm derives the thresholds from statistical examination of neighboring background pixels centered on the pixel under examination. It assumes that the surface temperatures of the background pixels obey a normal distribution, so a positive statistical anomaly in surface temperature distribution can be a forest fire in forest area. About 3 standard deviations above the mean of the background temperature is generally used for the threshold.

However, this is not always satisfied because there sometimes exist more than two land cover types whose surface temperature distributions are clearly different, so that the standard deviation of overall background pixels could be overestimated. This makes operational difficulties in small fire detection which is very important in early fire fighting.

To address this problem, we used a physical relationship between NDVI (normalized difference vegetation index) and surface temperature. The relation between the amount of vegetation and the surface temperature of a given region has been analysed (Lambin and Ehrlich 1996). Thermal energy of land surface is transferred to the atmosphere in the form of latent heat through bare soil evaporation, plant transpiration and direct evaporation of water intercepted by plant canopies. In general, soil water extraction by plant roots occurs more rapidly and at much greater depth than water diffusion to the soil surface. So there is a negative relation between NDVI and surface temperature. This relation is widely observed and used for various applications in environmental remote sensing areas. Considering previous phenomena, we developed a modified forest fire detection algorithm which uses a regression function between NDVI and surface temperature to detection forest fire. This algorithm was based on contextual algorithm and used statistical examination of residuals between an observed and an estimated surface temperature, instead of surface temperature in general contextual algorithm. The estimated surface temperature was calculated by regression function of non-fire pixels in background pixels. Details of the proposed method are explained after brief introduction of prediction interval of linear regression.

2. Prediction interval of linear regression

A linear regression model can be constructed as in Equation (1) (Montgomery et al. 2001). ε is an error term vector, and X, y, and β are the regressor value matrix, the response value vector and the regression coefficient vector, respectively. *n* and *k* represent the numbers of observations and coefficients in the model, and β is obtained by the least squares method as in Equation (2):

$$y = X\beta + \varepsilon \tag{1}$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{x_{11}} & \mathbf{x_{12}} & \cdots & \mathbf{x_{1k}} \\ \mathbf{1} & \mathbf{x_{21}} & \mathbf{x_{12}} & \cdots & \mathbf{x_{2k}} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{1} & \mathbf{x_{n1}} & \mathbf{x_{n2}} & \cdots & \mathbf{x_{nk}} \end{bmatrix}, \quad y = \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \vdots \\ \mathbf{y_n} \end{bmatrix}, \quad \beta = \begin{bmatrix} \boldsymbol{\beta_0} \\ \boldsymbol{\beta_1} \\ \vdots \\ \boldsymbol{\beta_k} \end{bmatrix}$$

$$\beta = \left(X^{\mathbf{T}} X\right)^{\mathbf{-1}} X^{\mathbf{T}} y \tag{2}$$

An important application of this regression model is to specify a prediction interval that represents the uncertainty of predicting the value of a single future observation from a population based on the distribution of past observations. Given a specific regressor value vector, X_0 and confidence level $\boldsymbol{\alpha}$, there is a 100(1 - $\boldsymbol{\alpha}$)% probability that the real value of $\mathbf{y_0}$ corresponding to x_0 is within a prediction interval about the predicted value $\hat{\mathbf{y}}_{\mathbf{0}}$, which is obtained by $x_0^{\mathrm{T}}\beta$. This interval is obtained by a t-test of a random variable $\boldsymbol{\psi}$, which represents the difference between the observed value and the estimated value in the regression model, as shown in Equation (3):

$$\boldsymbol{\psi} = \mathbf{y_0} - \hat{\mathbf{y}}_{\mathbf{0}} \tag{3}$$

This variable obeys the normal distribution with zero mean as Equation (4). Variance, $\boldsymbol{Var}(\psi)$ in the equation is derived from Equations (5) and (6):

$$\boldsymbol{E}(\psi) = \mathbf{0},\ \boldsymbol{Var}(\psi) = \boldsymbol{\sigma}^2\left(1 + x_0^{\mathrm{T}}\left(X^{\mathbf{T}}X\right)^{-1}x_0\right) \tag{4}$$

$$\boldsymbol{Var}(\psi) = \boldsymbol{Var}(\mathbf{y_0} - \hat{\mathbf{y}}_{\mathbf{0}}) = \boldsymbol{Var}(\mathbf{y_0}) + \boldsymbol{Var}(\hat{\mathbf{y}}_{\mathbf{0}}) = \boldsymbol{\sigma}^2 + \boldsymbol{Var}\left(x_0^{\mathrm{T}}\beta\right) \tag{5}$$

$$\begin{aligned} \boldsymbol{Var}\left(x_0^{\mathrm{T}}\beta\right) = \boldsymbol{Var}\left(x_0^{\mathrm{T}}\left(X^T X\right)^{-1} X^T \mathbf{y}\right) &= x_0^{\mathrm{T}}\left(X^T X\right)^{-1} X^T \boldsymbol{Var}(\mathbf{y}) X \left(X^T X\right)^{-1} x_0 \\ &= \boldsymbol{Var}(\mathbf{y}) x_0^{\mathrm{T}}\left(X^T X\right)^{-1} X^T X \left(X^T X\right)^{-1} x_0 \\ &= \boldsymbol{\sigma}^2 x_0^{\mathrm{T}}\left(X^T X\right)^{-1} x_0 \end{aligned} \tag{6}$$

Under the normality assumption of the error term of $\boldsymbol{\psi}$, whose mean is zero, and the t-statistic in Equation (7), the confidence interval of (1 - α)100% on the regression prediction of $\mathbf{y_0}$ can be expressed as Equation (8), where *p* is the number of degrees of freedom in the model, and a standard deviation of error $\boldsymbol{\sigma}$ is defined as in Equation (9). Thus, given confidence level $\boldsymbol{\alpha}$, the upper limit of the prediction interval for the temperature threshold in Figure 2 can be obtained as the right-side term in Equation (8):

$$\frac{\mathbf{y_0} - \hat{\mathbf{y}}_{\mathbf{0}}}{\sqrt{\boldsymbol{Var}(\psi)}} \sim \mathbf{t}_{\alpha/2,\mathbf{n-p}} \tag{7}$$

$$\hat{\mathbf{y}}_{\mathbf{0}} - \mathbf{t}_{\boldsymbol{\alpha/2,n-p}}\sqrt{\boldsymbol{\sigma}^2\left(1 + x_0^{\mathrm{T}}\left(X^T X\right)^{\mathbf{-1}} x_0\right)} \le \mathbf{y_0} \le \hat{\mathbf{y}}_{\mathbf{0}} + \mathbf{t}_{\boldsymbol{\alpha/2,n-p}}\sqrt{\boldsymbol{\sigma}^2\left(1 + x_0^{\mathrm{T}}\left(X^T X\right)^{\mathbf{-1}} x_0\right)} \tag{8}$$

$$\boldsymbol{\sigma} = \sqrt{\frac{\mathbf{y^T y} - \beta^{\mathrm{T}} X^T \mathbf{y}}{\mathbf{n-p}}} \tag{9}$$

3. Proposed method

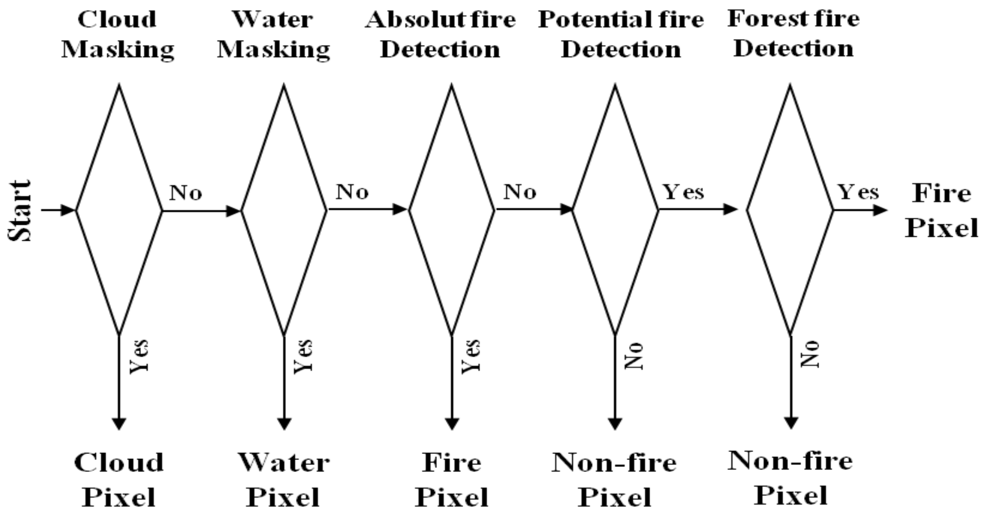

Figure 1. Workflow of proposed algorithm.

The proposed forest fire detection algorithm performed following workflow in Figure 1. Except temperature threshold determination from NDVI and regression analysis, other processes are based on NASA's MODIS contextual fire detection algorithm.

3.1 Cloud masking

The fire detection algorithm begins with the cloud masking stage and the absolute fire detection stage. Cloud pixels must be excluded in statistical examination of background pixels because their NDVI and surface temperatures are clearly different from those of land surface pixels. Cloud detection was performed using a technique based on that used in the MODIS contextual algorithm such as Equation (10).

$$\begin{aligned}&(\rho_{0.65}+\rho_{0.86} > 0.9) \text{ or } (T_{12} < 265\text{K}) \text{ or}\\&(\rho_{0.65}+\rho_{0.86} > 0.7) \text{ and } T_{12} < 285\text{K}\end{aligned} \tag{10}$$

where, $\rho_{0.65}, \rho_{0.86}$: reflectance of 0.65 μm and 0.86 μm band

T_{12} : brightness temperature of 12 μm band

3.2 Water masking

Water masking is done with MODIS NDVI product where water pixels are identified.

3.3 Absolute fire detection

Pixels whose surface temperatures are above 320 **K** in T_4 must be fire pixels because this temperature is rare in natural phenomena unless there is a forest fire:

$$(T_4 > 320\text{K}) \tag{11}$$

where, T_4: brightness temperature of 4 μm band

3.4 Potential fire detection

The potential fire detection stage is a test to identify all pixels plausible to be fires. This is a kind of pre-screening stage that reduces processing time significantly by eliminating obvious fire and non-fire pixels from further processing. In Equation (12), $T_4 - T_{11}$ indicates difference between brightness temperature of 4 μm and 11 μm.

$$(T_4 > 308\text{K}) \text{ and } (T_4 - T_{11} > 8\text{K}) \tag{12}$$

where, T_{11}: brightness temperature of 11 μm band

3.5 Forest fire detection

If more than 2 land cover types whose NDVI and surface temperature distribution are clearly different coexist in background pixels, simple linear regression method cannot explain a complex scatter plot of NDVI and surface temperature distributions. So, we choose 2nd polynomial function. So, a temperature threshold θ_4 is determined by Equation (13)

$$\boldsymbol{\theta_4} = \left(1\ \mathbf{N_0}\,\mathbf{N_0^2}\right)\beta + \mathbf{t}_{\boldsymbol{\alpha}/\mathbf{2},\mathbf{n}-3}\sqrt{\boldsymbol{\sigma}_4{}^{2}\left(1+\left(1\ \mathbf{N_0}\,\mathbf{N_0^2}\right)\left(N^T N\right)^{-1}\left(1\ \mathbf{N_0}\,\mathbf{N_0^2}\right)^{\mathbf{T}}\right)} \tag{13}$$

where, $\left(1\ \mathbf{N_0}\,\mathbf{N_0^2}\right)$ is regressor value vector of the second-order polynomial function with the NDVI of the pixel being tested and $\boldsymbol{\sigma}_4$ is standard deviation obtained by Equation (9). n and p are number of valid background pixels and degree of freedom, respectively. And α is a confidence level of prediction interval analysis. α is an important control parameter of the proposed method's capability to find small forest fire. After experiment, 0.1 % was chosen for the parameter.

4. Result and discussion

The data used in this research was a set of MODIS L1B imageries covering the South Korean peninsula as shown in Table 1. As for the reference data for accuracy assessment, data from the KFS (Korea Forest Service) forest fire information system was used. The KFS data include locations of forest fires, the time of breakout and extinction, the size of damaged areas, and the cause of forest fires.

Table 1. MODIS images used as experimental data in this study.

No.	Local acquisition time	File Name	Terra/Aqua
1	04/02/2003 11:50	MYD021KM.A2003092.0425.005.hdf	Aqua
2	04/03/2003 13:25	MYD021KM.A2003093.0510.005.hdf	Aqua
3	04/05/2003 14:10	MYD021KM.A2003095.0455.005.hdf	Aqua
4	04/06/2003 13:00	MYD021KM.A2003096.0400.005.hdf	Aqua
5	02/09/2004 13:20	MYD021KM.A2004040.0420.005.hdf	Aqua
6	02/16/2004 13:25	MYD021KM.A2004047.0425.005.hdf	Aqua
7	02/19/2004 13:00	MYD021KM.A2004051.0400.005.hdf	Aqua
8	04/05/2004 14:10	MYD021KM.A2004096.0510.005.hdf	Aqua
9	04/16/2004 13:50	MYD021KM.A2004107.0450.005.hdf	Aqua
10	04/18/2004 13:40	MYD021KM.A2004109.0440.005.hdf	Aqua

The proposed algorithm detected 8 more pixels as forest fires rather than general contextual algorithm in Figure. 1. In Table2, the proposed method found 8 more forest fires. Of these eight pixels, two pixels are true forest fire pixels and three pixels are hot facility pixels, meanwhile, another three pixels are false detection.

Table 2 Accuracy Test Results (Total number of true forest fires is 28)

	Proposed method	Contextual algorithm
Total Number of Detected Pixels	54	46
Number of incineration site in detected pixels(*)	12	9
Total Number of Detected Pixels except *	42	37
Number of True Fire Pixels	18	16
Number of Omission Error	10	12
Number of Commission Error	24	21
User Accuracy (%)	42.9(18/42)	43.2(16/37)
Producer Accuracy (%)	64.3(18/28)	57.1(16/28)

5. Conclusion

In this paper, we proposed a new contextual algorithm with threshold optimization to resolve some problems existing in the contextual fire detection algorithm. For this, we assumed that an appropriate consideration of disparate temperature distributions in the background pixels can resolve the edge problem, and improve detection capacity was set up and tested. The results showed that the proposed algorithms yielded higher producer accuracies than that of the MODIS fire product and general contextual algorithm in spite of similar user accuracies. It validate that the proposed algorithm is an effective approach with improved detection capacity. However, there occurred some additional false detection. To solve this problem, another validity index or complementary criteria for optimum cluster number is necessary. Thus, future research may put its focus on relieving such problem.

Acknowledgement

This research was supported by the Ministry of Land, Infrastructure, and Transport in Korea under the Urban Planning & Architecture (UPA) research support program and was supervised by the Korea Agency for Infrastructure Technology Advancement (KAIA) (13 Urban Planning & Architecture 02).

Reference

Giglio, L., Descloitresa, J., Justicec, C.O. and Kaufman, Y.J. 2003. An enhanced contextual fire detection algorithm for MODIS. Remote Sensing of Environment, 87, pp. 273-282.

Lambin, E.F. and Ehrlich, D. 1996. The surface temperature-vegetation index space for land cover and land-cover change analysis. International Journal of Remote Sensing, 17, pp. 463-487.

Manyangadze, T. 2009. Forest fire detection for near real-time monitoring using geostationary satellites. Msc Thesis, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands.

Montgomery, D.C., Peck, A.E. and Vining, G.C. 2001. Introduction to linear regression analysis 3rd edition, John Wiley & Sons, Inc. New York, USA. pp.101–109.