A New Forest Fire Detection Algorithm using Outlier Detection Method on Regression Analysis between Surface temperature and NDVI

Yong Huh, Young-Gi Byun, Jeong-Hoon Son, Ki-Yun Yu, Yong-Il Kim

Spatial Informatics and Systems Lab, Seoul National University,
Sillim-dong, Kwanak-gu, Seoul, Korea
Email : hy21262@dreamwiz.com

ABSTRACT :

In this paper, we developed a forest fire detection algorithm which uses 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.

KEY WORDS: Forest fire detection, Outlier detection, Regression analysis, Land surface temperature, NDVI

1. INTRODUCTION

An occurrence of the forest fire makes an abrupt increase of the land surface temperature comparing to its neighborhood. So, forest fire detection algorithms in the remote sensing techniques use filtering methods with temperature thresholds to discriminate fire pixels from non-fire pixels in surface temperature images.

In the present time, forest fire detection algorithms of MODIS use a contextual algorithm. This algorithm derives the temperature thresholds from statistical examination of neighboring background kernel pixels centered on the pixel under examination whether fire pixel or not. Contextual algorithms assume that the surface temperatures of background pixels obey a normal distribution, so a positive statistical anomaly in surface temperature distribution means forest fire occurrence. In many literatures, about 3σ (standard deviation) above the mean of these background temperatures is used for the threshold.

However, this assumption is not always satisfied in actual implementation of the contextual algorithm. Because there sometimes exist more than two land cover types whose surface temperatures are clearly diverse in background pixels, then the standard deviation value becomes increased resulting in overestimated threshold. This makes operational difficulties in small fire detection which is very important in early fire fighting.

To ease these problem, we used a physical relationship between NDVI(normalized difference vegetation index) and surface temperature. 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 VI 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.

2. METHODOLOGIES

An important application of the regression model is prediction of a new observation to a specified confidence level of the regressor variables. In this paper, a new observation means NDVI of a potential fire pixel. And the predicted confidence interval of surface temperature of the potential fire pixel derived from regression analysis between surface temperature and NDVI of background pixels means is used for the temperature threshold for fire detection.

2.1 Confidence interval estimation of the mean response

We may construct a general linear regression model as equation (1) and regression coefficients vector β is calculated by the method of least squares.

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

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$$

Then, estimated response variable $\hat{y}$ is calculated by equation (2).

$$\hat{y} = x'\hat{\beta} = x'(X'X)^{-1}X'y \tag{2}$$

In other words, this is an unbiased estimator of $E(y \mid x_0)$, since $E(\hat{y}) = x_0'\beta = E(y \mid x_0)$ and the variance of $\hat{y}_0$ is

$$Var(\hat{y}_0) = \hat{\sigma}^2 x_0'(X'X)^{-1}x_0 \tag{3}$$

Therefore, a $100(1-\alpha)$ percent confidence interval on the mean response at the point x_0 with n observations and p degree of freedom is

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 x_0'(X'X)^{-1}x_0} \le E(y \mid x_0) \le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 x_0'(X'X)^{-1}x_0} \tag{4}$$

2.2 Prediction of new observation

If there occurs a forest fire in potential fire pixel, then NDVI and surface temperature of the potential fire pixel disobeys the regression relation of NDVIs and surface temperature s of background non-fire pixels. So, calculation of regression coefficients β is implemented with only background pixels except potential fire pixel. As a result, a prediction of confidence temperature interval of potential fire with the regressor variable, NDVI is statistical process of new observation.

Obtaining an interval estimate of this new observation y_0, the confidence interval on the mean response at $x = x_0$ as equation(4) is inappropriate for this problem because it is an interval estimate on the mean of y, not a probability statement about future observations from that distribution.

So, developing a prediction interval for the future observation y_0 with the random variable $\psi = y_0 - \hat{y}_0$ is necessary. ψ obeys normal distribution with variance as equation (5), because the future observation y_0 is independent of $\hat{y}_0$.

$$Var(\psi) = Var(y_0 - \hat{y}_0) = \hat{\sigma}^2 + Var(y_0) \tag{5}$$

If we use $\hat{y}_0$ to predict y_0, then the standard error of $\psi = y_0 - \hat{y}_0$ is the appropriate statistic on which to base a prediction interval. Thus, the $100(1-\alpha)$percent prediction interval on a future observation at x_0 is

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1 + x_0'(X'X)^{-1}x_0)} \le E(y \mid x_0) \le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1 + x_0'(X'X)^{-1}x_0)} \tag{6}$$

2.3 Application of confidence interval estimation for fire detection

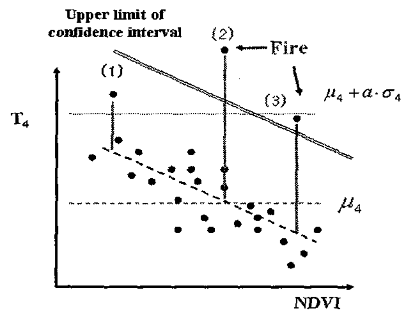

Figure 1. Conceptual representation of threshold determination of the proposed and a general contextual fire detection algorithm.

Figure 1 shows the conceptual representation of threshold determination of the proposed and a general fire detection algorithm. A contextual fire detection algorithm assumes normal and random distribution of temperatures in a background area. However, the proposed algorithm assumes that the land surface conditions such as land surface types or amount of vegetation affect surface temperature. So, the assumption of normal distribution and statistical determination of temperature threshold should be applied to similar surface condition. This implies that there is a distribution of temperature at each land surface condition and that the variance of this distribution is normal distribution. The difference of these approaches is described in figure 1.

In a contextual algorithm, point (1) and (2) will be identified as fire pixel because their temperatures are relatively high comparing to distribution of all background pixels though point (1) is less high comparing to the pixels whose NDVIs are similar to the pixel under examination. And point (3) is not identified as fire pixel in contextual algorithm, however its temperature is very high comparing to pixels whose NDVIs are similar.

This difference is negligible if an algorithm is designed to detect somewhat large forest fires to prevent problem of false alarm. However, if we want to detect small fires which are very important for early fire fighting, the proposed algorithm can improve detection capacity.

3. IMPLEMENTATION

3.1 Workflow of proposed algorithm

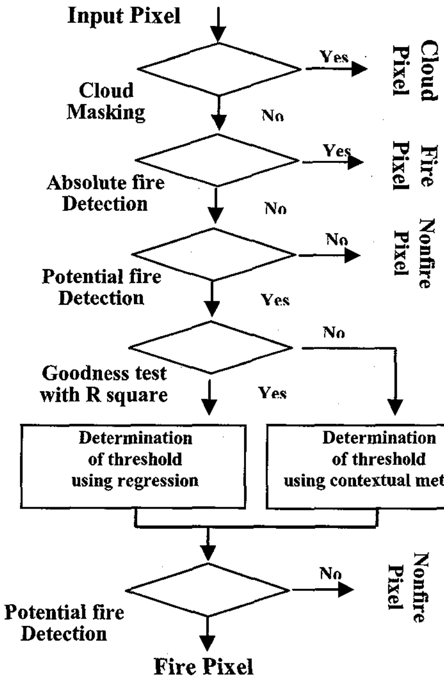

Figure 2. Workflow of proposed algorithm.

The proposed forest fire detection algorithm performed following workflow in figure 2. Except temperature threshold determination from NDVI and regression analysis, other processes are based on NASA's MODIS contextual fire detection algorithm.

3.2 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 (7).

$$(\rho_{0.65} + \rho_{0.86} > 0.9) \ or \ (T_{12} < 265K) \ or \\ (\rho_{0.65} + \rho_{0.86} > 0.7) \ and \ T_{12} < 285K \tag{7}$$

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

T_{12} : brightness temperature of 12 μm

3.3 Potential fire detection

The potential fire detection stage is a test to identify all pixels plausible to be fires using equation (8). 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 (8), $T_4 - T_{11}$ indicates difference between brightness temperature of 4 μm and 11 μm.

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

3.4 Threshold for fire detection

3.4.1 Background characterization

Background characterization is an attempt is made to use the neighbouring pixels to estimate the characteristics of non fire pixels in the absence of fire pixels. Valid neighbouring pixels in a window centered on the potential fire pixel are identified using equation (9).

$$(T_4 < 315K) \ and \ (NDVI > 0.08) \tag{9}$$

Temperature threshold of 315 K is used to discriminate fire pixels and NDVI is used to discriminate water pixels. In contrast to normal land surface, water pixels have very low temperature and NDVI. So if water pixels are contained in background characterization process, inappropriate linear relation will be derived.

3.4.2 Threshold determination with regression method

The regression relation is only fulfilled in ideal conditions. If more than 2 land cover types whose NDVI and surface temperature are clearly different coexist in background pixels, simple linear regression method can not explain a complex scatter plot of NDVI and surface temperature distributions. So, we used quadratic function (2nd polynomial function) after several examinations. In spit of a capability of explanation for complex distributions, more than 2nd polynomial function sometimes made severe error especially in a region of far from a central area.

So, a temperature threshold θ_{T4} with a 95% confidence level is determined by equation (10) where, n is number of valid background pixels and p is degree of freedom.

$$\theta_{T4} = \hat{y}_0 + t_{0.05,n-p}\sqrt{\hat{\sigma}^2(1 + x_0'(X'X)^{-1}x_0)} \tag{10}$$

In equation (10), $\hat{\sigma}^2$ is an unbiased estimator of σ^2 and is derived by following equation (11).

$$\hat{\sigma}^2 = \frac{y'y - \hat{\beta}'X'y}{n-p} \tag{11}$$

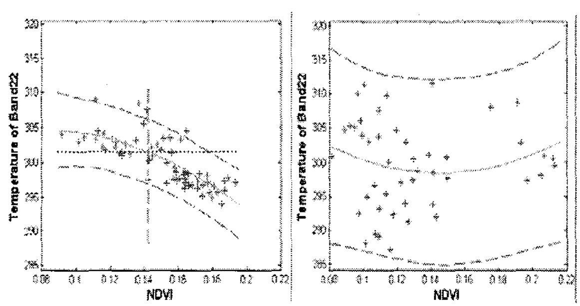

Figure 3. NDVI and T_4 distribution of an ideal forest area and a complex land types area.

Figure 3 shows T_4 threshold estimated by this approach. Red dotted line indicates 95% confidence interval of temperature distribution and the upper limit value is used for temperature threshold.

3.4.3 Threshold determination with contextual method

Results of regression are not always adaptable because any of each regression models can't explain diverse kinds of data distribution. Particularly when 2 types of land surface types whose surface temperatures are very different each other such as figure 3, residuals of regression model becomes large and the model can't explain nor estimate surface temperature distribution.

The proposed algorithm is valid only when the regression model is well fit to NDVI and surface temperature distribution. So, our algorithm can't appropriately detect fire pixel. In this case, we used contextual algorithm instead of the proposed algorithm. If R-square value of regression model is under 0.4, threshold of T_4 is derived by a contextual algorithm such as equation (12).

$$\theta_{T4} = \overline{T}_4 + 3.5\,\sigma_4 \qquad (12)$$

where, $\overline{T}_4$: mean of T_4 in background pixels
σ_4 : std of T_4 in background pixels

3.4.4 Threshold determination of $T_4 - T_{11}$ with contextual method

General contextual algorithms use not only T_4 threshold, but also $T_4 - T_{11}$ threshold for better detection accuracy. In most cases, about 3 times of standard deviation of $T_4 - T_{11}$ in background pixels was used.

$$\theta_{T4\text{-}11} = \overline{T}_{4\text{-}11} + 3.0\,\sigma_{4\text{-}11} \qquad (13)$$

where, $\overline{T}_{4\text{-}11}$: mean of $T_4 - T_{11}$ in background pixels
$\sigma_{4\text{-}11}$: std of $T_4 - T_{11}$ in background pixels

In this research, this additional fire detection criteria is used for both proposed and contextual algorithm.

4. RESULT AND DISCUSSION

The data used in this research was a set of MODIS L1B imageries covering the South Korean peninsula. We used 10 imageries taken on April of 2003, February, March and April of 2004, and April of 2005.

The user accuracy and producer accuracy are calculated by comparing their detection results with the ground data provided by the Korea Forest Service (KFS). To show the feasibility of the proposed algorithm, we compared to the result of a contextual algorithm.

As shown in Table 1, there is little discrepancy in detection capacities and stability of proposed and contextual algorithms. A higher temperature threshold made better result in user accuracy in spite of worse result in producer accuracy and a lower temperature threshold made inverse results. So, it is hard to say that the proposed algorithm is better than contextual algorithm.

Table 1. Accuracy test results
(Total number of true forest fires is 29)

	User Accuracy	**Producer Accuracy**
Contextual Algorithm	39.00% (16/41)	55.17% (16/29)
$t_{0.005}$	35.85% (19/53)	65.51% (19/29)
$t_{0.0025}$	41.46% (17/41)	58.62% (17/29)
$t_{0.0001}$	45.71% (16/35)	55.17% (16/29)
$t_{0.00005}$	51.61% (16/31)	55.17% (16/29)
$t_{0.000025}$	52.00% (13/25)	44.83% (13/29)

The optimal detection result was shown when $t_{0.00005}$ was used for the T_4 threshold determination. In this case, the producer accuracy was same as that of contextual algorithm. However, user accuracy was 12% better than that of contextual algorithm.

Thus, we can conclude that the proposed algorithm reduced false alarm. Most of these reduces false alarm pixels were located on urban or bare soil area. As a result, the proposed algorithm had an adjusting effect on temperature threshold to high value in urban and bare soil area.

Acknowledgments

This research is supported in part by the Korea Aerospace Research Institute (KARI) and Institute of Engineering Science at Seoul National University.

References

Boles, Stephen H. and Verbyla, David L., 2000. Comparison of three AVHRR-based fire detection algorithms for Interior Alaska, *Remote Sensing of Environment*, 72, pp.1-16.

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.

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

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