Proceedings of the 5th WSEAS International Conference on Signal Processing, Istanbul, Turkey, May 27-29, 2006 (pp59-64)

Threshold Optimization of Contextual Fire Detection Algorithm using Fuzzy Clustering

YONG HUH
Urban engineering
Seoul National Univ.
Sillim, Kawnak, Seoul
REP. of KOREA

YOUNG GI BYUN
Urban engineering
Seoul National Univ.
Sillim, Kawnak, Seoul
REP. of KOREA

KI YUN YU
Urban engineering
Seoul National Univ.
Sillim, Kawnak, Seoul
REP. of KOREA

YONG IL KIM
Urban engineering
Seoul National Univ.
Sillim, Kawnak, Seoul
REP. of KOREA

Abstract: - A contextual algorithm that is widely used for identifying forest fire pixels uses a threshold derived from statistical examinations of temperature distributions of background pixels. In general, about 3σ (standard deviation) above the mean of these background temperatures is used for the threshold. In case where land cover types are multifaceted in the background pixels, whose distributions of surface temperatures are clearly diverse, the σ value becomes overestimated resulting in increased threshold. This is a typical edge problem encountered in the current contextual algorithm which explains why relatively small fires are often omitted. Therefore, in this paper, a new algorithm to optimize the threshold is proposed to overcome the above problem. In this algorithm, statistical inferences of various land covers in the background pixels as well as its center pixel are used. For this, a fuzzy clustering is applied to the background pixels and corresponding statistical distribution of temperatures of each cluster is examined to derive a series of thresholds of the clusters. Then the characteristics of the center pixel are analyzed based on memberships of different clusters. Lastly, an optimum threshold is calculated throughout some arithmetic operations to the derived thresholds and memberships information. In this study, the proposed algorithm was applied to the MODIS imageries that were attained during several fire seasons. The results were compared with those from the current algorithm developed by NASA MODIS science team. The proposed algorithm showed relatively high accuracy by detecting several pixels that the current algorithm failed to sense. For a ground truth, forest fire data provided by the Korea Forest Service were used.

Key-Words: Forest fire detection, Contextual algorithms, Threshold optimization, Fuzzy clustering, MODIS

1 Introduction

A contextual algorithm identifies a fire pixel based on the level of contrast between the potential fire pixel and its background pixels [1]. As opposed to the traditional fire detection algorithms, in which a pre-fixed threshold for a given region and season is used, the contextual algorithm determines a threshold from statistical examinations of temperature distributions of background pixels at the time the imagery is analyzed [2]. In general, about 3σ (standard deviation) above the mean of these background pixels' temperatures is used for the threshold. In that respect, the contextual algorithm is believed to be more flexible and effective in changing seasons and different environmental conditions [3], and thus, used for global forest fire monitoring programs such as IGBP (International Geosphere Biosphere Program) or NASA's EOS(Earth Observing System) program.

Even though the contextual algorithm is useful, there are still some room for improvements. Surface feature that produces a sharp temperature transition often results in either omission or commission errors [4]. This means that when the common assumption that surface temperatures are in Gaussian normal distribution is rejected due to the feature, the errors are introduced. In practice, existence of diverse land cover types of which surface temperatures are clearly diverse accordingly and change abruptly at the borders is one main reason.

For example, if there are two land cover types such as a dense forest and a dry bare soil in background pixels whose surface temperatures are different, then a standard deviation is overestimated and the threshold is increased. This is a typical case of an edge (or boundary) problem in the contextual algorithm that explains why relatively small fires are, in many cases, difficult to detect. This problem is referred to as omission error.

In this research, a new algorithm that optimizes the threshold is proposed to resolve this problem. In this proposed algorithm, statistical inferences of various land covers types in the background pixels as well as its center pixel are used. A widely used fuzzy clustering process is applied to the background pixels and corresponding statistical distribution of temperature of each cluster is examined to derive a series of thresholds of the clusters. Then, the center pixel characteristic is analyzed based on memberships of different clusters. This is a necessary step due to the coarse spatial resolution of the imagery used, which in many case, contains multiple land covers in a pixel. Finally, an optimum threshold is calculated throughout some arithmetic operations to the derived thresholds and memberships information.

To evaluate its effectiveness, the proposed algorithm was tested using the MODIS imageries, in which the imageries were attained during several fire seasons. Then, the results were compared with those from applying the current algorithm developed by NASA MODIS science team.

2 Major Methodological Ideas

To separate and cluster different land cover types, the Gustafson-Kessel (GK) algorithm is used. This fuzzy clustering method is widely used, because it is the first introduced method which can detect clusters of various geometrical shapes such as ellipsoidal distributions in a spectral dimension [5]. With this algorithm, the number of clusters should be pre-defined and has a critical effect on clustering results. To find an optimal number as well as to check the quality of clustering results, the validity index test proposed by Mohamed is used. This method is believed to be reliable and produced good results [7]. Then, some arithmetic operations to derive the optimum threshold are applied.

2.1 Gustafson-Kessel algorithm

The GK algorithm is based on FCM (Fuzzy C Means) approach which is a kind of iterative optimization of an objective functional J [5]. Provided $U = [\mu_{ik}] \in [0,1]^{k \times N}$ is a fuzzy partition matrix of the data $X \in R^{n \times N}$, then $V = [v_1, v_2, \Lambda , v_k]$, $v_i \in R^n$ is a k tuple of cluster centers and $m \in [1, \infty)$ is a scalar parameter that determines the fuzziness of the resulting clusters.

$$J(U,V,X) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ik})^m \left\| x_k - v_i \right\|^2 \qquad (1)$$

where, c and N are the number of clusters and pixels, m is the weighting exponent.

For $m > 1$, Bezdek [6] gave the following necessary conditions derived from making the gradient of J zero with respect to U,V respectively for the minimization.

$$v_i^* = \frac{\sum_{k=1}^{N} (\mu_{ki})^m x_k}{\sum_{k=1}^{N} (\mu_{ki})^m} \quad \forall i \qquad (2)$$

$$\mu^*_{ki} = \left(\sum_{j=1}^{c} \left(\frac{\| x_k - v_i \|^2}{\| x_k - v_j \|^2} \right)^{\frac{1}{(m-1)}} \right)^{-1} \quad \forall i \qquad (3)$$

As a result, the FCM algorithm is a simple iteration through equation (2) and equation (3) until a certain terminal condition is reached [6]. Hence the FCM measures distance by Euclidian distance that it can only detect clusters with the same shape and orientation.

Gustafson and Kessel [5] extended the standard FCM algorithm by employing a new distance norm, D_{ikA}^2, instead of $\|x_k - v_i\|^2$ in order to detect clusters of different geometrical shapes in a data set as follows.

$$D_{ikA}^2 = (x_k - v_i)^T A_i (x_k - v_i), \quad 1 \le i \le c, \; 1 \le k \le N \qquad (4)$$

However, the objective function $J(U,V,X)$ can not be directly minimized with respect to A_i, since it is linear in A_i. This means that $J(U,V,X)$ can be made as small as possible by simply making A_i less positive definite. To obtain a feasible solution, A_i must be constrained in some way. Normally, this is done by constraining the determinant of A_i. That is, the matrix A_i is allowed to vary with its determinant fixed corresponds to optimize the cluster shape while its volume remains constant. Gustafson and Kessel let ρ_i be fixed for each cluster to solve this problem as is explained in equation (5).

Here, $\sum_i$ indicates the fuzzy covariance matrix of the i^{th} cluster that D_{ikA}^2 is generalized squared Mahalanobis distance norm between x_k and v_i.

$$\| A_i \| = \rho_i , \quad A_i = \left[\rho_i \det(\textstyle\sum_i) \right]^{1/n} \textstyle\sum_i^{-1}, \quad \rho > 0 \qquad (5)$$

$$\sum_i = \frac{\sum_{k=1}^{N} \mu_{ik}^m (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^{N} \mu_{ik}^m} \quad (6)$$

2.2 Mohamed validity index

Mohamed proposed a new validity index based on the concept of the ratio between separation and compactness in fuzzy clustering result [7]. The index measures compactness of a cluster by the properties of the fuzzy covariance matrix which explains the dispersion of the data for each cluster and take into account the shape and density of clusters. Similarly, as a measure of the isolation of the cluster centers from one another, the properties of the between-cluster fuzzy matrix is used as a validity index as follows.

$$V_{sc}(U,V,X) = \frac{S(c)}{GComp(c)} \quad (7)$$

where, $GComp(c)$ is the global compactness of the fuzzy clustering that is defined as follows.

$$GComp(c) = \sum_{i=1}^{c} trace(\sum_i) \quad (8)$$

As is shown in equation (8), the $GComp(c)$ is the sum of cluster variance. A small value of this term indicates a compact partition exists. The fuzzy separation of fuzzy cluster can be defined as

$$S(c) = trace(S_B) \quad (9)$$

where, S_B is the between-cluster fuzzy matrix given by

$$S_B = \sum_{i=1}^{c}\sum_{k=1}^{N} \mu_{ki}^m (v_i - \bar{v})(v_i - \bar{v})^T \quad (10)$$

The separation function S_B combines the fuzziness in μ_{ki} with the distance from the i^{th} cluster center v_i to the grand mean $\bar{v}$. Here, a large V_{sc} means a good result has been obtained, because a large $S(c)$ means a well-separating between clusters, whereas a small $GComp(c)$ means a well-compacting within each cluster.

2.3 Optimal thresholds

Once the fuzzy clustering and corresponding statistical attributes are analyzed, a threshold value for a potential fire pixel k can be calculated using the following equations.

$$\theta_{T4}(k) = (\bar{T}_{4(c1)}(k) + 3\,\sigma_{T4(c1)}(k))\mu_{1k} + (\bar{T}_{4(c2)}(k) + 3\,\sigma_{T4(c2)}(k))\mu_{2k} + \Lambda \quad (11)$$

$$\theta_{\Delta T}(k) = (\Delta\bar{T}_{(c1)}(k) + 3.5\sigma_{\Delta T(c1)}(k))\mu_{1k} + (\Delta\bar{T}_{(c2)}(k) + 3.5\sigma_{\Delta T(c2)}(k))\mu_{2k} + \Lambda \quad (12)$$

where, $\theta_{T4}(k)$ and $\theta_{\Delta T}(k)$ are optimized threshold T_4 and ΔT for a potential fire pixel k, $\bar{T}_{(ci)}(k)$ and $\Delta\bar{T}_{(ci)}(k)$ are means of T_4 and ΔT in c_i cluster of background pixels , $\sigma_{T4(ci)}(k)$ and $\sigma_{T4(ci)}(k)$ are standard deviations of T_4 and ΔT in c_i cluster of background pixels.

3 Implementation Details

In this section, we discuss the overall implementation details. The entire workflow is explained in Fig. 1, in which the implementation is divided into two parts: the general contextual fire detection process [4] and a new algorithmic process proposed.

The entire process begins with a test to identify all pixels plausible to be fires, i.e. potential fire pixels. This is a kind of pre-screening that reduces processing time significantly by eliminating obvious fire and non-fire pixels from further processing [4]. Pixels that are clearly non-fire include clouds and cold land pixels. In contrast, pixels that indicate fires are those whose surface temperatures are above $320\,K$, which is very rare in natural phenomena.

After the pre-screening, a test of normality is done using brightness temperatures of background pixels for each potential fire pixel. Thus, for each potential fire pixel, a 7 by 7 kernel is applied and pixels within this kernel constitute the background pixels. Generally in multivariate analysis, the test of normality is conducted by χ^2 test as follows [8].

1. Order the squared distance in Mahalanobis distance from the smallest to the largest one $d_{(1)}^2 \le d_{(2)}^2 \le \Lambda \le d_{(n)}^2 \cdot\cdot$
2. Graph the pairs of $(d_i^2, \chi_p^2((i-0.5)/n))$, where $\chi_p^2((i-0.5)/n)$ is the $100(i-0.5)/n$ percentile of the χ^2 distribution with p degree of freedom.

The plot should resemble a straight line whose slope is about 1. A systematic curved pattern suggests a lack of normality. This hypothesis verified by performing by a simple t-test on the estimated slope of the χ^2 plot. If the temperatures of the background pixels are statistically uniform, the general contextual fire detection process, as in Fig.1, is performed. However, if there are multiple temperature distributions after the normality test, the GK clustering is performed.

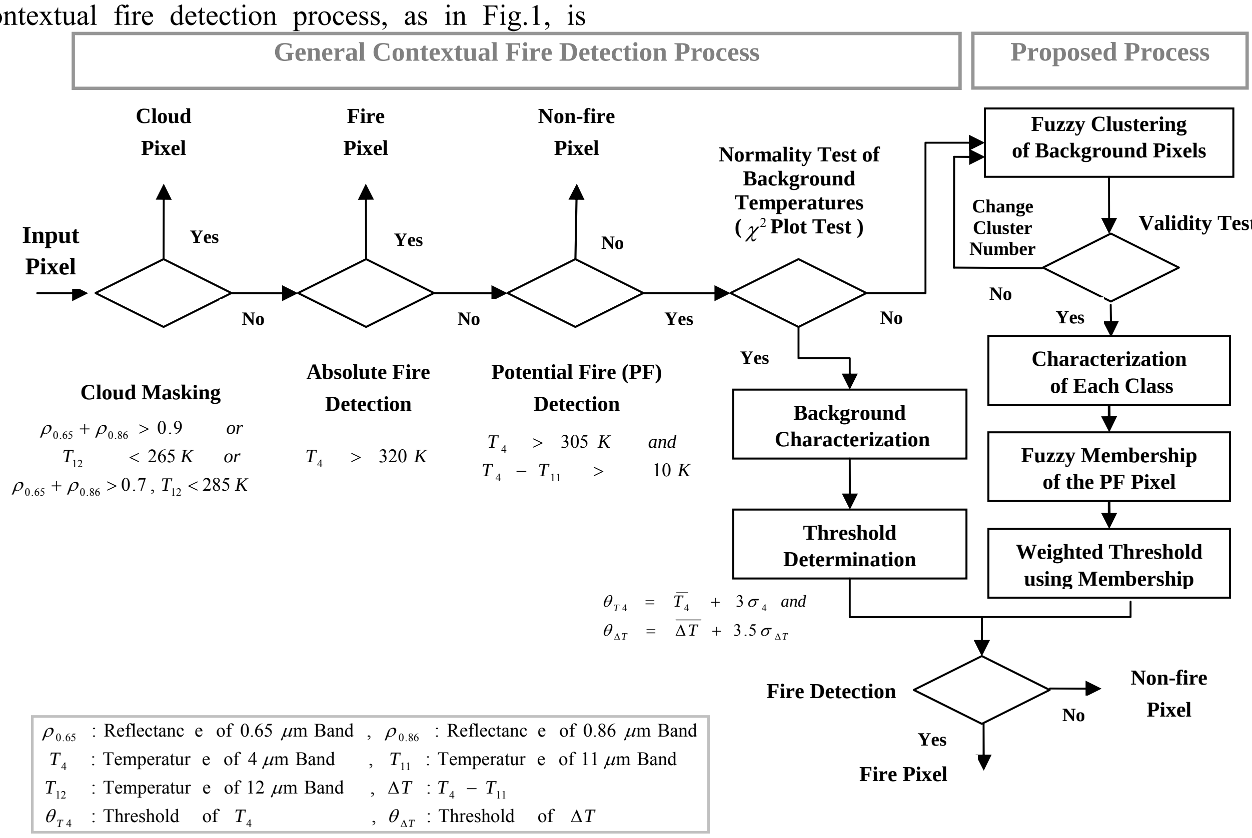

Fig.1 Workflow of this research

The weighting exponent was determined to be 1.5 empirically. The process begins with 2 clusters and the number increases until the validity index reaches the most optimum condition. The mean and standard deviation of each cluster is calculated using membership values of the partition matrix as weight. Then a threshold value for a potential fire pixel k is calculated.

4 Experiments and Results Analysis

4.1 Data Used

The data used in this research was a set of MODIS L1B imageries covering the South Korean peninsula. These images were taken by the NASA's AQUA satellite. We used seven imageries taken on April of 2003, February, March and April of 2004, and April of 2005.

The MODIS bands used in this study are listed in Table 1. The $\rho_{0.65}$, $\rho_{0.86}$ and T_{12} for cloud masking were acquired from band 1,2 and 32 respectively. To detect forest fire from the MODIS imagery, band 21 as well as band 22 was used to resolve the problem related to sensor saturation, which takes place when radiance that reaches the sensor exceeds beyond the upper limit of the sensor's sensing range. The problem of sensor saturation was solved by using band 21 when sensor saturation took place in band 22. Reflectances of Bands 1, 4, and 3 are used for GK algorithm for background pixels clustering.

Table 1 Band numbers and Bandwidth

Band number	Bandwidth (μm)	Band number	Bandwidth (μm)
1	0.620-0.670	21	3.930-3.989
2	0.841-0.876	22	3.930-3.989
3	0.459-0.479	31	10.780-11.280
4	0.545-0.565	32	11.770-12.270

4.2 Accuracy Test

User accuracy and producer accuracy were calculated by comparing detection results with the ground data provided by the Korea Forest Service (KFS). Regarding the fact that the AQUA satellite scans Korea Peninsula at around 13:30pm, the forest fires which were running through this time was selected for accuracy test.

The size of damage area recorded in the KFS data can be tens to hundreds times as large as that of burning area. Theoretically, MODIS can sense the forest fire with 50% detection accuracy, whose burning area is over 0.01ha (100m^2) when burning temperature is about 800~1000K. Therefore, another constraint condition for reference data is that the size of damaged area should be over 0.1ha.

To show improvement of detection capacity caused by threshold optimization with proposed fuzzy clustering process, we compared the result of the proposed algorithm to that of the MODIS fire product of which algorithms is developed by MODIS science team. At the same time, the results after applying the general contextual fire detection process were compared with one another.

Table 2 Accuracy Test Results
(Total number of true forest fires is 25)

	Proposed Algorithm	General Contextual Algorithm in Fig.1	MODIS Fire Product
Total Number of Detected Fire Pixels	30	30	18
Number of True Fire Pixels in Detected Pixels	13	12	8
Number of Omission Error	12	13	17
Number of Commission Error	17	18	10
User Accuracy (%)	43.3	40.0	44.4
Producer Accuracy (%)	52	48	32

As shown in Table 2, there is little discrepancy in user accuracies of these three algorithms. They are around 40% and that of MODIS fire product is slightly higher than the others. However, the producer accuracies are different. The proposed algorithm shows 4 % higher producer accuracy than that of the general contextual fire detection algorithm as in Fig. 1 and 20% higher than that of MODIS fire product.

5 Conclusions

In this paper, we proposed a new threshold optimization approach 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 proposed algorithm was implemented with fuzzy clustering method.

The results showed that proposed algorithms yielded higher producer accuracies than those of the MODIS fire product provided by the NASA MODIS Science Team and showed slightly better detection accuracy than the general contextual fire detection algorithm. It validate that the proposed algorithm is an effective approach with improved detection capacity.

The proposed algorithm starts from current contextual algorithm differentiating by considering background characteristics in which temperature threshold is determined using statistical inference. This means that slightly warm pixel likely to be identified as a fire pixel when temperature variance of background pixels is very small. Thus, future research will need to focus on relieving this problem.

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

References:

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

[2] Flasse, S. P. and Ceccato, P., A contextual algorithm for AVHRR fire detection, *International Journal of Remote Sensing*, Vol. 17, 1996, pp.419-424

[3] Pu, R., Gong, P., Li, Z. and Stephen, Scarborough J., A dynamic algorithm for wildfire mapping with NOAA/AVHRR data, *International Journal of Wildland Fire*, Vol. 13, 2004, pp.275-285.

[4] Giglio, L., Descloitresa, J., Justicec, C. O., and Kaufman, Y. J., An Enhanced Contextual Fire Detection Algorithm for MODIS, *Remote Sensing of Environment*, Vol. 87, 2003, pp. 273-282.

[5] Gustafson, D. and Kessel, W., Fuzzy clustering with a fuzzy covariance matrix, *Proceeding of IEEE*, San Diego, CA, USA, 1979, pp.761–766.

[6] Bezdek, James C. et al., *Analysis of fuzzy information Vol. 3 Applications in engineering and science*, Boca Raton, Florida:CRC Press, Inc., 2000, pp.123-131.

[7] Bouguessa, Mohamed and Wang, Sheng-Rui, A new efficient validity index for fuzzy clustering, *Proceeding of the 3th Int. Conference on machine learning and cybernetics*, Shanghai, 26-29 August 2004, pp.1914-1919.

[8] Johnson, Richard A. and Wichern, Dean W., *Applied multivariate statistical analysis 3rd edition*, Englewood Cliff, NJ:Prentice-Hall Inc., 1992, pp. 152-164.