

# Spatial Data Mining: Clustering of Hot Spots and Pattern Recognition

Seng Chuan TAY\*, Wynne HSU\*\*, Kim Hwa LIM\*

\*Centre for Remote Imaging, Sensing and Processing

\*\*Department of Computer Science

National University of Singapore, SOC-1, Level-2, Lower Kent Ridge Road, Singapore 119260

Tel : (65) - 6874 1548, Fax: (65) – 6775 7717, Email : crstaysc@nus.edu.sg

## Abstract

Spatial data mining is the extraction of implicit knowledge, spatial relations or other patterns not explicitly stored in spatial database. The focus of this paper is placed on the information derivation of spatial data. Geographical coordinates of hot spots in forest fire regions, which are extracted from the satellite images, are studied and used in the detection of likely fire points. Due to the saturation of spectral band, there are false alarms in the derived data set. We use clustering and Hough transformation to determine regular patterns in the derived hotspots and classify them as false alarms on the assumption that fires usually do not spread in regular patterns such as in a straight line. This project demonstrates the application of spatial data mining to reduce false alarm from the set of hot spots derived from NOAA images.

## I. INTRODUCTION

Forest fire is a global concern. It has deforested many vegetated regions and has caused serious adverse social, economic and ecological consequences. In year 1997, it was estimated that between 20 and 70 million people suffered health-related problems and 40 thousand people were hospitalized as a result of forest fires in Indonesia [6]. The economic damages have been estimated in billions of US dollars [7].

Satellite remote sensing provides key information useful in fires monitoring and management as well as damage assessment. The AVHRR sensors on board the NOAA Polar Orbiting Satellites are able to detect the presence of fire hot spots using channel 3 (3.8  $\mu\text{m}$ ) and provide images showing the spatial distribution and temporal evolution of fire hot spots [5]. Due to the saturation of channel 3 at a temperature of 320K, false alarms can occur in the set of hot spots derived from AVHRR sensor.

The objective of this piece of work is to use spatial data mining approach to reduce the false alarms in the derived hotspots. Clustering and Hough transformation are applied to the positional attributes of hotspots. This approach assumes that forest fires usually do not burn in regular patterns such as in a straight line. Therefore, we classify hot spots positioned on a straight line as false alarms.

## II. SPATIAL DATA MINING

Spatial objects may be either point objects or spatially extended objects such as lines, polygons or polyhedrons. There are three basic types of spatial relations – topological, distance and direction relations, which are binary relations between pairs of objects [3]. The spatial

relation involved in this piece of work is the distance relation and it is used in the clustering procedure where computed distance between two points is compared with the intra-cluster distance threshold. Hence, useful information can be mined from the spatial data set and the information is further processed to fulfill the objectives of the work.

### A. Region Growing Method for Clustering

Region growing method often gives very good segmentations that correspond well to the observed edges [2]. The steps used in the application of region growing method in forming the clusters are as follows:

1. First point in the data set is chosen as a seed and it is compared with the remaining points in the data set.
2. Cluster is grown from the seed point by adding in neighbouring points which are close to it (i.e. absolute difference of distance between any existing points in the cluster and the point under consideration is within the specified intra-cluster threshold value).
3. When the growth of one cluster stops, i.e. the absolute distances of all other points to any existing point in the cluster are all greater than the intra-cluster threshold value, another seed point which does not belong to any clusters yet is chosen from the same data set, and steps 1 and 2 are repeated for the formation of other clusters.
4. This whole process is repeated until all points have been placed in clusters.

Once the clusters for hotspots are formed, we proceed to detect false alarm in each cluster based on the assumption that true fire points are usually not positioned in regular patterns such as in a straight line. For this purpose the use of Hough Transform is suitable. Hough Transform is an algorithm used to link edge points by determining whether they lie on a curve of specified shape. The method is particularly useful in cases when the number of solutions is not known. Hough Transform can also be used for straight-line detection, which is what this piece of work needs to do.

### B. Hough Transform

Consider a point  $(x_i, y_i)$ . There are infinite number of lines, represented by  $\mathbf{y}_i = \mathbf{a}\mathbf{x}_i + \mathbf{b}$ , that pass through  $(x_i, y_i)$ . In Hough transform we rearrange the equation as  $\mathbf{b} = -\mathbf{a}\mathbf{x}_i + \mathbf{y}_i$ . In the  $\mathbf{ab}$  plane (also called parameter or transform space), the equation of a single line containing a fixed point  $(x_i, y_i)$  represents all the possible lines passing through the point in  $\mathbf{xy}$  plane. A second point  $(x_j, y_j)$  will also have a line in the parameter space and this line will intersect the line associated with  $(x_i, y_i)$  at  $(a', b')$  if both

lines are not parallel, where  $a'$  is the gradient and  $b'$  the y-intersect of the line containing both  $(x_i, y_i)$  and  $(x_j, y_j)$  in the  $xy$  plane. Fig. 1 shows the transformation of a line in  $xy$  space to a point in parameter space.

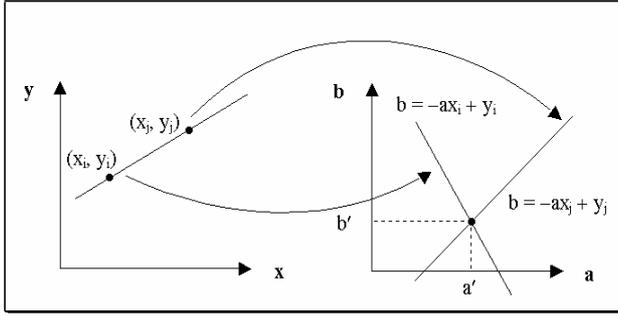


Fig. 1. Transformation of a line in  $xy$  plane to a point in  $ab$  plan

As theoretically the gradient and the y-intersect of a straight line are unbounded, the Hough Transform algorithm cannot be implemented in a program. We therefore bound the values of  $a$  and  $b$ , and divide the space into a finite number of accumulator cells (see Fig. 2). Using this approach, the granularity of a cell will directly affect the closeness of the points located on or near to a straight line.

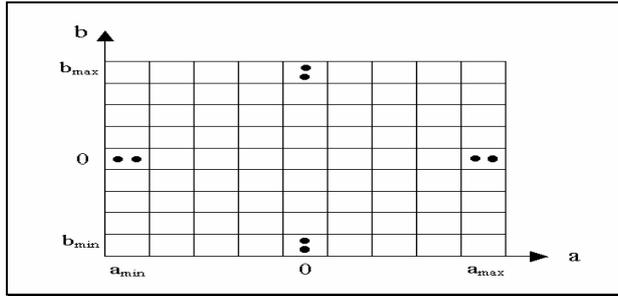


Fig. 2. Accumulator cell with expected ranges of gradient and y-intersect

The following procedure is an exhaustive attempt to place all the hotspots within a cluster to the best suited accumulator cell:

1. Set all counters for accumulator cells to 0.
2. For every point  $(x_k, y_k)$ , let the parameter  $a$  equals each of the allowed discrete subdivision values on the  $a$  axis, and the corresponding  $b$  is solved using the equation  $b = -ax_k + y_k$ .
3. The resulting  $b$  is rounded off to the nearest allowed value.
4. If a choice of  $a_p$  results in solution  $b_q$ , we increment the counter for the accumulator cell for  $a_p$  and  $b_q$  by 1, and include the hotspot in the accumulator cell.
5. The transform space is searched for the largest value of all the counters. All the hotspots in the accumulator cell with the largest counter value are classified as false alarms.
6. Steps 1 to 5 are repeated for the hotspots in all other clusters for other false alarm detection.

### III. IMPLEMENTATION

Window platform is used in the implementation. The user interface is programmed in Visual Basic, and clustering and Hough Transform are programmed in C language. In the first step, the user is required to select the data file to upload for clustering. Each of the data files is named after the NOAA satellite number, and the date and time at which the image is taken. These files contain many pairs of latitude, longitude value and each pair indicates the location of a derived hotspot. In the following interface, the user has to select the correct drive (label 1), paths or folders (label 2) and the file name (label 3). Once a data file is loaded via double-click on label 3, the visual basic program will save the full path of the uploaded file.

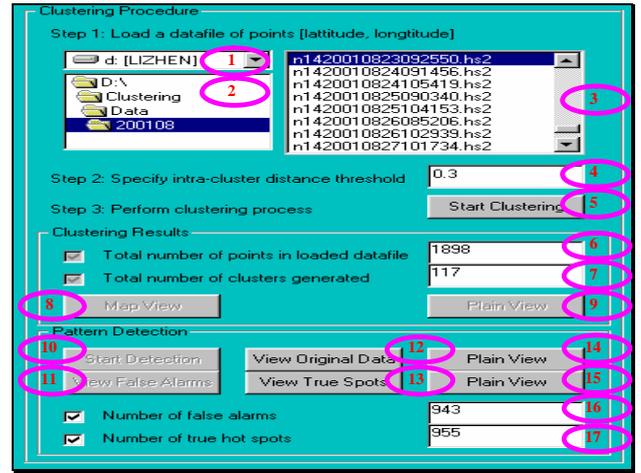


Fig 3. User interface

The next step is to specify a value in the *intra-cluster distance* field (label 4). This threshold value is used in the region growing clustering process. With the click on the *Start Clustering* button (label 5), the C program will be invoked to make use of the parsed-in distance threshold through shell to perform the actual clustering procedure.

#### A. Clustering the Hotspots

The program computes the distance between two points using Pythagoras formula. If the computed distance is less than or equal to the parsed-in intra-cluster distance, the point is deemed as a member of the current cluster and will not be considered again for other clusters. After the clustering procedure is done, the points' information and their cluster labels will be sorted. Finally, all the sorted information is output to a text file for future analysis. The output information, such as the total number of points in the uploaded file and clusters formed, are available upon clicking the checkboxes labeled 6 and 7 in Fig. 3. Two functional buttons for viewing purpose are available, i.e. *Map View* (label 8) and *Plain View* (label 9) for visualization purpose.

#### B. Regular Pattern Detection and Visualizing the Hotspots

This procedure will be performed when the clustering procedure is completed. The detection for hotspots on straight line is performed based on per-cluster basis. The

user can click on the *Start Detection* button (label 10 in Fig. 3) for the program to activate the Hough Transform to detect any straight line of hotspots.

The visualization function buttons (labels 11 to 15 in Fig. 3) provide users with a few option to display all the hotspots, to display only the false alarms, to display only the true hotspots, or to display with or without a background map. A snapshot of the visualization for the false alarms is illustrated in Fig. 4. Useful results or information from the pattern detection, such as numbers of detected false alarms and true hotspots, are displayed in the fields labeled 16 and 17 in Fig. 1.



Fig. 4. Visualization for the location of false alarms

#### IV. VERIFICATION WITH SPOT IMAGES

In order to verify that the set of points which are categorized as straight line are false alarms, we verify the false alarm information derived from the data mining approach with SPOT images at 20-meter resolution. Our objective is to verify that there is no fire or burn scar on the false alarmed hotspots.

30 datasets of NOAA and SPOT data acquired in year 2001 are used in the verification. Though SPOT satellite is unable to zoom into the whole stretch of the recognized straight line due to its limited 60 km swath width, there is no sign of any burn scar from extracts of the SPOT footpaths for the experimental datasets. Fig. 5 and Fig. 6 show an example of the false alarm verification.

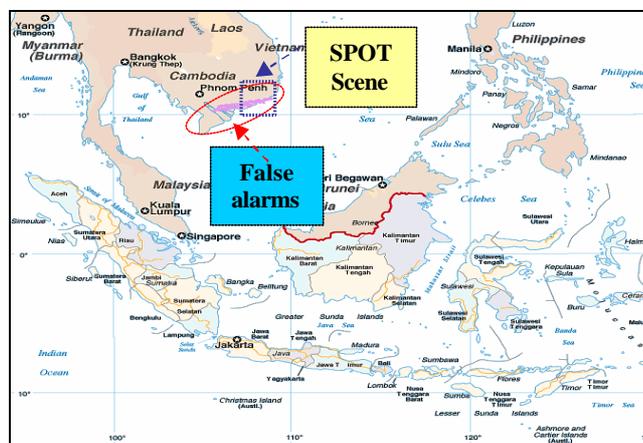
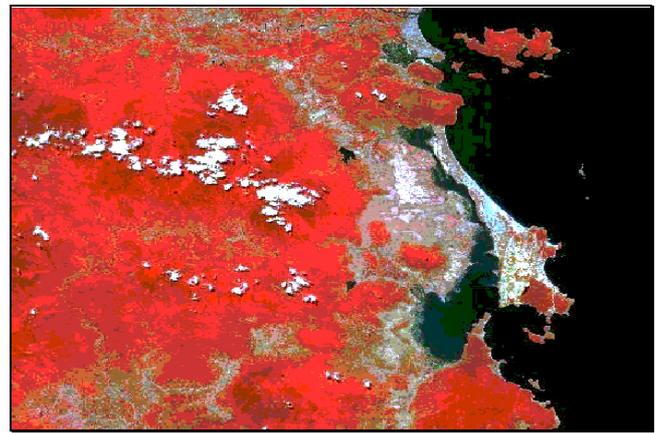


Fig. 5: False alarms derived by Hough's method using NOAA's processed data acquired on 13 July 2002



© CNES 2001

Fig. 5. SPOT data at 20-meter resolution acquired on 17 July 2001 showing no signs of fire or burnt scars on the false alarmed region

#### V. CONCLUSION

The clustering on the hotspot datasets using region growing algorithm and the removal of false alarms using pattern recognition algorithm are demonstrated. An interface is designed to integrate all functions into a useful application. This piece of work can be enhanced to recognize other regular patterns such as rectangle and triangle so as to remove more false alarms in hotspots since fires usually do not spread in such regular patterns.

#### Acknowledgement

This paper is based on a project done by Ms Yap Lee Chen for her partial fulfillment of the requirements for the Degree of Bachelor of Engineering (Computer Engineering), National University of Singapore, in 2002.

#### References

- [1] Centre for Remote Imaging, Sensing and Processing, "Space Views of Asia", CD-ROM, Second Edition, 2001.
- [2] David Marshall, "Region Growing", [http://www.cs.cf.ac.uk/Dave/Vision\\_lecture/node35.html](http://www.cs.cf.ac.uk/Dave/Vision_lecture/node35.html)
- [3] Hans-Peter Kriegel, Jorg Sander, Martin Ester and Stefan Gundlach, "Database Primitives for Spatial Data Mining".
- [4] Jan Kucera and Yoshifumi Yasuoka, "Regional Monitoring of Forest Disturbances and their Potential Effects to Carbon Cycling".
- [5] Liew SC, Lim OK, Kwok LK, Lim H, "A Study of the 1997 Forest fires in South East Asia Using SPOT Quicklook Mosaics", Proc. 1998 International Geoscience and Remote Sensing Symposium, Vol 2, pp. 879-881.
- [6] Robert J Fisher and William J Jackson, "Forest Fires in Rural Communities".
- [7] Shen CM, Liew SC, Kwok LK, "Spatial and Temporal Pattern of Forest or Plantation Fires in Riau, Sumatra from 1998 to 2000", Proc. ACRS 2001 - 22nd Asian Conference on Remote Sensing, 5-9 November 2001, Singapore. Vol. 1, pp. 520-525.