IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 06, Issue 04 (April. 2016), ||V2|| PP 57-60

Comparison of Three Distances In K-Means Clustering On Satellite Imagery

Barnali Goswami[1*] and Sanjay Goswami[2]
[1]*Symbiosis Institute of Computer Studies and Research, Model Colony, Pune, INDIA*
[2]*Narula Institute of Technology, Agarpara, Kolkata, INDIA*

Abstract: One of the preliminary works in the field of flood prediction, due to heavy rainfall, is the detection and identification of convective clouds using satellite imagery. Thermal infra-red (TIR) band images have been extensively used for this purpose. In order to identify the convective cloud, the image has to be clustered so that cloudy pixels can be identified. In this paper k-means clustering has been used for clustering pixels in a TIR image. From the image, four features such as mean, standard deviation, entropy, and busy-ness were obtained. Based on these features, clouds were segmented using k-means clustering algorithm. Finally, using a threshold value, cloudy pixels are extracted. Generally Euclidean distance is used in k-means clustering, but in this paper two more types of distances, Manhattan and Mahalanobis, have been used and the results have been observed using skill score analysis.

Keywords: *convective cloud detection, TIR images, segmentation, k-means clustering, Mahalanobis distance*

I. INTRODUCTION

Floods are one of the major disasters in India and cause heavy amount of damages. If a flood event can be predicted much before its occurrence then taking preventive measures can minimize the amount of damage. Detection and identification of rain clouds play a key role in predicting heavy rainfall that can eventually lead to flood in a particular area. Satellite images have been used for detection of heavy rain clouds because it provides useful information without the requirement of visiting the area physically.

The convective clouds can be identified from thermal infrared (TIR) images (10.5-12.5μm) because clouds are associated with extremely low temperature (Mandal et al. 2005; Turiel et al. 2005). Ample numbers of literature are available in the area of cloud detection from satellite images.

The methods for detection of cloud from IR images can be viewed in two broad categories. The first one includes those that are based on the brightness temperature obtained from the IR images (Arnaud et al. 1992; Carvalho & Jones 2001; Donovan et al. 2008; Endlich & Wolf 1981; Feidas & Cartalis 2005; Levizzani & Setvak 1996; Raut et al. 2008; van Hees et al. 1999); however, the other one consists of those models that directly identify cloud from gray level values of IR images (Azimi-Sadjadi & Zekavat 2000; Brad & Letia 2002; Brad & Letia 2002a; Das et al. 2009; Mandal et al. 2005; Mukherjee & Acton 2002). The second category models exploit the fact that high gray level values in an IR image represent areas of low temperature.

In this paper, dense rain clouds are identified from a TIR image using k-means clustering algorithm with Euclidean, Manhattan, and Mahalanobis distance and results are compared using a skill score.

This paper is organized into following five sections. Section 2 describes feature extraction and segmentation. Section 3 describes coldest cluster extraction. Section 4 compares the result and finally Section 5 concludes the paper.

II. SEGMENTATION USING K-MEANS CLUSTERING

Since, the objective is to extract the cloudy pixels so a hard clustering technique is more suitable. Here the simple k-means clustering was used for segmentation (Hartigan et al., 1979). Fig 1 shows a sample TIR image

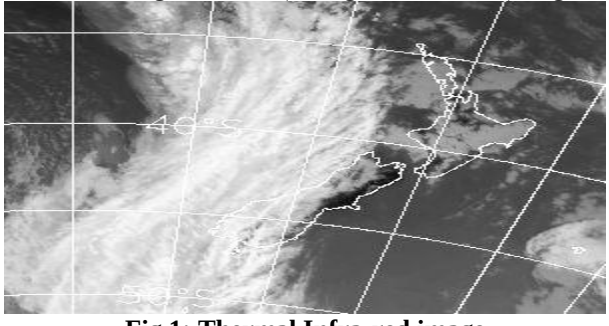

Fig 1: Thermal Infra-red image

Four important features like mean, standard deviation, busyness and entropy (Mandal et al. 2005) were computed using eight neighborhoods, before the segmentation of the image. Busyness gives the direction of variation in intensity. Entropy measures the local homogeneity. For each pixel, a feature vector of length four was obtained. Then the entire image was segmented based on these features using k-means clustering.
The k-means clustering partitions the feature matrix into k clusters where each pixel belongs to any one of the clusters. Every cluster is represented by a centroid, which is the mean of all the pixels belonging to the cluster.
Generally, k-means is implemented using Euclidean distance. But in this study two more distances, Manhattan and Mahalanobis, have been used with k-means.

Euclidean Distance
It is the most common type of distance. It examines the root of square differences between coordinates of a pair of objects. Distance between points p ($p_1,p_2,\ldots,p_n$) and q ($q_1,q_2,\ldots,q_n$), where n is the dimension of the points, can be defined as follows:

$$d(p,q) = \left(\sum_{i=1}^{n} (p_i - q_i)^2 \right)^{1/2} \qquad (1)$$

Manhattan Distance
Also known as city block distance, taxicab distance, absolute value distance. It represents distance between points in a city road grid. It examines the absolute differences between coordinates of a pair of objects. The distance between two points p ($p_1,p_2,\ldots,p_n$) and q ($q_1,q_2,\ldots,q_n$) is:

$$d(p,q) = \sum_{i=1}^{n} \left| p_i - q_i \right| \qquad (2)$$

Mahalanobis Distance
In statistics, Mahalanobis distance is a distance measure introduced by P.C. Mahalanobis in 1936 (Mahalanobis, 1936). It measures the separation of two groups of objects. Suppose we have two groups x_i and x_j, then Mahalanobis distance between them is given by-

$$d_{ij} = \left((\overline{x_i} - \overline{x_j})' S^{-1} (\overline{x_i} - \overline{x_j}) \right)^{1/2} \qquad (3)$$

where, $\overline{x_i}, \overline{x_j}$ are means of the two groups x_i and x_j, respectively. S is the covariance matrix of the groups.

Extraction of Coldest Segment
After clustering, the cluster having highest centroid value is selected which comprises of pixels having very high values. Because high pixel values means coldest and that in turn means dense cloud (Mandal et al. 2005).
Fig 2, 3, 4 shows the coldest segment when the distance used is Euclidean, Manhattan, Mahalanobis, respectively.

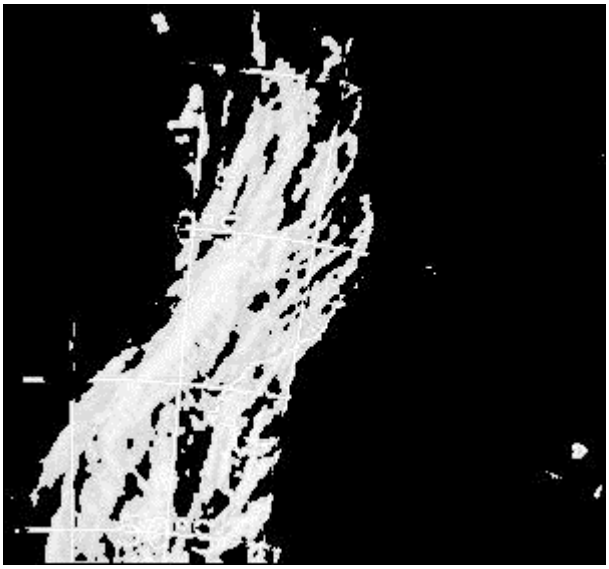

Fig 2: Coldest segment - Euclidean distance used in k-means clustering.

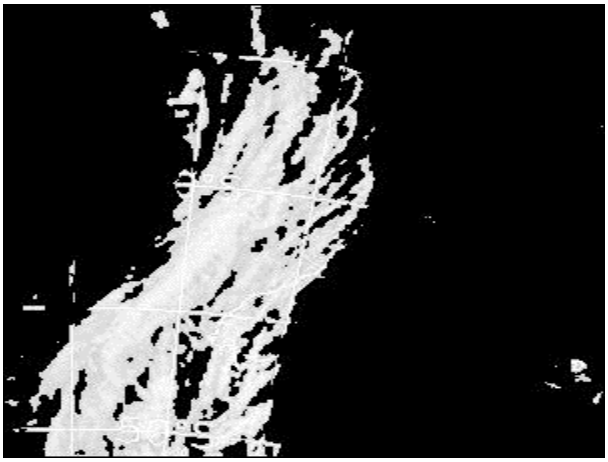

Fig 3: Coldest segment - Manhattan distance used in k-means clustering.

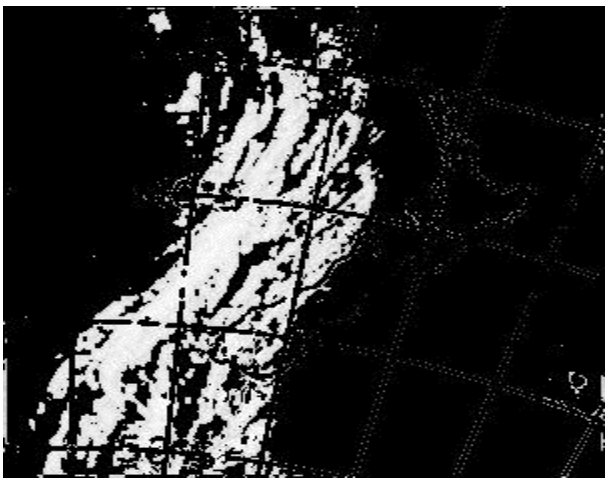

Fig 4: Coldest segment - Mahalanobis distance used in k-means clustering.

III. COMPARISON OF THE RESULTS

The skill score analysis was done by using a Cluster validity technique, Dunn's index (Theodoridis et al., 2006). Dunn's index is calculated as-

$$D_m = \min_{i=1..m} \left\{ \min_{j=i+1..m} \left(\frac{d(C_i, C_j)}{\max_{k=1..m} (diam\ (C_k))} \right) \right\} \quad (4)$$

If image X contains *m* clusters then dissimilarity function between two clusters C_i and C_j is-

$$d(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y) \quad (5)$$

And diameter of cluster C is-

$$diam\ (C) = \max_{x,y \in C} (d(x, y)) \quad (6)$$

If X contains compact and well-separated clusters, Dunn's index will be large, since the distance between the clusters is expected to be "large" and the diameter of the clusters is expected to be "small". Dunn's index $D_m >$ 1 means clustering contains compact and well-separated clusters.

When Euclidean distance was used with k-means clustering D_m was 0.1018, with Manhattan distance D_m was 0.0977 and clustering with Mahalanobis distance gave Dunn's index as 1.0007.

Mahalanobis distance gave Dunn's index > 1 so, clearly it is better than Euclidean distance to use with k-means clustering.

IV. CONCLUSION

Infra red image has been clustered to identify heavy rain clouds. K-means clustering has been used for the purpose. Euclidean distance, Manhattan distance, and Mahalanobis distance has been used with k-means clustering and the results have been compared. Dunn's index for clustering validity suggests that Mahalanobis distance, in comparison to Euclidean and Manhattan, is better for k-means clustering. After using this method dense cloud can be identified for predicting heavy rain that can lead to flood.

REFERENCES

[1]. **Arnaud, Y., Desbois, M. & Maizi, J. (1992).** Automatic Tracking and Characterization of African Convective Systems on Meteosat Pictures. AMS Journal of Applied Meteorology, vol. 31, no. 5, pp. 443-453.

[2]. **Azimi-Sadjadi, M. R., & Zekavat, S. A. (2000).** Cloud Classification using Support Vector Machines. Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS), vol. 2, pp. 669-671.

[3]. **Brad, R. & Letia, I. A. (2002).** Cloud Motion Detection from Infrared Satellite Images. IEEE Proceedings of Second International Conference on Image and Graphics, vol. 4875, no. 1, pp. 408-412.

[4]. **Brad, R. & Letia, I. A. (2002a).** Extracting Cloud Motion from Satellite Images. IEEE Proceedings of 7th International Conference on Control, Automation, Robotics and Vision, vol. 3, pp. 1303-1307.

[5]. **Carvalho, L. M. V. & Jones, C. (2001).** A Satellite Method to Identify Structural Properties of Mesoscale Convective Systems Based on the Maximum Spatial Correlation Tracking Technique. Journal of Applied Meteorology, vol. 40, pp. 1683-1701.

[6]. **Das, S. K., Chanda, B., & Mukherjee, D. P. (2009).** Prediction of Cloud for Weather Now-casting Application using Topology Adaptive Active Membrane. PReMI '09 Proceedings of the 3rd International Conference on Pattern Recognition and Machine Intelligence, pp. 303-308.

[7]. **Donovan, M. F., Williams, E. R., Kessinger, C., Blackburn, G., Herzegh, P. H., Bankert, R. L., Miller, S. & Mosher, F. R. (2008).** The Identification and Verification of Hazardous Convective Cells over Oceans using Visible and Infrared Satellite Observations. Journal of Applied Meteorology and Climatology, vol. 47, pp. 164-184.

[8]. **Endlich, R. M. & Wolf, D. E. (1981).** Automatic Cloud Tracking Applied to GOES and METEOSAT Observations. Journal of Applied Meteorology, vol. 20, pp. 309-319.

[9]. **Feidas, H. & Cartalis, C. (2005).** Application of An Automated Cloud-tracking Algorithm on Satellite Imagery for Tracking and Monitoring Small Mesoscale Convective Cloud Systems. International Journal of Remote Sensing, vol. 26, no. 8, pp. 1677-1698.

[10]. **Hartigan, J.A. and Wong, M.A. (1979).** Algorithm AS 136: A k-means Clustering Algorithm. J. of Royal Statistical Society, Series C (Applied Statistics), Vol. 28, No. 1.

[11]. **Levizzani, V. & Setvak, M. (1996).** Multispectral, High-resolution Satellite Observations of Plumes on Top of Convective Storms. Journal of the Atmospheric Sciences, vol. 53, no. 3, pp. 361-369.

[12]. **Mahalanobis, P.C. (1936).** Mahalanobis distance. Proc. Of the National Institute of Science of India, 49(2), pp 234-256.

[13]. **Mandal, A.K., Pal, S., De, A.K. and Mitra, S. (2005).** Novel Approach to Identify Good Tracer Clouds from a Sequence of Satellite Images. IEEE Transactions on Geoscience and Remote Sensing, Vol. 43, No. 4.

[14]. **Mukherjee, D. P. & Acton, S. T. (2002).** Cloud Tracking by Scale Space Classification. IEEE Transactions on Geoscience and Remote Sensing, vol. 40, no. 2, pp. 405-415.

[15]. **Raut, B. A., Karekar, R. N. & Puranik, D. M. (2008).** Wavelet-based Technique to Extract Convective Clouds from Infrared Satellite Images. IEEE Geoscience and Remote Sensing Letters, vol. 5, no. 3, pp. 328-330.

[16]. **Theodoridis, S. and Koutroumbas, K. (2006).** Pattern Recognition. Academic Press, London, U.K., pp 752-753.

[17]. **Turiel, A., Grazzini, J. & Yahia, H. (2005).** Multiscale Techniques for the Detection of Precipitation using Thermal IR Satellite Images. IEEE Geoscience and Remote Sensing Letters, vol. 2, no. 4, pp. 447-450.

[18]. **Van Hees, R. M., Lelieveld, J. & Collins, W. D. (1999).** Detecting Tropical Convection using AVHRR Satellite Data. Journal of Geophysical Research, vol. 104, no. D8, pp. 9213-9228.