International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 04, April-2015

An Investgation in to Rainfall over Kerala using Satellite Images Analyzed using Discrete Cosine Transform

Rishidas S
Associate Professor,
Department of Applied Electronics and Instrumentation
Govt. College of Engineering,
Kozhikode Kerala

Maheswary V
Department of Applied Electronics and Instrumentation
Govt. College of Engineering,
Kozhikode, Kerala

Sujith Kumar T P
Asst Professor,
Department of Applied Electronics and Instrumentation
Govt. College of Engineering,Kozhikode,kerala

***Abstract*—Rain, the liquid form of precipitation, plays a significant role in hydrologic cycle. This work is to do rainfall estimation based on satellite images in visible channel using classification techniques. The features of satellite images are taken using the image moments, so that reliable rainfall prediction of the day to be a rainy day or not is made using same day images.**

Keywords –Visible channel, image moments, SVM, DCT

I. INTRODUCTION

India is a tropical country with diverse climatic features depending on the monsoon. The word 'monsoon' is derived from an Arabic word '*mausim*', meaning the seasonal reversal of winds. Apart from the direction of the rain bearing winds, the presence of a long coastal plain and the position of the mountain ranges in India also constitute the incredible climatic variations in India.

The main landmass of India lies between 8°4' N and 37°6' N latitude and 68°7' E and 97°25' E longitude i.e., the southern India and some parts of northern India lies in torrid zone and rest in temperate zone. The coastline of India is about 7500 km long and the regions nearer to the sea gets moderate climate throughout the year. So the larger geographical region of India comes under the trophies and the climate is greatly influenced by the Monsoon. These endow the country with a tropical monsoonal climate. Only regions that are far away from the sea have extremes.

Unlike other states of India, Kerala has the most gifted geography which lies stretched north –south along the coast of Arabian Sea. Its western coastal plain is about 580 km long with a varying width of 35 to 120km. The Western Ghats borders the eastern boundary of the state. The strip of hills and valleys, close to the Ghats, encompasses of steep mountains and deep valleys covered with dense forests. The diverse climate of Kerala can be divided into 4 seasons – summer, winter, North-east monsoon and south west monsoon.

The short range climatic condition of atmosphere is called weather. The science of weather is known as meteorology. It is an integrated inter-disciplinary system of land, ocean and atmosphere. The basic aspects of meteorology are observation, understanding and prognosis of weather. The weather has been predicted conventionally by common man using observation and by nineteenth century, with the advent of science and technology, it has grown to higher dimensions. The world networks the weather observations through fast communications channels. The weather charts are analyzed by professional meteorologists at forecasting centers. Weather forecast models are then made with the help of modern computers and supercomputers. Weather forecast and timely reports are of key importance to many activities like agriculture, industrial projects, defence, shipping, aviation, fisheries, tourism, water management and disaster mitigation. Recent advances in computer technology and satellite have led to noteworthy headway in meteorology. However, still the weather forecast and information are incomplete.

Kerala being a land with long coastal area, back waters and rivers, a sizable population depends on marine and freshwater fishing. The accurate prediction of rainfall has a great importance on this front. Navigation also demands an efficient weather forecast, particularly, rainfall.

A. Satellite imagery and their role in weather forecasting:

The satellite imagery has an important role in modern weather prediction systems. Satellite images are taken using two types of sensors. Visiblelight sensor called the 'imager', gathers information on cloud movements and patterns. The different surface features reflect light in different ways and thus can be distinguished on images. The second one is an infrared sensor called the 'sounder'. This sensor measures the amount of energy radiated by the earth's crust, air, clouds and so on. Unlike the former, IR sensors can be used at night as it does not depend on daylight. Depending on the sensor being used to capture the information, images are categorized to visible channel, IR channel and water vapour channel.

For the entire world, the large weather systems are heavily depended on weather satellites. They show cloud patterns, large weather events like hurricanes, etc. They help forecasters to foresee weather across the whole earth including poles, land and water. Modern satellite data is detailed, even to the perspective of states (district wise) and countries. These data are given to super computers, which use statistical equations to create forecast models of the atmosphere. These forecast models may be correct failing which causes a wrong prediction.

The work proposed in this paper has application in weather prognosis especially rainfall prediction. This method uses the principles of pattern recognition for classifying the day to be a rainy day or not. The patterns are collected from satelliteimages for the rainfall prediction. It's a short range prediction method, which uses same day visible spectrum images and the image samples are transformed using Discrete Cosine Transform (DCT). The classification results are compared with spatial domain classification.

II. PROPOSED WORK

The proposed work is an application of pattern classification using support vector machine(SVM). The schematic block diagram is shown in figure 1. The main stages are data collection, data reduction and classification.

A. Data collection:

For the work documented here, the samples are collected from Indian Meteorology Department website (www.imd.gov.in)[6] and also archive images are collected from the website of University of Wisconsin-Madison Space Science and Engineering Centre [7]. Samples of images are shown in figure 1 (taken by the satellites, Kalpana 1 and INSAT – 3D at 6.45 am, +5.30 IST).

The process of image capturing and transmission involves noise such as speckle noise. Therefore pre-processing is needed. The images are subdivided into sub images so that each sub-image contributes to the feature vector.

B. Data Reduction:

1) Feature Extraction:

Feature is any minute attribute of a signal which can represent the signal proficiently reduced dimension. This process is called feature extraction.When the data input is very large to be processed or is doubted to be redundant, it can be characterized with a reduced set of features called features vector.The feature vector so obtained can be used to perform the required task instead of the complete initial data. Feature is the minute attribute of an image giving its characteristics. The first, second, third and fourth order image moments [8] are taken as the measures for feature extraction.

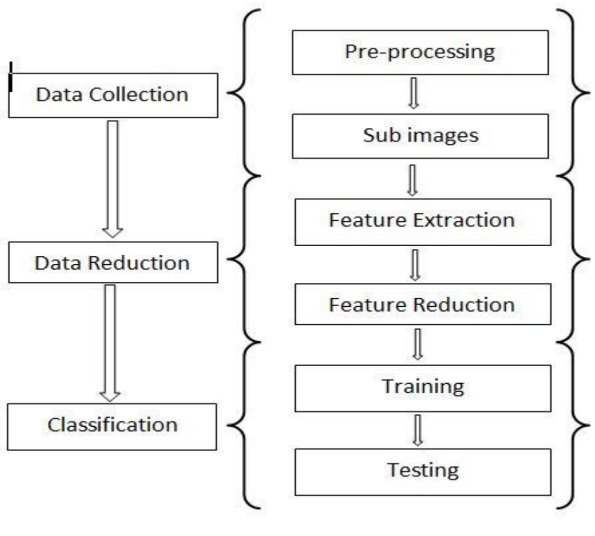

Fig 1 Schematic block diagram of proposed method

Image moments are measures that can be used differentiate images based on their pixel intensities, providinga measurement for intensity similarity between images.

a. Mean: Mean can be understood as the average intensity value of image.

$$mean = \frac{1}{N}\sum_{i=1}^{N} x_i$$

b. Variance: A measure of average variability of the data points and standard deviation is the square root of the variance of the distribution.

$$Var(x1 \ldots xN) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2$$

c. Skewness: Skewness is defined with the equation given below and it can be interpreted as a measure of the degree of asymmetry in the distribution.

$$Skew(x1 \ldots xN) = \frac{1}{N}\sum_{j=1}^{N}\left[\frac{x_j - \bar{x}}{\sigma}\right]^3$$

d. Kurtosis: The kurtosis is also a non-dimensional quantity. It measures the relative peaked-ness or flatness of a distribution

$$Kurt(x1 \ldots xN) = \frac{1}{N}\sum_{j=1}^{N}\left\{\left[\frac{x_j - \bar{x}}{\sigma}\right]^4\right\} - 3$$

On extracting these four moments of sub-images makes the vector of large length. Therefore the dimensions are reduced to significant features, using principal component analysis.

2) Principal component analysis:

Principal component analysis (Karhunen-Loeve or Hotelling transform) [4] - PCA belongs to lineartransforms based on the statistical techniques. This is a powerful tool fordata analysis and pattern recognition, often used in signal and image processing, as a technique for data compression, data dimension reduction or their decorrelation. Properties of PCA can be used for characterizing selected object orientation or its rotation.

PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components. The principal components are orthogonal as they are the eigenvectors of the covariance matrix.

3) Discrete Cosine Transform (DCT):

DCT is an orthogonal transformation that is very widely used in image compression and is widely accepted in the multimedia standards. DCT belongs to a family of 16 trigonometric transformations. The type-2 DCT transforms ablock of image of size N x N having pixel intensities s (n1, n2) into a transform array of coefficients S (k1, K2), described by the following equation:

$$S(k_1, k_2) = \sqrt{\frac{4}{N^2}} C(k_1) C(k_2)$$

$$\sum_{m_1=0}^{N-1} \sum_{n_2=0}^{N-1} s(n_1, n_2) \cos\left(\frac{\pi(2n_1+1)k_1}{2N}\right) \cos\left(\frac{\pi(2n_2+1)k_2}{2N}\right)$$

Where $k_1, k_2, n_1, n_2 = 0,1,2, \ldots N-1$ and

$$C(k) = \begin{cases} \frac{1}{\sqrt{2}} \; for\ k = 0 \\ 1 \; otherwise \end{cases}$$

The transformed array $S(k_1, k_2)$obtained through equation is also of the size N x N, same as that of the original image block.

C. Classification:

The statistical learning theory provides an outline for studying the problem of gathering information, predictions and making decisions from a set of input data. In short, it aids the choosing of the hyper plane space in such a way that it closely characterizes the underlying function in the target space.

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. SVM is a kernel-based algorithm. A kernel is a function that transforms the input data to a high-dimensional space where the problem is solved. Kernel functions can be linear or nonlinear. In practical use of SVM, the user specifies the kernel function. Kernel function, being an inner product, is really a similarity measure between the objects.

- Linear kernel: reduces to a linear equation
- Polynomial kernel:

$$K(x, y) = (x^T y + 1)^d$$

- Radial Basis Function : with width σ

$$K(x, y) = \exp(-(\|x - y\|^2 / 2\sigma^2))$$

The extracted data are trained using Support Vector Machine (SVM). The training set is 80% of the data and testing set is 20%. A rainy day is taken as the positive class and sunny day without rain is taken as the negative class.

III. EXPERIMENTAL RESULTS

Visible spectrum image samples of have been collected form the Indian meteorology department website.The images samples are from the month of June – December which includes the two monsoon seasons in India, i.e., the North East monsoon and South West Monsoon. The region of interest is cropped, i.e., latitude 0-15 and longitude 70-85.This includes regions of Kerala, the Western Ghats and the Arabian Sea. The cloud patterns are obtained from the visible spectrum images. Also archive images are collected for training as well as testing from the website of University of Wisconsin-Madison Space Science and Engineering Centre. These are GIF images, taken at 6.45am (IST +5.30). A median filter is applied for pre-processing. The images are resized and subdivided. The four central image moments are taken as the features. Since the so obtained vector is higher dimensional, a dimensionality reduction is done using principal component analysis (PCA). On selecting, the significant features using PCA, the feature vector dimension can be reduced considerably. Plot of Eigen values are shown in figure 2 and 3, from which it can be seen that the significant values are of the order of 70. The similar feature extraction and reduction has been performed for the transformed domain vectors obtained after applying two-dimensional Discrete Cosine Transform (DCT). The feature vector has been reduced to a 65 dimensional vector.

The extracted data are trained using Support Vector Machine (SVM). The training set is 80% of the data and testing set is 20%. A rainy day is taken as the positive class and sunny day without rain is taken as the negative class, therefore applying a binary classifier. As per the standard classification of rainfall, a day with less than 10mm rain is taken as a negative class and the days having more than 10mm of rain is taken as positive class.

The extracted feature vectors of about 196 day-samples are the patterns taken for classification. 118 samples are taken as the training set with a dimension of 70 features. 39 samples are taken as validation data for cross validation of kernels and the remaining 39 samples are taken as the testing data set.

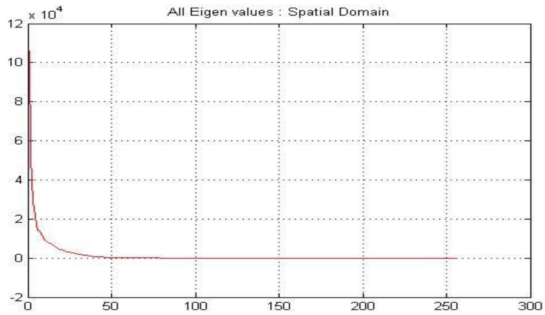

Fig. 2 Plot of Eigen Values of Covariance matrix in Spatial Domain

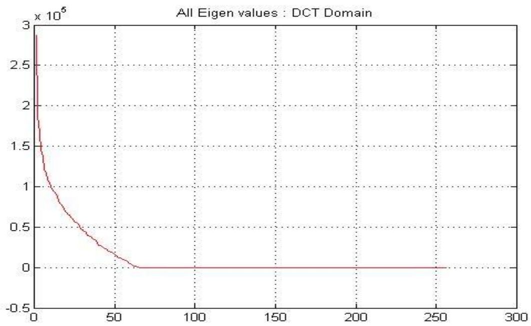

Fig. 3Plot of Eigen Values of Covariance matrix in DCT Transformed Domain

Table 1 SVM results for Test data in Spatial Domain

Cross Validation	Linear Kernel	Polynomial Kernel with d=2	Polynomial Kernel with d=3	Polynomial Kernel with d=5	Polynomial Kernel with d=8	RBF with Std= 1	RBF with Std= 0.1	RBF with Std= 0.01	RBF with Std= 0.001	RBF with Std= 2
100	61.54	38.46	41.03	35.9	33.33	66.67	100	66.67	66.67	66.67
500	48.72	35.9	41.03	35.9	33.33	69.23	100	66.67	66.67	66.67
10	66.67	46.15	30.77	35.9	33.33	66.67	100	66.67	66.67	66.67
0.1	66.67	46.15	64.1	35.9	33.33	66.67	66.67	66.67	66.67	66.67
0.01	66.67	46.15	66.67	35.9	33.33	66.67	66.67	66.67	66.67	66.67

Table 2 SVM results for Test data in DCT Domain

Cross Validation	Linear Kernel	Polynomial Kernel with d=2	Polynomial Kernel with d=3	Polynomial Kernel with d=5	Polynomial Kernel with d=8	RBF with Std= 1	RBF with Std= 0.1	RBF with Std= 0.01	RBF with Std= 0.001	RBF with Std= 2
100	43.59	33.33	35.9	38.46	41.03	41.03	66.67	66.67	66.67	33.33
500	43.59	33.33	35.9	38.46	41.03	43.59	66.67	66.67	66.67	43.59
10	51.28	35.9	35.9	38.46	41.03	43.59	66.67	66.67	66.67	64.1
0.1	66.67	35.9	35.9	38.46	41.03	43.59	66.67	66.67	66.67	64.1
0.01	66.67	35.9	35.9	38.46	41.03	43.59	66.67	66.67	66.67	64.1

The experimental results are shown in the Table 1and table 2. The tables show the test results of different kernels in spatial domain as well as the cosine transformed domain. These include the cross validated classification accuracy. From the above tables it is seen that for spatial domain, the classification accuracy is maximum (=100 %) for the Radial Basis Kernel with a standard deviation equal to 0.1. When transformed with discrete cosine transform, the maximum accuracy is obtained on using RBF kernel, though comparatively less with respect to spatial domain. On cross validation, the accuracy improved for all the kernels.

IV. CONCLUSION AND FUTURE WORKS

Prognosis of rainfall with same day visible channel image gives a trust worthy result on testing. The features used for generating patterns of satellite images are the first, second, third and fourth order moments of the images in spatial and DCT domain.

Also the work so far deals with same day images for short range prediction which can be extended by using previous days' images for prediction so that forecasting is done one or two days in advance. This requires more precise feature vector using other feature extraction methods, including higher order moments, co-occurrence matrix as well as transforms like Discrete Wavelet Transform (DWT). Also images taken at different instances of the day can be used as more samples per day. This work implements a binary classifier, which can be extended as a multi class problem, taking the amount of rain on each day, thereby giving an accurate prediction with estimated rainfall.

REFERENCES

[1] V.S. Rathnayake, H.L. Premaratne1 and D.U.J. Sonnadara, "Performance of neural networks in forecasting short range occurrence of rainfall", June 2011

[2] B.Syam Prasad Reddy, K.Vagdhan Kumar, B.Musala Reddy, N.Raja Nayak, "ANN Approach for Weather Prediction using Back Propagation", International Journal of Engineering Trends and Technology- Vol 2012

[3] Anil K. Jain, "Statistical Pattern Recognition: A Review", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 1, January 2000

[4] Gonzalez and Woods, "Digital Image Processing 3rd Ed. (DIP/3e)"

[5] Simon O. Haykin, "Neural Networks and Learning Machines"

[6] Govt. of India, Indian Meteorological Department, http://imd.gov.in/

[7] Space Science & Engineering Centre, University of Wisconsin-Madison: http://www.ssec.wisc.edu/data/kalpana

[8] Noah Keen, "Color Moments" February 10, 2005.