International Journal of Pure and Applied Mathematics
Volume 118 No. 8 2018, 191-196
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

Implementation Of Classification System Using Density Clustering Based Gray Level Co Occurrence Matrix (DGLCM) For Green Bio Technology

P.Santhi
Professor, Department of CSE
M. Kumarasamy College of Engineering
Karur, Tamilnadu,India.
Mail-Id : santhipalanisamypec@gmail.com

R.Vikram
Assistant professor, Department of CSE
M. Kumarasamy College of Engineering
Karur, Tamilnadu, India.
Mail- Id:vikramr.cse@mkce.ac.in

***Abstract* - Green Bio Technology is the most focused research area in image processing for improving the process of agriculture. Herewith image classification is used to solve the problems identified in green bio technology. Classification is a process of grouping the similar data based on its features. Images are having the different types of features, i.e. Texture, shape and color. Normally, image classification is performed using Feature extraction, Selection and similarity measurements. In classification, most of the researchers are using the Gray level Co Occurrence Matrix (GLCM) for texture feature extraction. The main drawback of existing algorithm of GLCM is global features generation. To overcome the above mentioned drawback, this paper proposes Density clustering based GLCM (DGLCM) for extracting the local features of an in image. When compared to global, the local features are having more information of an image. The similarities of features are calculated using Euclidean distance measure. Finally, the features are classified using Fisher kernel based Support Vector Machine (FSVM).The performance of this algorithm is calculated using precision and recall. The result is evaluated using the images collected from the agricultural processes.**

Keywords – Texture; Co-Occurrence; Support Vector Machine; Selection; Extraction.

I. INTRODUCTION

Agriculture is a backbone for all countries. Now days most of the agriculture lands are destroyed for building construction and some of the lands are not used for agriculture. This paper mainly focuses to increase the agriculture process in all the areas. This process is established using image classification technique.

Image classification has earned more attention due to the initiation of modern day applications involving image based information. Nowadays lot of research is being carried out in this field. Image has been classified or the set of images have been retrieved based on the salient features of an image such as texture, shape and color. Image classification is an important area of research in the applications like bio technology, medical, military applications and etc. This paper focuses on texture based image extraction for image classification.There are three approaches to analyse an image texture which are namely structured approach, transform approach and statistical approach. In structured approach, the image texture is viewed in the form of primitive texture in some regular or repeated patterns. Statistical approach is the widely used technique to find out the texture of an image in which the texture is viewed as a quantitative measure of the arrangement of intensities in a region. The methods used in statistical approach are edge detection, co-occurrence matrices, laws texture energy measures, auto correlation and power spectrum.

In statistical method, GLCM and tamura features are used to collect only the global features of an image. As the nearest pixels are not the same, when GLCM is applied on whole image, it leads to many zero frequencies in the offset. Gabor filter and wavelet transform are the methods used for the texture model transformation. The spectrum model contains Local Binary Pattern (LBP), Local Derivative Pattern (LDP) and Local Ternary Pattern (LTP). Figure 1 shows different methods for texture extraction and color feature extraction.

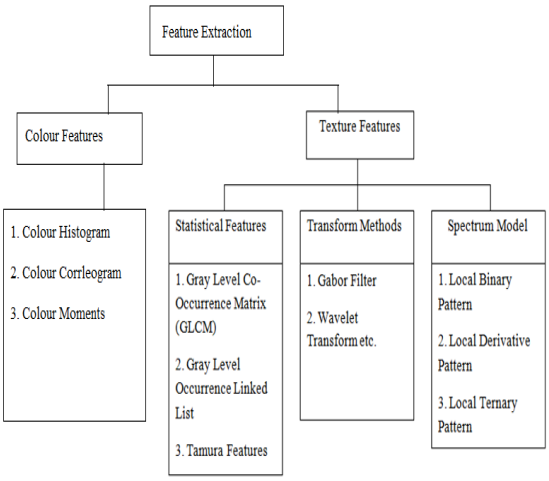

Fig 1. Different types of Feature Extraction

The extracted features are classified using classification.Classification process uses supervised learning technique. Here, the training set is given in the form of description or features that can be used to predict the previously unseen images. This supervised classification uses two third of images for training and one third of images for testing. Supervised learning contains many algorithms such as Support Vector Machine, Bayesian Classification, Naïve bayes classifier, Neural Network based classification, Fuzzy rule based classifier, etc. In this paper concentrates on GLCM for feature extraction and support vector machine based classifier for classification.

II. RELATED WORKS

Commonly used texture extraction techniques are co-occurrence matrix, fractals, gabor filters and variations of wavelet transform (Saha et al 2004). The texture spectrum is introduced to characterize the edge information and co-occurrence matrix properties. Image texture is achieved by modelling the image feature into two dimensional arrays. This array is called as Gray level Co- occurrence matrix (GLCM) [6].

The GLCM matrix is used to obtain texture feature from an image. Many features such as angular second moment, correlation, inverse difference moment and entropy can be extracted using GLCM. These features provide high dimensional accuracy to find the patterns of an image [3]. GLCM is widely used in many applications. Haralick proposed gray-level co-occurrence matrices (GLCM) to extract the features from an image. Main limitation of GLCM is that, it is computationally very intensive and it produces only the global features of an image.

Generally, co- occurrence matrices are sparse matrix and many of the calculations are done using unnecessary zero frequencies [4]. Bastos has proposed a novel method called gray level co-occurrence linked list (GLCLL) which stores only the non – zero frequencies. Howarth et al (2004) considers three different approaches for computing the texture of an image. They are co-occurrence matrix using statistical approach, feature computation using Tamura's features and signal processing using Gabor wavelets.

In remote sensing, dimension of a feature space increases significantly which increases the complexity of the classification process. SVM classifier also solves the problems encountered in multi-class classification [1]. Yongjun Maet al(2002) have proposed the ISD (Integrated SVM and Distance Calculation) algorithm for texture based image classification. These two algorithms act as two serial layer of classification. SVM algorithm is used in traditional classification system but it consumes more time for processing the images [9]. Distance calculation is mainly used to avoid the problems faced in SVM and to classify the rejected input images using rejection rule [8].

Jianming Li (2008) proposed the fuzzy SVM method for classification. It provides a membership function to classify the images which are unclassifiable using conventional SVM. Image feature histogram is used to input the SVM and the fuzzy SVM. SVM is used to find out the normal and abnormal brain in terms of brain tumor. It uses a feature vector of MRI to classify the brain [2]. It considers the kernel constituted by fuzzy basis functions. It combines both the characteristics of SVM and fuzzy SVM in order to provide high generalization and better performance [7].

III. OBJECTIVE

The main objective of this research work is as follows:

- To produce the local information of texture extraction, the density based clustering is combined with GLCM features.
- To handle the non linear data and to reduce the training time of classifier, the Fisher kernel based Support Vector Machine (FSVM) is with DGLCM.

IV. METHODOLOGIES

A. Density Clustering based GLCM

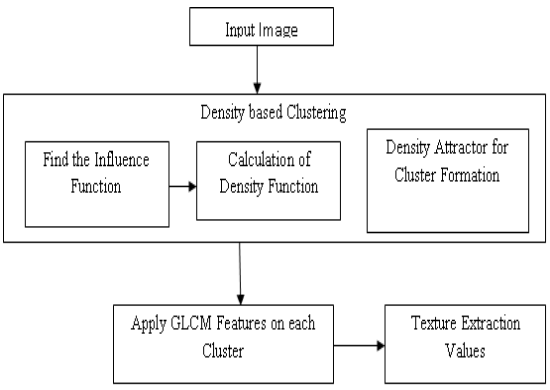

Fig.2. Texture Extraction

Normally, the photographic images contain lot of noises due to its disordered values. These noises are eliminated and the disordered images are converted into ordered images using distribution function (Cao et al 2006). This function is used during clustering for grouping similar pixels of an image. In DENCLUE method, the clustering is based on a group of density distribution function. Rosenblatt (1956) & Parzen (1962) describe a kernel function for converting the images into a distribution function. This method is based on non-parametric models. Here, the distribution function is calculated without any prior knowledge. The following ideas are used in density clustering algorithm (Tan et al 2010).

- The Influence function for each data point is to be calculated using mathematical function and it describes the influence of data points to its neighbour.
- Density of data space is modeled as a total influence function of each data point.
- The clustering can be done using density attractor.

Using GLCM, Haralick proposed thirteen statistical features which are known as Haralick texture features.These features are computed from the clustering image. The five features are as follows:

Angular second moment (ASM) feature

This feature is used to measure the uniformity or energy from an image. Angular Second Moment is very large, if the pixels are very similar.

$$\text{Angular Second Moment} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{d,\theta}(i,j)^2 \quad (1)$$

Contrast Feature

Generally, contrast is defined as a difference between the color and the brightness of an object. Contrast is a measure of intensity or gray level variations between current pixels and its neighbor.

$$\text{Contrast} = \sum_{n=0}^{N-1} n^2 \left\{ \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{d,\theta}(i,j) \right\}, \text{where } n = |i-j| \quad (2)$$

If the value of i and j are equal, then the cell is on the diagonal. These values indicate that the pixels entirely similar to their neighbor. Here the weight 'n' is also 0. If the difference between i and j is 1, then there is a small contrast between the pixels and the weight is 1. If i and j are differed by 2, then the pixels have more contrast with its neighboring pixels. Here, the weight value is 4. The weight and the value of i-j are directly proportional to each other. The value of i-j increases with the weight.

Entropy Feature

Entropy is a term used in thermodynamics to measure the proximity among the systems or it is used to define the disorder between systems. Higher entropy means greater disorder values. The following Equation is used to define entropy:

$$\text{Entropy} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{d,\theta}(i,j)\log(P_{d,\theta}(i,j)) \quad (3)$$

Variance Feature

Variance is one of the moments in the distribution. It is used to measure the variation between the pixels. It is similar to entropy measure. It is always a positive value. If the value of variance is zero, it indicates that the values are identical. A small variance indicates that the pixels are very close to the mean and the high variance indicates that the pixels are spread around the mean and from each other.

$$\text{Variance} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} (i-\mu)^2 P_{d,\theta}(i,j) \quad (4)$$

Correlation Feature

It shows the linear dependency of gray level values in the co- occurrence matrix. It is used to refer the reference pixels with its neighbour. It takes the value of 0 for no correlation and 1 for perfect correlation.

$$\textit{Correlation} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{d,\theta}(i,j)\frac{(i-\mu_x)(j-\mu_y)}{\sigma_x\sigma_y} \quad (5)$$

a. Fisher Kernel Based Support Vector Machine (FSVM)

It is one of the methods for classification of both linear and non linear data. SVM method uses non linear mapping to transform the original data into a higher dimension. In this new dimension, it searches the linear optimal separating hyperplane. Linear SVM is also extended to nonlinear SVM for the classification of linearly inseparable data. Two main steps involved in this approach are as follow:

i) Transform the original input into high dimensional space using nonlinear mapping.

ii) Find out the linear separable hyperplane in new space.

Consider a 3D input vector X= (x_1, x_2, x_3) which is mapped to a high dimensional space 'Z' such as $\phi_1(X) = x_1$, $\phi_2(X) = x_2$, $\phi_3(X) = x_3$, $\phi_4(X) = (x_1)^2$, $\phi_5(X) = x_1x_2$ and $\phi_6(X)$

$=x_1x_3$. Thus the nonlinear space is converted into a new space d (Z).

$$d(Z) = w_1x_1 + w_2x_2 + w_3x_3 + w_4(x_1)^2 + w_5x_1x_2 + w_6x_1x_3 \quad (6)$$

$$d(Z) = w_1z_1 + w_2z_2 + w_3z_3 + w_4z_4 + w_5z_5 + w_6z_6 \quad (7)$$

Equation 7 requires more number of dot products for MMH computation. To avoid this, the kernel function $K(X_i, X_j)$ is used in the original input data. The kernel functions are polynomial kernel, Gaussian radial basis function kernel, sigmoid kernel and fisher kernel. This chapter concentrates on fisher kernel method for computation of nonlinear space.

The fisher kernel is defined as follows:

$$K(X_i, X_j) = U^T X_i - U^T X_j \quad (8)$$

Where,

U^T_X – Fisher Score

I – Fisher Information Matrix.

The Fisher score is defined in Equation 9 as follows:

$$U_X = \nabla_\theta \log p\left(\frac{X}{\theta}\right) \quad (9)$$

∇_θ - Gradient function or the partial difference.

B. *Similarity Measurements*

Similarity measurement is a measure to find out similar images from a database using distance metrics. Here, the Euclidean distance is used to measure such similarities between the images. The following formula shows the Euclidean distance to measure the similarities between images.

$$ED = \sqrt{\sum_{i=1}^{n} (FQ_i - FD_i)^2} \quad (10)$$

In Equation 10, FQ_i and FD_i represent the features of query and the database images. The objects in an image are identified using this distance metrics. The object that has less distance between them is identified. Similarly all identical objects are identified from an image.

V. RESULTS & DISCUSSION

This work considers five Haralick features from GLCM at four directions such as 0o, 45o, 90o and 135o. This feature is applied on each cluster for obtaining the local descriptors of an image.

TABLE I. Texture Extraction Values

GLCM Features	Energy	Contrast	Entropy	Variance	Correlation
0^o Degrees	0.0732	0.8465	7.6762	143.7472	0.8786
45^0 Degrees	0.0683	1.04	8.2054	153.5296	0.8505
90^0 Degrees	0.0522	2.0712	8.0372	137.5723	0.7006
135^0 Degrees	0.0679	1.0104	8.7209	138.2476	0.0679

VI. PERFORMANCE ANALYSIS

The performance of an algorithm is analyzed using a ground truth. The ground truth refers to the accuracy of the training set in supervised learning technique. This truth is used in statistical models to prove or disprove the research hypothesis. It uses set of measurements to measure the accuracy of object. The ground truth can be calculated using the measures such as precision and recall. Recall and precision are defined as follows:

Precision is referred as a positive predictive value. It is the number of relevant images that are retrieved from the total number of retrieved images. Equation 11 shows the formula for calculating precision. The recall is a measure of relevant images retrieved from the total number of relevant images from a database. Equation 12 shows the formula for calculating recall.

$$\text{Precision} = \frac{\text{No.of Relevant Images Retrieved}}{\text{No.of Images Retrieved}} \quad (11)$$

$$\text{Recall} = \frac{\text{No.of Relevant Images Retrieved}}{\text{No.of Relevant Images in the DataSet}} \quad (12)$$

Table II. Performance of GLCM Vs DGLCM

GLCM - SVM	Execution Time	1.93s
	Precision %	88.16%
	Recall %	85.34%
	Accuracy %	90.25%
DGLCM - FSVM	Execution Time	1.53s
	Precision %	93.50%
	Recall %	90.60%
	Accuracy %	93.16%

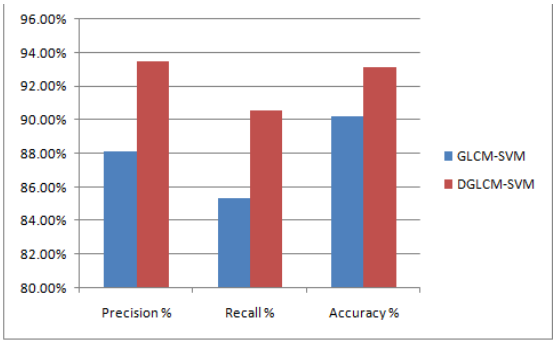

Fig 3. GLCM Vs DGLCM

Figure 3 shows the performance of GLCM and DGLCM.

VII. CONCLUSION

Bio technology is the most interesting area of research in image processing. Different kinds of Bio technology are medical, agriculture and industrial bio technology. This Paper focuses on agriculture bio technology is called green bio technology. In this technology, the classification of agriculture land is difficult. To resolve this problem, image classification system is implemented using

DGLCM for texture extraction and FSVM for classification. The performance of this classification system is calculated using precision and recall. Result shows that the performance of proposed system is having more precision and recall when compared to existing GLCM and SVM.

REFERENCES

[1] Angelos Tzotsos 2006, 'A Support Vector Machine Approach For Object Based Image Analysis', Lecture Notes in Geo information and Cartography, pp. 663-677.

[2] Jianming Li, Shuguang Huang, Rongsheng He & Kunming 2008, 'Image classification based on fuzzy support vector machine', Proceedings of International Symposium on Computational Intelligence and Design, pp.68-71.

[3] Mohanaiah, P, Sathyanarayana, P & GuruKumar, L 2013, 'Image Texture Feature Extraction using GLCM Approach', International Journal of Scientific and Research Publications, vol. 3, no.5, pp. 1-5.

[4] Nitish Zulpe1 & Vrushsen Pawar 2012, 'GLCM Textural Features for Brain Tumor Classification', International Journal of Computer Science, vol. 9, no. 3, pp. 354-359.

[5] Peter Howarth & Stefan Ruger 2004, 'Evaluation of Texture Features for Content-Based Image Retrieval', Proceedings of International Conference on Image and Video Retrieval, pp. 326–334.

[6] Saha, SK, Das, AK & Chanda, B 2004, 'CBIR using perception based texture and color measures', Proceedings of International Conference on Pattern Recognition, vol. 2, pp. 985–988.

[7] Spyrou, E, Stamou, G, Avrithis, Y & Kollias, S 2005, 'Fuzzy support vector machines for image classification fusing MPEG-7 visual descriptors', Proceedings of Integration of Knowledge, Semantics and Digital Media Technology.

[8] Sumiti Bansal & Rishamjot Kaur 2014, 'A Review on Content Based Image Retrieval using SVM', International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 7, pp.232–235.

[9] Yongjun Ma, Tingjian Fang, Kai Fang & Dingcheng Wang 2002, 'Texture image classification based on support vector machine and distance classification', Published in Intelligent Control and Automation, Proceedings of the 4th World Congress, vol.1, pp. 551–554.