Plant Leaf Segmentation Using Non Linear K means Clustering

N.Valliammal[1] and Dr.S.N.Geethalakshmi[2]

**[1] Assistant Professor, Department of Computer Science,
Avinashilingam Institute for Home Science and Higher Education for Women,
Coimbatore, Tamilnadu, India.**

**[2] Associate Professor, Department of Computer Science,
Avinashilingam Institute for Home Science and Higher Education for Women,
Coimbatore, Tamilnadu, India.**

Abstract

This paper presents a new approach for plant leaf image segmentation by applying non linear k means algorithm. The segmentation process presents a clustering mechanism for high resolution images in order to improve the precision and processing time. Plant image, however, always contain complicated background objects that interfere with the examination process and must be removed from the image prior to species classification. K means clustering is applied at the first level of segmentation to detect the structure of the plant leaf. At the second level Sobel edge detector is used to remove the unwanted segments to extract the exact part of the leaf shape. The performance of the proposed method is compared with other traditional methods to analyze the efficiency of the system. Experimental result shows that this new approach simplifies the process to extract shape related features and measurements of the leaf for higher accuracy.

Keywords: *Edge Detection, Sobel Edge Detector, K means clustering, Plant Leaf Identification.*

1. Introduction

The goal of segmentation is to simplify and change the representation of an image into some other representation that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, edges etc.) in images. More precisely, it is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. Therefore, a need for optimal solution is must for the above mentioned approach that has the capability to analyze the pattern of distribution of the pixels.

Currently, major research groups focus on the leaf recognition problem. Leaf vein extraction is an important part of modeling plant organs. Assessing the appearance of plants is an important botanical skill, with many applications, ranging from plant recognition to health diagnosis. When identifying plant species, the whole plant, leaves, flowers, stem and fruit will be observed. Plant leaf has approximately two-dimensional nature. Therefore, it is most suitable for machine processing. As the shape of leaves is one of the most important features for characterizing various plants visually, the study of leaf segmentation will be an important stage for developing a plant recognition system.

The recognition process includes segmentation as the major step, but the segmentation results are not evaluated that good [16]. However, those recognition methods are based on several approaches could also be used for leaf segmentation. The approaches are distinguished as boundary enhancement, clustering, smoothening, edge detection. This work employs an enhancement method with k means clustering method to eliminate the over segmentation and false edges during the segmentation process using sobel edge detection.

Several approaches have been already introduced for image segmentation. The most popular method for image segmentation is K-means algorithm [1][2][5]. It is a widely used algorithm for image segmentation because of its ability to cluster huge data points very quickly [10]. The function of clustering is to group image pixels where the related feature vectors produces the similar images. One type of hierarchical clustering is widely applied for image segmentation [7][8][9].

Edge detection plays an important role in preprocessing step particularly, in object classification and recognition systems. Based on gray level properties the relative information about the edges helps to identify the contours. An edge is defined as the boundary between two regions

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

with relatively distinct gray level properties. Information about edges in an image helps to identify the contours of an image and help to retrieve regions enclosed by those contours. An edge image represents a higher level of abstraction (i.e. less information to process) and edges are features invariant to absolute illumination (as opposed to color information) [3][4][6]. Edge detectors are algorithms which perform edge detection. Edge detectors have gained popularity in leaf classification because the classifiers are usually over-fitted to one particularly lighting / intensity condition and edge detection is fairly robust to changing illumination conditions. The result of applying an edge detector to an image is a set of connected curves that indicate the boundaries of surface markings as well curves that correspond to discontinuities in surface orientation. Thus, applying a sobel edge detector to a leaf image significantly reduces the amount of data to be processed and filters out non-relevant information, while preserving the important structural properties of an image. Edge detection produces an edge map that contains important information about the image. If the edge detection step is successful, the subsequent task of interpreting the information contents in the original image may be substantially simplified.

This paper proposes a new hybrid approach for image segmentation that utilizes sobel edge detector and K means clustering. The sobel operator performs the edge detection for plant leaf shape analysis, which should be located as far as possible from each other to withstand against the edge distribution of a plant leaf, as identical to the number of edges amongst the data distribution. It designates the edge positions by calculating the accumulated metric between each data point and all previous line and Sharpe edge corners, and then selects data points which have the maximum value. Here several number of edges are formed which is unable to locate particular leaf shape. Plant image, however, always contain complicated background objects that interfer with the examination process and must be removed from the image prior to species classification. Sobel edge detector is applied at the first level of segmentation to detect the structure of the plant leaf. At the second level k means clustering is used to remove the unwanted segments and false edge detection [14][15]. The performance of the proposed method is compared with other traditional methods to analyze the efficiency of the system in order to improve precision and reduce computation time.

In this paper, Section 2 describes the image segmentation using sobel edge operator and K-means algorithm. Proposed approach will be discussed in Section3. Section4 describes the experimental setup and results by comparing two algorithms, and then followed by conclusion with references in Section 5.

2. Image Segmentation

The main idea of the image segmentation is to group pixels in to identical regions. Features can be represented by the space of colour, texture and gray levels, each exploring similarities between pixels of a region (Kurugollu et al., 2001, Navon et al., 2004, Kim et al., 2002). In Mery (2005) the segmentation is treated as an image division of regions which are not coincident. Among the algorithms normally used to segment a digital image, the K-means algorithm and the edge detection are possible alternatives. The following diagram shows the automatic plant leaf image segmentation system.

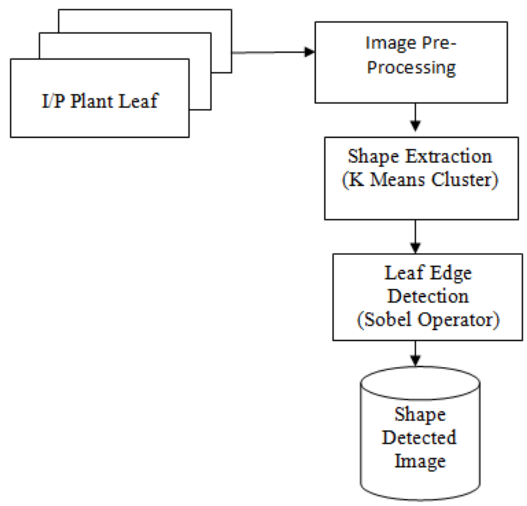

Fig. 1 Architecture for Automatic Plant Image Segmentation

2.1 K Means Clustering

Clustering algorithm has been widely used in image segmentation and database organization. Clustering algorithms can be grouped into two i.e) hierarchical and partitional. Hierarchical clustering algorithms recursively find nested clusters which starts with the data point that merges the most related pairs of cluster data successively Compared to hierarchical clustering, partitional clustering differs by partioning the data which are not imposed of hierarchical structure. The hierarchical algorithm is an n*n similarity matrix, which is derived from the pattern matrix that supports the Multi-Dimensional Scaling (MDS). The main steps are

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

(1) Noise Removal
(2) Data Normalization
(3) K means Cluster

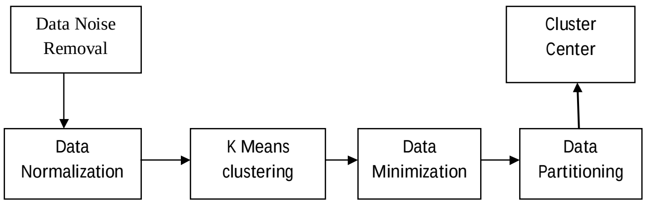

Fig. 2 Modified K means Clustering for Plant leaf segmentation

The above figure 2 shows the modified K means Clustering for Plant leaf segmentation. An adaptive noise removal filtering using the Wiener filter is applied for noise removal of images. The Wiener filter can be considered as one of the most fundamental noise reduction approaches and widely used for solution for image restoration problems [13]. In our system, we use 3x3 neighborhoods of filtering size. So, need to normalize the datasets.

$$f(a_i) = \frac{1}{1+e^{-a_i}} \quad (1)$$

The logistic function presents the needed S-curve with the range of values which are not linear. In order to solve, {a} should be first transformed linearly to vary around the mean *x* in the following way: $a_i^1 = \frac{a_i - \bar{a}}{\lambda(\sigma_x / 2\pi)}$ (2)

where: $\bar{a}$ is the mean value of variable *x*

σ_x is the standard deviation of variable *x*

λ is the linear response measured in standard deviation. It describes in terms of how many normally distributed standard deviations of the variables are to have a linear response. In our case, we set λ=10 in order to make smoother for normalizing the datasets.

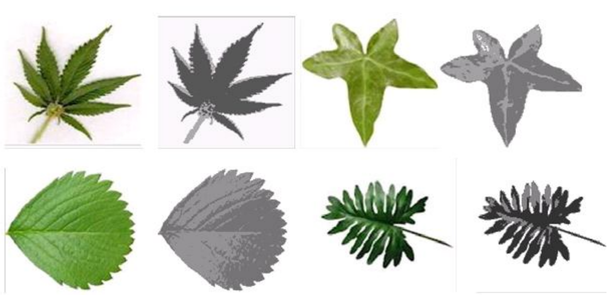

Fig. 3 Plant leaf image segmentation result using modified K means Clustering method.

The hierarchical algorithms are single-link and complete-link in which the K-means independently works with clustering. In the paper, we apply k-means algorithm to analysis images similarities in the database.

Let A={a_i },i=1,…..n be the set of n d-dimensional points to be clustered into a set of k clusters, C={ D_k ,k=1,….k}; k-means algorithm finds a partition such that the squared error between the empirical mean of a cluster and the points in the cluster is minimized. Let M_k be the mean of cluster D_k. The squared error is defined as

$$J(D_k) = \sum_{x_i \in c_k} \left\| a_i - M_k \right\|^2 \quad (3)$$

The goal of K-means is to minimize the sum of the squared error over all K clusters,

$$J(D_k) = \sum_{k-1}^{k} \sum_{x_i \in c_k} \left\| a_i - M_k \right\|^2 \quad (4)$$

Minimizing this objective function is known to be an Non deterministic polynomial time (NP-hard) problem (even for k =2). Thus K-means converges to a local minimum when clusters are well separated (Meila, 2006). Initial partition begins with k clusters and assign to reduce the squared error. Since the squared error always decrease with an increase in the number of clusters k (with J(C)= 0 when k= n), it can be minimized only for a fixed number of clusters. The main steps of k-means algorithm are as follows:

1) Select an initial partition with k clusters; repeat steps (ii) and (iii) until cluster membership stabilizes.
2) A new partition is generated by assigning each pattern to its closest cluster.
3) Compute new clusters.

The objects are classified into K number of classes based on a set of features. The classification process is carried minimizing the sum of squares of distances between the objects. The clustering is partitioned into four clusters. Multiple values of clusters have been tested.

2.2 Sobel Edge Detection

Edge detection is more popular for identifying discontinuities in gray level than detecting isolated points and thin lines .The edge is the boundary between two regions with relatively distinct gray level properties. It is assumed here that the transition between two regions can be properties. The transition between two regions can be determined based on the gray level discontinuities .The Sobel operator performs a 2-D spatial gradient

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

measurement on an image and so emphasizes regions of high spatial frequency that correspond to edges. In the input grayscale image ,approximate gradient magnitude is also identified at each point by the edge detector. The operator consists of a pair of 3x3 convolution kernels which is rotated by 90 degree. The convolution masks of the Sobel detector are given in figure 4

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

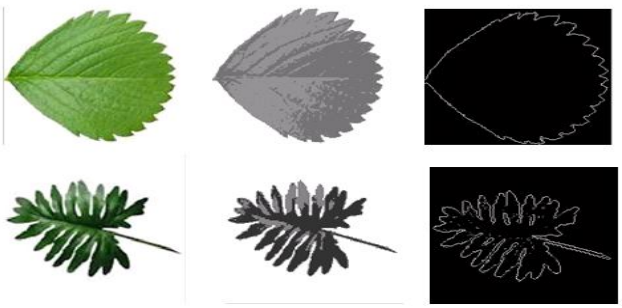

Fig. 4 Sobel Mask and Image Results for Proposed approach

3. Proposed Methodology

In the current study, the proposed automated segmentation technique was tested to detect the shapes of plant leaf images. Edge detection is known as the most popular method to detect the shape of different plant leaf present in the database. Based on sobel edge detected plant leaf images, botanist differentiates plant leaf based on several edge detection technologies. Since the extracted image consist of complicated edge line to detect the exact part of the plant leaf image. Hence k means clustering is used to optimize the cluster center and to detect the exact part of the plant leaf shape. The determination of plant leaf can sometimes be missed in certain situation. Generally, the accuracy of detection test depends on the quality of the plant leaf samples. Thus, the botanist may have difficulty in extracting the shape of the plant leaf due to these problems. Therefore, the current study will apply the proposed automated edge detection technique on combination of sobel and K means algorithm to detect the edges of leaf images. In the current study, the detected plant leaf will have 0 and 255 of grey level value respectively. This will provide clearer segmented plant leaf images to assist botanists for better detection and classification ratio extraction.

4. Experimental Results

The experiments are conducted using matlab 7.1. Ten leaf images are taken as benchmark images. To test the accuracy of the segmentation algorithms, four steps are followed.

i) First, a pre-processed Leaf image is taken as input.
ii) Secondly proposed approach is applied to a leaf image.
iii) Third, the performance evaluation is estimated based on the parameters Energy, Mutual Information, UQI, Entropy and ET.

Normally the value of Energy, Mutual Information, UQI & Entropy must be high which produces good quality image. Whereas the ET value must be minimum to prove the efficiency of the proposed approach.

a.) Energy

The gray level energy is calculated using the equation,

$$E(x) = \sum_{i=1}^{x} p(x) \quad (5)$$

where E(x) represents the gray level energy with 256 bins and p(i) refers to the probability distribution functions, which contains the histogram counts. The larger energy value corresponds to the lower number of gray levels, which means simple. The smaller energy corresponds to the higher number of gray levels, which means complex. The following figure 5 shows the energy value for different segmentation method. The proposed method proves efficient by displaying high energy level. The figure 6 shows the sample dataset.

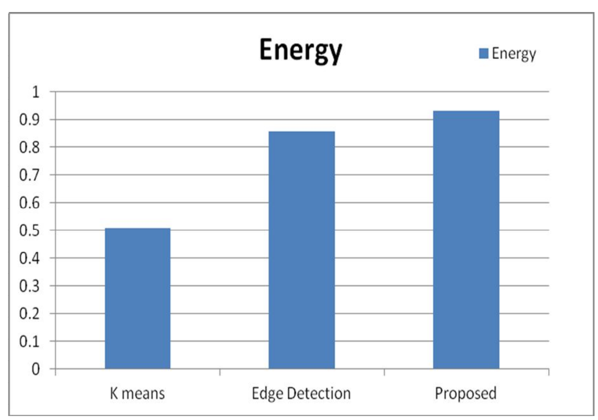

Fig 5 Energy Level for Plant Leaf Image

IJCSI
www.IJCSI.org

Fig. 6 Sample Plant Leaf Image Dataset

b.) UQI

UQI measures image similarity across distortion types. The universal quality index is defined as

$$UQI = \frac{4\sigma_{xy}\overline{xy}}{\left[\sigma_x^2+\sigma_x^2\right]\left[\left(\overline{x}\right)^2+\left(\overline{y}\right)^2\right]} \quad (6)$$

The dynamic range of UQI is [-1,1]. The following figure 7 shows the UQI value for different segmentation method. The proposed method proves efficient by displaying better value.

Table 1 Parametric Evaluation for Segmentation Technique

Methodology	Energy	UQI	Entropy	Mutual Information	Evaluation Time
K means	0.509	0.5354	0.1352	0.67	1.32
Edge Detection	0.8582	0.5034	0.1034	0.52	1.63
Proposed	0.9320	0.7325	0.1524	0.73	1.41

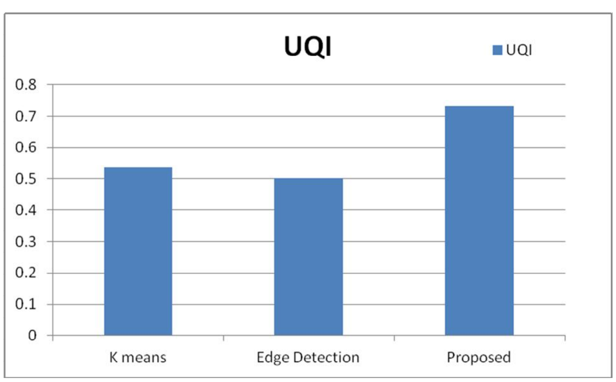

Fig.7 UQI value for Plant Leaf Image

c.)Entropy

Entropy is used to characterize the texture of the input image based on the statistical measure of randomness. This measure quantifies the expected value of the information contained in a image. Considering two discrete probability distributions of the images p and q the entropy is formulated as,

$$d = \sum_{i=1}^{k} p(i)\log_2 \frac{p(i)}{q(i)} \quad (7)$$

The following figure 8 shows the Entropy value for different segmentation method. The proposed method proves efficient by displaying better entropy value.

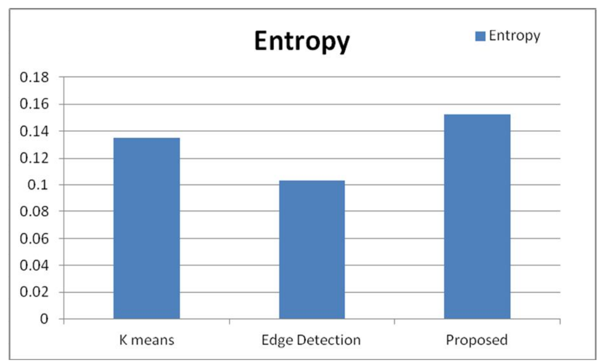

Fig. 8 Entropy Level for Plant Leaf Image

d.) Mutual Information (MI)

MI is the measure used between two random variables that are mutually dependent. The symmetric function is formulated as,

$$I(X,Y)=\sum_{XY}P_{XY}(X,Y)\log_2\frac{P_{xy}(X,Y)}{P_x(X)P_y(Y)}$$
$$=-\sum_{x}P_x(X)\log_2 P(X)+\sum_{x,y}P_{xy}(X,Y)\log_2\frac{P_{xy}(X,Y)}{P_x(X)P_yY)} \quad (8)$$
$$=H(X,)-H(X|Y)$$

where I(X; Y) represents the mutual information; H(X) and H(X|Y) are entropy and conditional entropy values. It is interpreted as the information that Y can tell about X and the measure of reduction in uncertainty of X due to

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.
IJCSI
www.IJCSI.org

the existence of Y. The following figure 9 shows the MI value for different segmentation method.

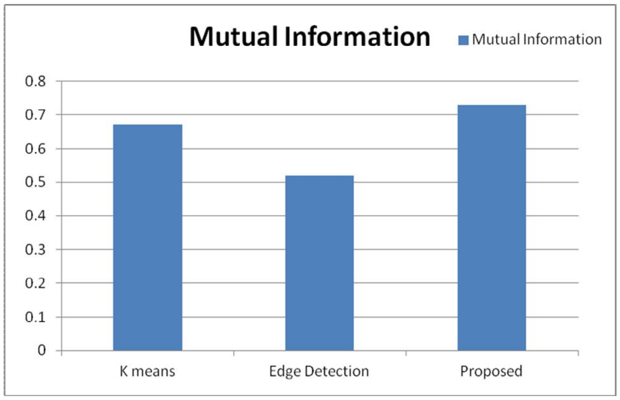

Fig. 9 Energy Level for Plant Leaf Image

e.) Evaluation Time

Evaluation Time (ET) is defined as the time taken by a system to process the operation. The measurement of any event depends on the memory-size, the input data size, and the memory access time. The execution time taken should be low for online and real-time image processing applications. Hence lower ET is better than higher ET value when all other performance-measures are identical.

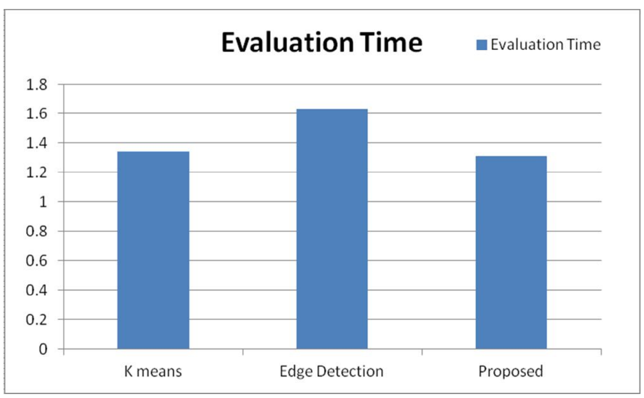

Fig. 10 Elapsed Time for Plant Leaf Image Segmentation

The above figure 10 shows the ET value for different segmentation method. The ET value must be low for a better image. The proposed method proves with efficient results.

5. Conclusions

This paper presents a new approach for plant leaf image segmentation by applying non linear k means algorithm. The segmentation process presents a clustering mechanism for high resolution images in order to improve the precision and processing time. K means clustering is applied at the first level of segmentation to detect the structure of the plant leaf. At the second level Sobel edge detector is used to remove the unwanted segments to extract the exact part of the leaf shape. The performance of the proposed method is compared with other traditional methods. Experimental result shows that this new approach simplifies the process to extract shape related features and measurements of the leaf for higher accuracy.

References

[1] Wan-Ting Lin, Chuen-Horng Lin, Tsung-Ho Wu, and Yung-Kuan Chan, Image Segmentation Using the K-means Algorithm for Texture Features, World Academy of Science, Engineering and Technology, 65, 2010.

[2] Jarosław Gocławski, An Automatic Segmentation Method for Scanned Images of Wheat root Systems with Dark Discolourations, Int. J. Appl. Math. Comput. Sci., Vol. 19, No. 4, pp.679–689, 2009.

[3] Wenchao Cai, Jue Wu, Albert S Chung, Shape-Based Image Segmentation Using Normalized Cuts, Proceedings of IEEE conference ICIP (2006) Vol. 1, ,pp. 1101-1104, 2006.

[4] D E Ilea, P F Whelan, Color Image Segmentation Using a Spatial K-Means Clustering Algorithm, proceeding 10th International Machine Vision and Image Processing Conference (2006) ,Vol.1, 2006.

[5] Nor Ashidi Mat Isa, " Automated Edge Detection Technique for Pap Smear Images Using Moving K-Means Clustering and Modified Seed Based Region Growing Algorithm", International Journal of The Computer, the Internet and Management Vol. 13.No.3 ,pp 45-59, 2005.

[6] Song Wang, Feng Ge, Tiecheng Liu, "Evaluating Edge Detection Through Boundary Detection", Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208, USA, 2006.

[7] Mat-Isa, N. A., Mashor, M. Y. & Othman, N. H. (2002). Pap Smear Image Segmentation Using Modified Moving K-Means Clustering. Proceeding of Kuala Lumpur Int. Conf. on Biomedical Engineering. pp. 44-46, 2002.

[8] Arun Kumar et.al, Leaf Disease Grading by Machine Vision and Fuzzy Logic, Int. J. Comp. Tech. Appl., Vol 2,No.5, pp.1709-1716,2011.

[9] Dheeb Al Bashish, Malik Braik and Sulieman Bani-Ahmad, “Detection and Classification of Leaf Diseases using K-means-based Segmentation and Neural-Networks-based Classification”, Information Technology Journal, Vol.10, No.2, pp. 267-275, 2011.

[10] Rafael C.Gonzalez, Richard E.Woods, Steven L.Eddins, “Digital Image Processing”, *Pearson Education*, 2nd Edition.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[11] B.Ramamurthy, CBMIR: Shape-Based Image Retrieval using Canny Edge Detection and K-means Clustering Algorithms for Medical Images**,** International Journal of Engineering Science and Technology (IJEST), Vol. 3 No. 3 pp.1870-1877, 2011.

[12] N. Senthilkumaran and R. Rajesh, Edge Detection Techniques for Image Segmentation – A Survey of Soft Computing Approaches, International Journal of Recent Trends in Engineering, Vol. 1, No. 2, pp.250-254, 2009.

[13] N. Senthilkumaran and R. Rajesh, "Edge Detection Techniques for Image Segmentation - A Survey", Proceedings of the International Conference on Managing Next Generation Software Applications (MNGSA-08), 2008, pp.749-760.

[14] Hong Liu , and Xiaohong Yu , Application Research of k-means Clustering Algorithm in Image Retrieval System, Proceedings of the Second Symposium International Computer Science and Computational Technology(ISCSCT '09) *, pp. 274-277, 2009.*

[15] H.P. Ng, S.H. Ong, K.W.C. Foong, P.S. Goh, W.L. Nowinski, Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm, Proceeding on IEEE Southwest Symposium on Image Analysis and Interpretation, pp. 61 – 65,2006.

[16] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh, Fast and Accurate Detection and Classification of Plant Diseases, International Journal of Computer Applications , Vol. 17,No.1, pp. 0975 – 8887, 2011.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.