Image Processing In Data Mining By Using Color K-Means Clustering

[1]KRISHNA MOORTHY, [2]KANNEDARI GOPIKRISHNA, [3]RAMBABU M
[1]B Tech Student, [2]B Tech Student, [3] M.Tech, Associate professor
Department of Computer Science and Engineering, Department of Computer Science and Engineering, Department of Computer Science and Engineering
[1]KG Reddy College of Engineering &Technology, Moinabad, RR District,Telangana, [2]KG Reddy College of Engineering &Technology, Moinabad, RR District,Telangana, [3]KG Reddy College of Engineering &Technology, Moinabad, RR District,Telangana.

Abstract: Now-a-days people are interested in using digital images. There is a great need for developing an efficient technique for finding an image. Image mining is a vital technique which is used to mine knowledge straight forwardly from image. Image segmentation is the primary phase in image mining. It has been a difficult task in computer vision. The role of image segmentation is to decompose an image into parts that are meaningful with respect to the particular applications. The goal of data mining process is to extract information from a data set and transform it into an understandable structure for further use. The pixel discovery is critical issue for any specialists. In this work a strategy to recognize pixels by utilizing picture preparing methods with incorporates picture pre-handling , picture division, extraction, morphological operations and grouping system is used. Abnormal pixels are identified in filtered image by using marker controlled watershed segmentation and gradient magnitude. This present work proposes a technique to identify the pixel point adequately from the filter pictures. It will minimize blunder made by the human exposed eye. MATLAB is broadly utilized programming for the investigation of data mining identification from filtered pictures. We utilized MATLAB 2017b for building up this anticipate.

Keywords: **picture identification, Gaussian mixture, watershed transform, morphological functions, L*a*b color based, k-means clustering, Thresholding, MATLAB 2017b.**

1.INTRODUCTION:

The visual appearance of an image may be significantly improved by emphasizing its frequency contrast of an image is determined by its dynamic range which is defined as the ratio between the both the brightest and dark pixel intensities. Image mining is a technique which handles the mining of information , image data association, or additional patterns not unambiguously stored in the images. it utilizes methods from computer vision, image processing, image retrieval, data mining. The main intention of the image mining is to produce all considerable patterns. In this case we lose the detail information in either light or dark areas when we take a photo in the scences with dark shadows and bright light areas for this obviously we use some enhancement methods. Enhancement methods are divided into two types spatial domain method and frequency domain method. For this method we first pre-process the image, it reduces the noisy in an image. Next, image segmentation is the most important issues on automatic image analysis and pattern recognition according to some features of the image or similar criteria of feature set. An image is divided into series of meaningful and different region. The of image segmentation can be divided into edge based detection and region based. Finally using the K-Mean clustering. This paper presents a survey in the next section on various image mining techniques that were proposed earlier. Also, this paper provides a marginal overview for future research and improvements.

Problem:

Image processing can be broadly defined as the manipulation of signals which are inherently multidimensional. The goal of processing or manipulation can be compression for storage in transmission, image enhancement. Segmentation, analysis recognition and understanding or visualization for human observers. In this paper the goal is image enhancement, segmentation and recognition of data.

Existing System:

The existing system is the image segmentation. In this structure is the Image segmentation not given the extract pixel size. Correct pixel based segmentation gives the correct pixel size, by using the morphologies operations. It is easy to identify the pixel in data mining.

Proposed System:

In this problem, I have converted the image in to four ways, enhancement to gradient magnitude, this values are converted into morphological operation. From this applying the color based segmentation. Opponent can easily recognize the given data using K - Means clustering finally display the picture clarity.

2. LITERATURE SURVEY:

Information and communication technologies (ICT'S) have penetrated all areas of contemporary life. In this context, digital literacy has become much more than the ability to handle computers. Digital image processing [4]is an important role because it target all areas of contemporary existences one of the key aspects of our knowledge is the ability to locate, identity, retrieve, process and digital information Numerous researches have been carried on this image mining. [3].This section of the paper presents a survey on various image mining techniques that were proposed earlier. Development in area of image acquisition and storage technique have shown the way for incredible growth in extensively large and detailed image databases. Even though the growth of several applications and techniques in the individual research. It uses common pattern identical, pattern identification and data mining models with the intention that a real life scene/image.

can be associated to a particular category, assisting in different prediction and forecasting mechanisms. Image mining approach using clustering and color based segmentation. [2]

The images which are available in these data base examined can provide valuable information to the human users. Image mining facilitates the extract of hidden information. Decision tree based image processing and image mining technique was projected by Kun-che et al. (2009). The purpose of the mining is to produce all considerable patterns without prior knowledge of the patterns. The fundamental thought of this work is to identify the pixels and choose the malignant or not. It additionally finds the pixel malignancy stage and gives more precise result by utilizing diverse upgrade and division systems.

3. IMAGE SEGMENTATION:

Picture division is a crucial procedure for most picture investigation consequent assignments. It is a process of partitioning a digital image into multiple segments. It is typically used to locate objects and boundaries in images. The result of the image segmentation is a set of segments that collectively cover the entire image. Each of the pixels in a region are similar with respect to same characteristic or computed property, such as color, intensity and texture. The consequences of picture division is an arrangement of sections that cover the whole pictures, or an arrangement of forms separated from the picture. Division calculations depend upon the one of two essential properties of power qualities intermittence and closeness. Content based image retrieval is the application of the computer version technique searching for digital images in large data bases. Next, stages classified into the edge detection, Thresholding approach, Binarization.

Implementation:

K-means clustering is a method of vector quantization , originally from signal processing that is popular for cluster analysis. K means clustering aim to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. The main idea is to define the k- centroids, one for each cluster, one for each cluster. The next step is to take each point belonging to a given data set and associate it to the nearest [8] centroids. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroids. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step.. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function, where is a chosen distance measure between a data point and the cluster centre c_j , is an indicator of the distance of the n data points from their respective cluster centers. The algorithm is composed of the following steps:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

$\left\| x_i^{(j)} - c_j \right\|^2$ $x_i^{(j)}$

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

Although it can be proved that the procedure will always terminate, the k-means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. The algorithm is also significantly sensitive to the initial randomly selected cluster centers. The k-means algorithm can be run multiple times to reduce this effect.

Standard Algorithm:

The most widely recognized calculation utilizes an iterative refinement method because of its omnipresence it is regularly called the k-implies calculation; let the X={X1,X2……,Xn} be the set of data points and V ={V1,V2,V3,…Vc} be the set of centers.

1) Randomly select ‘c’ cluster centers.
2) Calculate the distance between each data point and cluster centers.
3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.
4) Recalculate the new cluster center using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where, ‘c_i’ represents the number of data points in i^{th} cluster.

5) Recalculate the distance between each data point and new obtained cluster centers.
6) If no data point was reassigned then stop, otherwise repeat from step 3

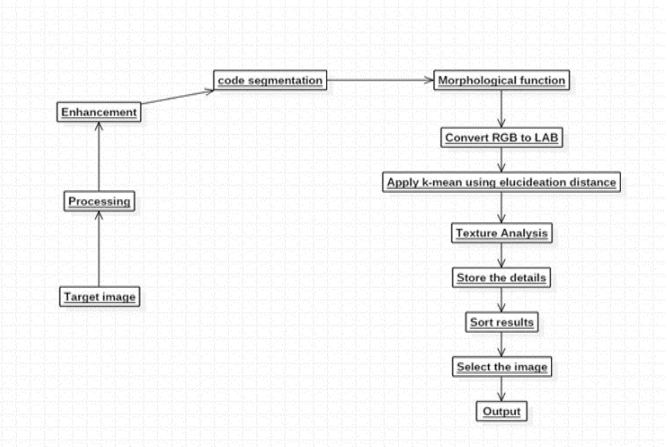
The calculation is regularly displayed as appointing items to the closest group by separation. The standard calculation goes for the minimizing the target and along these line done out the “minimum aggregate of squares”, which is precisely proportionate to dealing out by the Euclidean separation.

In Our Work:

As a result of the computational simplicity of the k-implies calculations over other grouping calculations it was chosen to utilize the k-mean bunching in the proposed work. The k-mean bunching calculation is an exceptional instance of summed of the summed up hard grouping.

calculations. It is connected when point delegates are utilized and the squared Euclidean between vectors xi and bunch agents. The k-implies calculation is given beneath.

Algorithm:



Thresholding Approach:

Thresholding is a process of converting a grayscale input to a bi-level image by using an optimal threshold. The purpose of Thresholding is to extract those pixels from some image which represent an object, Through the information is binary the pixels represent a range of intensities. Thus the objective of Binarization is to marks pixels that belong to true foreground regions with a single intensity and background regions with different intensities. This Threshold value each pixels set to eighter 0 or 1. i.e. Background or foreground. Thus here the change of image takes place only once. The pixels are divided into 2 classes,C_1 with gray levels[1,….,t] and C_2 with gray level [t+1,….L].Otsu's strategy depends on limit choice by measurable criteria.Ots in suggested minimizing the weighted sum of with-in the class variance of the object.

Binarization Approach:

Binarization approach relies on the way that the quantity of dark pixels is much more prominent than white pixels in typical pictures , so we began to check the dark pixels for normal and abnormal pictures to get an average that can be used later as a threshold , if the quantity of the dark pixels of greater than the threshold value, then it shows that the image is normal , generally, if the quantity of the dark pixels is less than the threshold value it demonstrates picture is abnormal. The limit esteem is utilized as a part of this examination is 17178.48 and the true acknowledgement rate(TAR)is(92.86%) and false acknowledgement rate(FAR)is(7.14%).

4. PROPOSED SYSTEM:

In this available image are passed through the system which is having following stages: pre-processing stage, segmentation stage, color based segmentation, morphological operations, k-means clustering. The Gaussian channel is utilized to smooth the info picture in the preprocessing stage. And in addition Thresholding, morphological operations and watershed is used to change the division reason. After picture division, the components are normal power, border, zone and unpredictability are removed from the picture. Binarization approach is used to choose the picture is normal or blurred.

A. Input image:

The filtered picture which are utilized for preparing are gathered from the all pixel quantity. This picture dataset contains filter pictures with normal and abnormal. thefigure 5 demonstrates a portion of the filter picture with normal and abnormal.

Figure5: filtered image (a) white pixel (b) black pixel

GAUSSIAN MIXTURE:

Image enhancement is the technique to improve the image quality so that output image is better than the original image.Themainproposeofimageenhancementistobringoutthedetailsofhiddenimagesortoincreasecontra stina low contrast image.The Gaussian filter works by using the 2D distribution as a point-spreader function. This is achieved by convolving the 2D Gaussian distribution function with the image. We need to produce a discrete approximation to the Gaussian function. This theoretically requires an infinitely large convolution kernel, as the Gaussian distribution is non-zero everywhere. Fortunately the distribution has approached very close to zero at about three standard deviations from the mean. 99% of the distribution falls within 3 standard deviations. This means

we can normally limit the kernel size to contain only values.

where x is the distance from the origin in the horizontal axis, y is the distance from the origin in the vertical axis, and σ is the standard deviation of the Gaussian distribution.

B. Image segmentation:

Division is utilized to isolate a picture into various little locales or items. It has a numerous applications in various fields for the division of the 2D restorative pictures. It is a vital procedure for the most picture investigation taking after methods. There are different techniques accessible for picture division. In this paper Thresholding and marker controlled division strategies are utilized. Thresholding is the best instrument for the picture division reason. It is utilized to change over the dim scale picture in to a paired picture.

C. Gradient magnitude:

Image gradient is a directional change in the intensity or color in an image. The gradient of the image is one of the fundamental building blocks. the gradient tells us the direction in which the image is changing most rapidly. To illustrate this, think of an image as like a terrain, in which at each point we are given a height, rather than intensity. For any point in the terrain, the direction of the gradient would be the direction uphill. The magnitude of the gradient would tell show rapidly our height increases when we take a very small step uphill. Because the gradient has a direction and a magnitude, it is natural to encode this information in a vector. The length of this vector provides the magnitude of the gradient, while its direction gives the gradient direction. Because the gradient may be different at every location, we represent it with a different vector at every image locate

D. Watershed Transform:

Considers the gradient magnitude of an image as a topographic surface. Pixels having the highest gradient magnitude intensities (GMIs) correspond to watershed lines, which represent the region boundaries. Water placed on any pixel enclosed by a common watershed line flows downhill to a common local intensity minimum (LIM). Pixels draining to a common minimum form a catch basin, which represents a segment. Over segmentation is the problem of watersheds. The transformation makes a number of regions as an output. This problem comes mostly from the noise and quantization error. To eliminate the effect of local minima from noise or quantization error in the final results, first the gradient of the original image is computed as a preprocessing and then watershed transformation is applied.Marker controlled watershed based segmentation:

To beat the downsides of this watershed division i.e,over division marker based watershed division system is utilized. It can fragment limits from a picture. Separating touching objects in image is one of more difficult image processing operations. The watershed transform is often applied to this problem. The watershed transform finds catchment basins CB and watershed ridge lines in an image by treating it as a surface where light pixels are high and dark pixels are low. [14, 15] The segmentation using the watershed transform works better if we can identify, or mark, the foreground objects and the background.

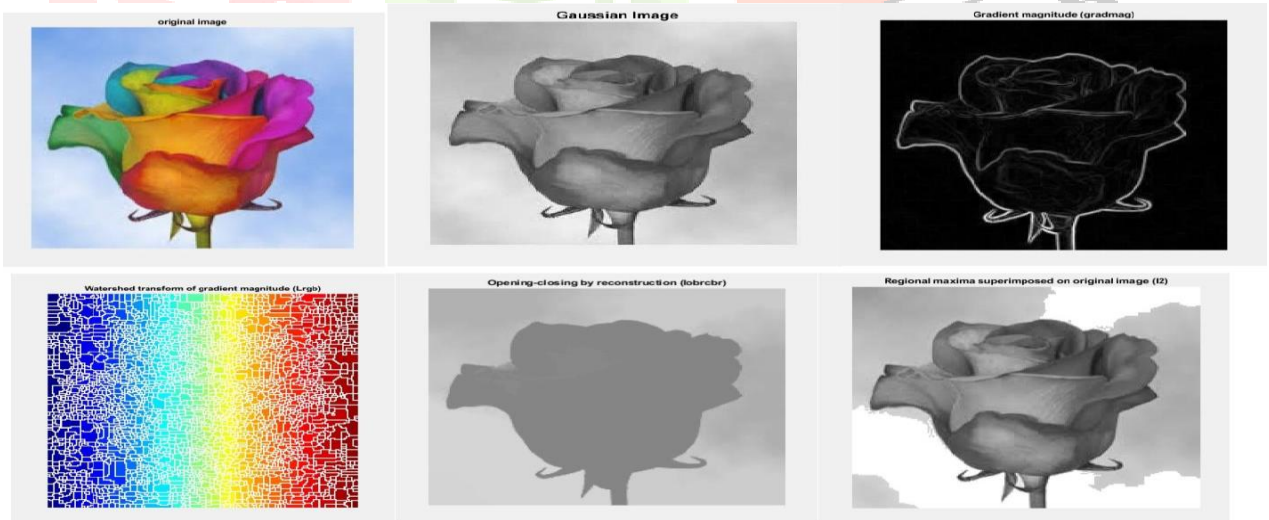
Algorithm:

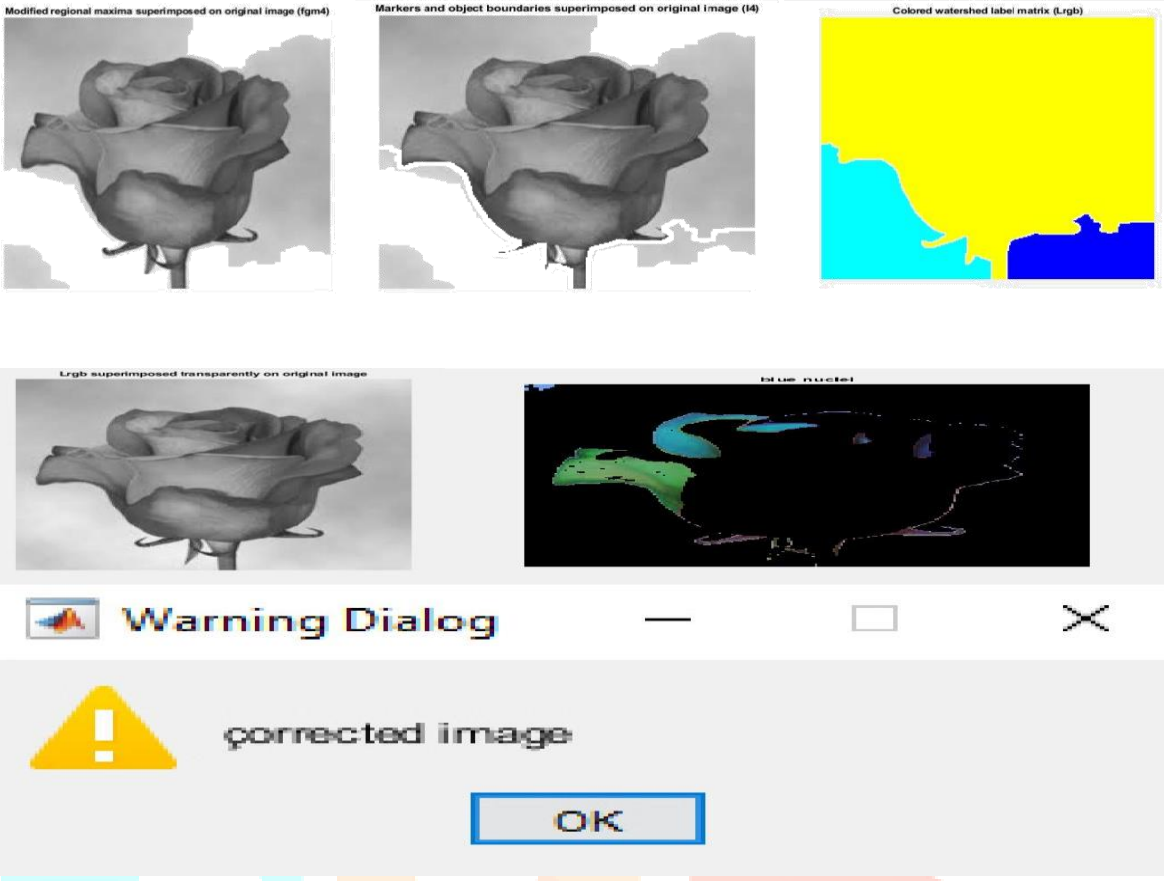
Step 1: Read Image

Step 2: Convert Image from RGB Color Space to L*a*b* Color Space Step 3: Classify the Colors in 'a*b*' Space Using gradient magnitude Step 4: Label Every Pixel in the Image Using the Results from KMEANS Step 5: Segment the Nuclei into a Separate Image

The cluster that contains the blue objects. Extract the brightness values of the pixels in this cluster and threshold them with a global threshold using imbinarize. After k-means clustering, Binarization procedure is finished. In this approach, the aggregate number of dark pixels and white pixels are checked. In the event that the aggregate number of dark pixels of info picture is more than limit, then it displays the pixels are normal or abnormal.

Outputs:





Conclusion:

From observation we know that the Thresholding technique gives us an output of two segments. But, in K-Means technique we can get various segmented according to the clustering size, It increases the cluster size and more accuracy. In feature this technique is to be applied for better performance in the voice recognition of moving objects. Final result we have segmented the image at the some time improve the quality.

References:

1.International Journal of Emerging Science and Engineering (IJESE) September 2014 " A Novel Approach Towards Clustering Based Image Segmentation" Dibya Jyoti Bora, Anil Kumar Gupta.

2.International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014) "Color Based Image Segmentation Using Data Mining Functionalities" I .Nishi Sharma, II Mr. Anurag Joshi
I .CSE, Northern India Engineering College, Shastri Park, Delhi, India II .CSE, BIT Mesra, Noida Campus, Noida.

3.(IJIREEICE) INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING , 2015 " Image Retrieval Using Data Mining and Image Processing Techniques" Preeti Chouhan1 , Mukesh Tiwari2 M. Tech Research Scholar, Digital Electronics, LNCT, Jabalpur, India1 Assistant Professor, Electronics & Communication Engineering, LNCT, Jabalpur, India2 Scott E umbaugh 1998, Computer vision and image processing, prenitic Hall, New Jersey of

5. Application of k-Means Clustering algorithm for prediction of Students' Academic Performance Oyelade, O. J Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria. Ola2000faith@yahoo.co.uk. Oladipupo, O. O Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria. frajooye@yahoo.com. Obagbuwa, I. C Department of Computer Science Lagos State University, Lagos, Nigeria. ibidunobagbuwa@yahoo.com

6.A.K.Jain and R.C.Dubes Algorithms for clustering data.pretentice Hall 1988. C.Solomon and T.Breckon,'Fundamentals Digital image processing,Iefrn 2011 IJCA. Digital Image processing 3rd edition (DIP/3e) by Genzalez and woods.

7.Ms.chinki chandhok,Mrs.soni chaturvedi,"An Approach to image segmentation using K-Means clustering approach algorithm" International journal of information technology(IJIT) August 2012