International Journal of Current Engineering and Technology
ISSN 2277 - 4106

Available at http://inpressco.com/category/ijcet

Research Article

Automatic Grayscale Classification using Histogram Clustering for Active Contour Models

M.Ramesh Kanthan[a] and S.Naga Nandini Sujatha[b*]

[a]Akzo Nobel, Singapore
[b]K.L.N.College of Engg ,Tamil Nadu, India

Accepted 4 March 2013, Available online 1June 2013, **Vol.3, No.2 (June 2013)**

Abstract

The problems of image segmentation using active contours are the minimization of energy criterion, involving both edge and region functional. Automatic initialization of level set function in geometric active contour model makes the process fast and convergence towards the object boundary in minimal iterations. A new technique of automatic region classification using clustering method is proposed in this paper. The approximate region count can be estimated using histogram peak finding method. The elbow statistical method is used to find the number of region using mean and standard deviation measures. The knee point in the elbow gives the accurate number of cluster. Automatic seed points of the clusters are created by Euclidian distance measure. A promising retrieval performance is achieved especially in particular examples.

Key words : *elbow algorithm, segmentation, mean and standard deviation measures, active contours.*

1. Introduction

Image segmentation is an important preprocessing step in most of the image analysis like pattern classification, pattern matching and object identification. Active contour snake model is an efficient technique for retrieving image boundary information. Originally snake(Chenyang Xu et al,1998), balloon (W.N.Lie et al,2001), geodesic active contours are driven towards the edge of an image, through the minimization of boundary integral of features depending on edges. A multi-scale approach is needed for further simplification of image by removing more details such as texture noise while keeping major edges.

Region based methods mainly depends on the assumption of uniformity and homogeneity of luminance within the boundary of object segments. Then region merging and splitting techniques are needed for classification. The main drawback of aforementioned method is that they produce over segmentation or under segmentation which is not used for object based image retrieval. Histogram method of region classification is also one of the best methods of region classification for bimodal images. If the given image is a low contrast image then the histogram peak finding method process will not be given any accurate classification.

The proposed method is a Seeded Region Growing (SRG) method which partitions the image into regions where each connected region corresponds to one of the seeds. It is a quiet promising method provided initial seeds are sufficiently accurate. The main limitation of this method is how to select the number of seeds and appropriate seed value. This work proposes the solution for finding the number of seeds using elbow statistics, and the seed value by distance measure.

The following sections are organized as follows. Section II deals with the review of Elbow algorithm , image clustering fundamentals with k means algorithm and histogram method . In section III we present a proposed method of efficient and accurate segmentation procedure of number of cluster calculation based on histogram peak finding process and statistical measures. Experiments to test the performance of the proposed method, is presented in section IV. Finally in section V we give results and its performance analysis.

2. Back ground

2.1 Elbow algorithm

Determining the number of clusters in an image, (Qinpei Zhao et al,2008) a quantity often labeled *k*, is fundamental to the problem of image clustering and is a distinct issue from the process of actually solving the clustering problem. In most cases, *k* must be chosen somehow and specified as an input parameter to clustering algorithms, with the exception of methods such as correlation clustering, which are able to determine the optimal number of clusters during the course of the

M.Ramesh Kanthan is Asst. IT Systems Manager and
*__Dr.S.Naga Nandini Sujatha__ is Prof.

algorithm. The correct choice of *k* is often ambiguous,(Jan Puzkha) (Bashar Al-Shboul et al,2009)with interpretations depending on the shape and scale of the distribution of points in a data set and the desired clustering resolution of the user. If an appropriate value of *k* is not apparent from prior knowledge of the properties of the image data, it must be chosen automatically. There are several categories of methods for making this decision.

In statistics, explained variation or explained randomness measures the proportion to which a mathematical model accounts for the variation (= apparent randomness) of a given data set. Often, variation is quantified as variance; then, the more specific term explained variance can be used. The complementary part of the total variation/randomness/variance is called unexplained or residual.

The knee of a curve is loosely defined as the point of maximum curvature(Stan Salvador et al,2004). The knee in a # of clusters vs. evaluation metric graph can be used to determine the number of clusters to return. Various methods to find the knee of a curve are:

1. The largest magnitude difference between two points.
2. The largest ratio difference between two points
3. The first data point with a second derivative above some threshold value.
4. The data point with the largest second derivative.
5. The point on the curve that is furthest from a line fitted to the entire curve.

In our algorithm we use magnitude difference statistics to draw a L curve and knee point finding process.

2.2 Image clustering and K means algorithm

An automatic clustering process is considered an unsupervised learning process(Ted Pedersen et al,2006), because it can automatically reveals, intrinsic categorical patterns. Different similarity measure can result in different cluster results. Clustering classification falls in non-hierarchical and hierarchical clustering algorithms. The difference is hierarchical generates multiple level categorical structure but not hierarchical have one level structure. K Means is a non-hierarchical cluster algorithm. The algorithm starts with guess about the cluster centroid, and based on the simple iterative scheme and finds the local optimum.

$$J=\sum \sum \| x_i^{(j)} - c_j \|^2 \quad (1)$$

The function defines the shortest distance difference between c_j (centroid) and the item x_i.

(1) Choose random k points and set as cluster centers.
(2) Assign each object to the closest centroid's cluster.
(3) When all objects have been assigned, recalculate the positions of the centroids.
(4) go back to Steps 2 unless the centroids are not Changing.

One popular way to start K-Means (Ming Luo et al) is to randomly choose k points from data. Initial starting points are important in the clustering process; however, the results mainly depend on the initial means. The standard solution is to try a number of different starting points. Moreover, the results also depend on the metric used to measure distance which is not always easy to implement especially in the high-dimensional space.

3. Proposed Method

The process block diagram of the proposed system is given in fig 4. The input image has classified based on the cluster analysis for N times depends on the nature of image. The N value is calculated from peak finding process. The seed point values of each clustering types have stored for mean variance calculation. From the number of clusters and the mean variance the elbow graph has been drawn. The knee point has been found from the graph lowest value.

3.1. Histogram peak finding process

One of the clusters stopping measure can be derived from histogram peaks. The histogram peak finding process can be used the basic simple logic of Low-High-Low points in the chart. Since the graph of histogram is not having smooth curve, the logic can be extended with local maximum. Each peak can be checked with other nodes to minimum level (approximately 10) to find its local maximum. The number of peaks output from this process N is used as the limit for knee finding curve.

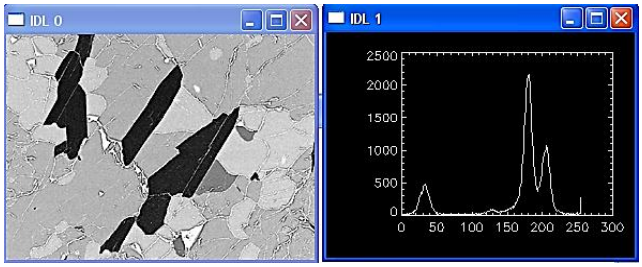

Figure 1 Histogram of sample image 1

The figure 1 shows the sample image having three different intensity and the corresponding histogram having three distinct peaks. But if it is low contrast image shown in figure 2 it is difficult to access the distinct peaks and also it should not give accurate classification. So from this method the approximate peak value N is calculated using local maxima algorithm.

3.2 Seed value finding process

This process starts with Random initiation of the cluster centroids. The process is repeated with grouping of pixels using distance measure until equilibrium obtained. The seed finding process is repeated for n times starts with the

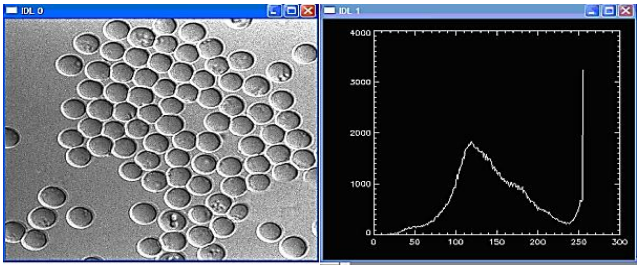

Figure 2 histogram of sample 2

initial k value as 2. The n value is calculated based on approximate peak finding process as a initialization step. Based on the initial cluster centroids the resultant Cluster Center value is calculated based on K-Means algorithm regional clustering. Using Euclidian distance measure the correct CCV is finalized. The resultant seed values ie, cluster center values (CCV) are stored in the array S for further processing.

$$CCV(k) = \sum_{j=0-i}^{i=2-n} S(i, j) \quad (2)$$

Where I = 2 to N and j= 0 to I
Here N is approximated with the number of peak value depends on the nature of the image. The sample images resultant values are shown in table 1.

3.3 Mean difference of cluster center

From the result of the CCV the mean difference at each cluster variables is calculated by the formula of image statistical measures

$$MDC(k) = \frac{1}{I-1} \sum S(i, j+1) - S(i, j) \quad (3)$$

From the mean difference which gives the small deviation is calculated and the corresponding cluster gives the number of cluster values. The knee graph has plotted with number of regions in X axis and the Mean Difference Cluster (MDC) of every regional classification in Y axis. The knee point shows the actual number of cluster in the image. The calculated seed value and its mean difference are discussed in the resultant section.

3.4 Finding the Knee of a Curve.

Histogram-based methods are very efficient when compared to other image segmentation methods because they typically require only one pass through the pixels. In this technique [4], a histogram is computed from all of the pixels in the image, and the peaks and valleys in the histogram are used to locate the clusters in the image. Color or intensity can be used as the measure.

From the graph plot of the data which is obtained from section b and c, the lower knee value can be find out. For this knee finding process the plot point differences are calculated. The difference points are stored in an arrays along with subscript of the number of cluster. If there is a

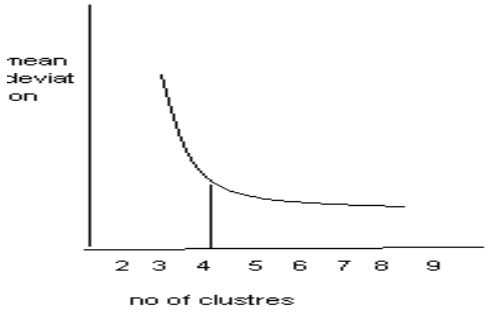

Figure 3 Finding Knee Value

long deviation compared with other deviations then that point is calculated as knee point. This process is explained with detailed sample data in section IV.

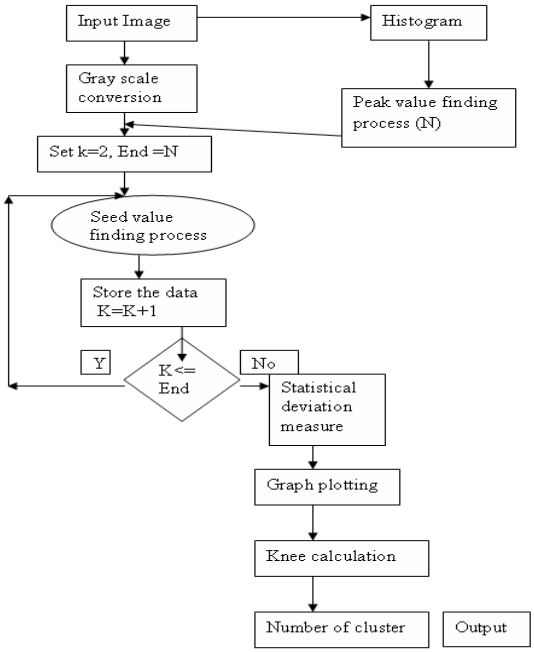

Figure 4 Proposed System Flow Chart

4. Implementation and results

The following table 1 shows the results obtained from the equation 2. Based on the result the mean differences of cluster center has find out and plot the graph between the number of cluster and the mean cluster center value. From the knee point the actual number of cluster has find out automatically.

Table 1 -Cluster center value table of Crane image

No of cluster	2	3	4	5
Cluster center values	68	63	50	48
	211	130	76	73
		222	143	117
			223	176
				227
Mean Avg difference	143	79.5	57.6	44.75

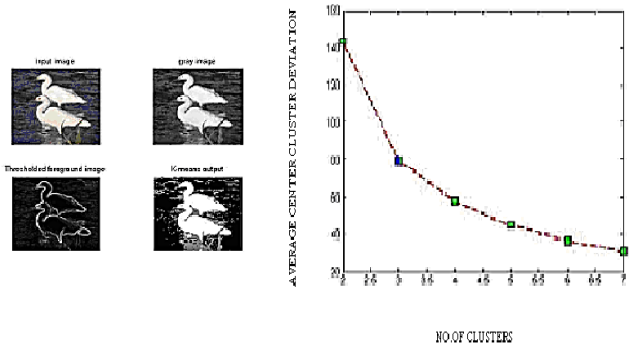

Figure 5. Output of the cluster regions No of cluster:3 (Example 1)

Table 2 -Cluster center value table of Sample image

No of cluster	2	3	4	5	6	7
Cluster center values	30	26	15	14	14	11
	208	150	62	53	52	35
		226	169	118	110	61
			232	180	164	113
				234	204	165
					239	205
						240
Mean Avg difference	178	100	72	55	45	38

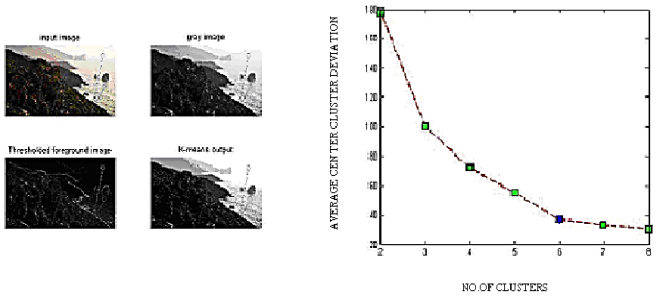

Figure 6. Output of the cluster regions No of cluster:6 (Example -2)

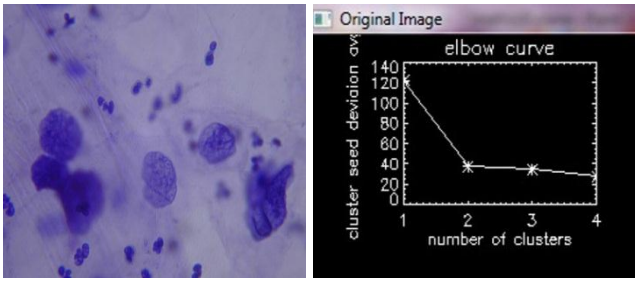

Figure 7 Sample hsil image and its knee curve

Table 3 -Cluster center value table of Hsilimage

Image Input	Seed Value					
	No of Regions					
Hsil	1	2	3	4	5	6
	127	93	90	79	71	0
	-	134	122	98	88	22
	-	-	137	125	103	44
	-	-	-	138	125	66
	-	-	-	-	138	89
	-	-	-	-	-	111
Mean Avg. Difference	0	20	15	14	13	0--(2)

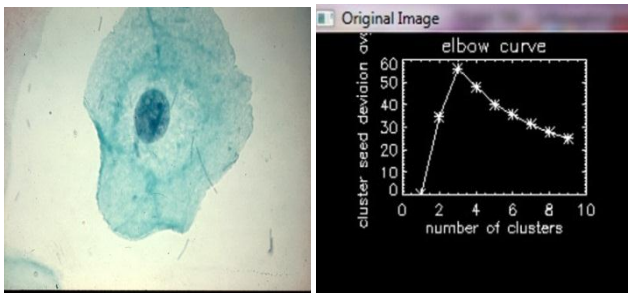

Figure 8 Sample cyto image and its knee curve

Image Input	Seed Value					
	No of Regions					
Cyto	1	2	3	4	5	6
	181	144	50	34	27	15
	-	213	157	133	122	92
	-	-	218	176	157	133
	-	-	-	255	186	164
	-	-	-	-	227	191
	-	-	-	-	-	228
Mean Avg. Difference	0	34	56	47	40	35--(3)

This method is used as a preprocessing for active contour model with level set function. Mainly aims for automatic cluster value finding for automatic region grid assignment of level set function using knee finding algorithm. The experimental results along with the graph plot value are shown in the figure 7 & 8 for sample medical images (cervical cancer cell images). From the mean difference average the number of cluster has been find out by low-high-low (or) high-low-high principle from the elbow plot. The seed center values for each classifications of cluster and its corresponding mean average difference values are listed in the table 1 to table 4. The elbow curve is plotted with number of clusters and its mean average cluster centers.

5. Conclusion

This paper presents region based approach using magnitude values to extract the object boundary from its background ie, object segmentation. Then numbers of clusters are automatically calculated using elbow statistics. The proposed technique only requires the information of input image in gray scale form, no other assumption to be considered. This is an adaptive method without human intervention the processing is carried out. For final segmentation the result need localization process. This can be used as a pre-processing step for level set grid formulation in geometric active contour models. This process can be extended with the combination of both results for multiple object detection or overlap region detection.

6. References

Qinpei Zhao,Ville Hautamaki, Pasi Fränti(Oct 2008), Knee Point Detection in BIC for Detecting the Number of Clusters, *Springer*

W.N.Lie and C.H.Chuang (2001), A fast and accurate snake model for object contour detection, *Election, Lett.*, Vol. 37. no.10, pp.624-626.

Chenyang Xu and Jerry L. Prince (March 1998), Snakes, Shapes, and Gradient Vector Flow, *IEEE Transactions on image processing*, vol. 7, no. 3.

ZHOU Xian Cheng SHEN (2008), New two dimensional Fuzzy C means clustering algorithm for image segmentation, *Qun-tai,J.Cent Sonth Univ Technology*, 15:882-887 Springer.

Stan Salvador, Philip Chan (2004), Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms, ictai, pp.576-584, 16th *IEEE International Conference on Tools with Artificial Intelligence* (ICTAI'04).

A Spatial Constrained K means Approach to Image Segmentation,Ming Luo ,Yu fei Ma,Hong.Jiang Zhang.

Histogram clustering for Unsupervised Image Segmentation,Jan Puzkha Thomas Hofm Hofmanin Joachim M Buhmam.

Bashar Al-Shboul, and Sung-Hyon Myaeng (2009), Initializing K-Means using Genetic Algorithms, *World Academy of Science, Engineering and Technology* 54.

Tapas Kanungo , David M. Mount , et. al.(July 2002), An Efficient k-Means Clustering Algorithm:Analysis and Implementation, *IEEE transaction on pattern analysis and Machine Intelligence* Vol 24, No 7.

Ted Pedersen and Anagha Kulkarni (June 2006) , Automatic Cluster Stopping with Criterion Functions and the Gap Statistic, *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume*, pages 276–279,New York City.