Volume 2, No. 11, November 2011
Journal of Global Research in Computer Science
RESEARCH PAPER
Available Online at www.jgrcs.info
ISSN-2229-371X

SEMI SUPERVISED IMAGE SEGMENTATION USING OPTIMAL HIERARCHICAL CLUSTERING BY SELECTING INTERESTED REGION AS PRIOR INFORMATION

L.Sankari *[1] and Dr.C.Chandrasekar[2]
[1*]Assistant Professor, Dept of computer science, Sri Ramakrishna CAS for Women, Coimbatore, India
sankarivnm@gmail.com
[2]Associate Professor, Department of computer science, Periyar University, Salem, India
ccsekar@gmail.com

Abstract: Image segmentation is to be the first step in image analysis, pattern recognition and feature extraction. It is a critical but primary component of image analysis since it only determines the quality of the final result of image analysis. This paper discuss about the semi supervised image segmentation using hierarchical clustering algorithm. The prior information for the clustering process is given as an interested area selection from image using mouse. Here intensity, color and texture of the image properties are considered. The proposed idea gives more clarity of segmented regions than the existing methods using the other semi supervised method.

Keywords: semi supervised clustering- PSO algorithm-image segmentation

INTRODUCTION

The intent of image segmentation is to cluster the pixels into relevant region and these regions may be surfaces, object or part of the image[6][7]. There are many methods available to perform this segmentation. Clustering is a common step to get the segmentation. In the data space, clusters are regarded as regions of similar data points. This work discuss about semi supervised clustering for image segmentation using prior information.

Semi supervised clustering:

Image segmentation using semi-supervised clustering [2] is a direct way to combine the prior information during the process [8] of clustering. For example, In Figure 1 the three clusters A, B, C are formed with some known label (prior information). The prior information may be given either by mouse clicks or by constraints.

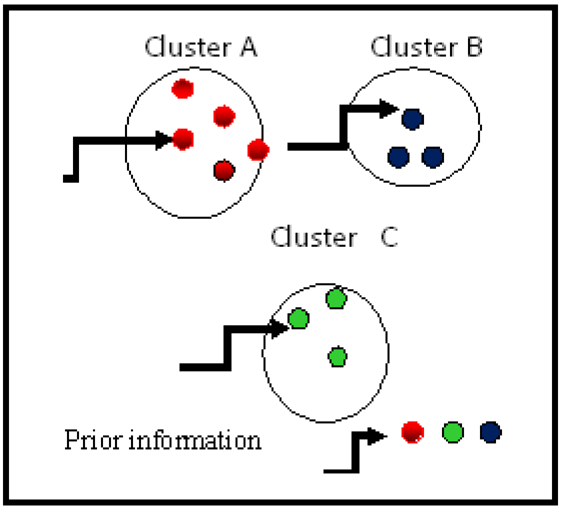


Figure.1. Semi supervised model

In this paper we present a segmentation algorithm base on semi-supervised clustering, which integrates limited human assistance. Instead of mouse clicks according to the paper [1], the user selects some window area by mouse. The selected object inside the window will be segmented and displayed. The proposed idea takes less time runs fast and is very convenient for a user to present the prior segmentation information[9][10]. Actually for many realistic applications this kind of limited human assistance can be obtained by unsupervised means [13]. For instance, in order to identify objects in a certain typical environment, the color distribution of an image background where objects locate can be known before segmentation, which can be regarded as the prior information in our proposed algorithm.

The region to be segmented is taken as prior information [11] and the Hierarchical clustering algorithm is applied with intensity, color and texture properties of the image[12]. Here the statistical Geometrical features of texture are used to extract the texture values. Among these, the optimal set of texture values are obtained using PSO (Particle Swarm Optimization algorithm). These values are indicated by yellow color markings as in the figure

PARTICLE SWARM OPTIMIZATION

Theory of particle swarm optimization (PSO) has been growing rapidly. PSO [3] [4] [5] has been used by many applications of several problems. The algorithm of PSO emulates from behavior of animals societies that don't have any leader in their group or swarm, such as bird flocking and fish schooling. Typically, a flock of animals that have no leaders will find food by random, follow one of the members of the group that has the closest position with a food source (potential solution). The flocks achieve their best condition simultaneously through communication among members who already have a better situation. Animal which has a better condition will inform it to its flocks and the others will move simultaneously to that place. This would happen repeatedly until the best conditions or a food source discovered. The process of PSO algorithm in finding optimal values follows the work of this animal society. Particle swarm optimization consists of a swarm of particles, where particle represent a potential solution.

The fuzzy membership function (gbellmf) is used as a fitness function in PSO. The algorithm of PSO is initialized with a particles obtained from SGF texture feature and then searches for optima by updating generations. The function is taken as a built in function from matlab.

$$f(x;a,b,c) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \quad \text{-----} \quad (1)$$

where the parameter b is usually positive. The parameter locates centre of the curve. The SGF values are given as the input values of X. Because quantitative evaluation functions deal with segmented images as a set of regions, the target image is divided into a set of regions and not to a set of classes during the different stages of our method (where a region is a group of connected pixels having the same range of gray levels).

Each particle is flown through the search space having its position adjusted based on its distance from its own personal best position and the distance from the best particle of the swarm.

Each particle, i, flies through an n–dimensional search space, Rn, and maintains the following information:

a. xi, the current position of ith particle (x - vector),
b. pi, the personal best position of ith particle (p - vector), and
c. vi, the current velocity of ith particle i (v - vector).

The personal best position associated with a particle, i, is the best position that the particle has visited so far. If *f* denotes the fitness function, then the personal best of i at a time step t is updated as:

$$p_i(t+1) = \begin{cases} p_i(t) & if \quad f(x_i(t+1)) \geq f(p_i(t+1)) \\ x_i(t+1) & if \quad f(x_i(t+1)) \leq f(p_i(t)) \end{cases} \quad (2)$$

If the position of the global best particle is denoted by *gbest*, then:

gbest ∈ { p1(t), p2 (t), . . . , pm(t) } = min{ f (p1(t)), f (p2(t)) , . . . , f (pm(t)) }   (2)

The velocity updates are calculated as a linear combination of position and velocity vectors. Thus, the velocity of particle i is updated using equation (3) and the position of particle i is updated using equation (4).

vi(t+1) = w* vi (t) + c1 r1(pi(t) – xi(t)) + c2 r2 (*gbest* –xi(t))   (3)

xi(t+1) = xi(t) + vi(t+1)   (4)

where w is the inertia weight, c1 and c2 are the acceleration constants, r1 and r2 are random numbers in the range [0,1] and Vi must be in the range [-Vmax, Vmax], where Vmax is the maximum velocity.

MERGING PROCESS

After finding the global best value using PSO the regions are merged using merging process. The merging is performed with the global best values. The adjusted regions with global best values are merged one by one at each step. After the global merging process the boundary of the objects in the image is obtained. But these boundaries are not refined boundaries. So boundary refinement technique is used for smoothing the boundaries.

BOUNDARY REFINEMENT

Boundary refinement is finally performed to smooth the jagged boundaries after merging. If an image pixel is on the boundary of at least two distinct regions, a discrete disk with radius 3 will be placed on it. The similarity between the two regions is calculated individually for smoothing the jagged boundaries.

Each particle keeps track of its coordinates in the problem space which are associated with the best solution (fitness) it has achieved so far. The fitness value is also stored. This value is called pbest. Another "best" value that is tracked by the global version of the particle swarm optimizer is the overall best value, and its location, obtained so far by any particle in the population. This location is called gbest.

The velocity of the particles is updated every time whenever the particle moves from one position to another. The velocity uses weight, current pixel position and global pixel position for calculating the velocity of the particle. The result is used for the next iteration to find the gbest particle.

ALGORITHM STEPS

STEP 1: Initialize a population (array) or particle with random position and velocities on dimensions on the problem space.

STEP 2: For each particle evaluate the desired optimization fitness function in d variables.

STEP 3: Compares particles fitness evaluation with particles pbest. If current value is better than the pbest, then set pbest value equal to the current location in d-dimensional space.

STEP 4: Compare fitness evaluation with the population's overall previous best. If current value is better than gbest then reset gbest to the current particle's array index and value.

STEP 5: Change the velocity and position of the particle according to equation(1) and (2), respectively:

$V_{id} = V_{id}$ +c1*rand()*(p_{id} x_{id})+c2*rand()*(p_{gd} - x_{id})   (5)

$X_{id} = x_{id} + v_{id}$ (6)

Where c1 and c2 are constants.

STEP 6: Loop to step (2) until a criterion is met, usually a sufficiently good fitness or maximum number of iterations.

The acceleration constants c1 and c2 in equation (1) represents the weighting of the stochastic acceleration terms that pull each particle towards pbest and gbest positions. Particles velocities on each dimension are clamped to a maximum velocity Vmax. Vmax is an important parameter. It determines the resolution, of fitness, with which regions between the present position and the target positions are searched.

Semi Supervised Hierarchical Image Segmentation:

In this paper the image is read first and the prior information for clustering is given by selecting the specific region. Then the clustering process starts. In order to perform clustering, the given image is converted to L*a*b color space and normalize the colors. It is converted to Related Works Jun Tang [1] proposed a color image segmentation algorithm based on region growing. In the field of image processing, image segmentation is a common topic. Also it is a more concentrated and most focused in the field of image processing techniques. Color image segmentation methods are more concerned by the researchers due to the growth of computer processing facilities and the increased application of color image. Color image segmentation techniques are an extension of the gray image segmentation technique, but most of the gray image segmentation techniques can not be applied directly to color images. This necessitates the improvement in the technique of real gray image segmentation technique according to the color image. This research introduces a color image segmentation technique of automatic seed region growing on basis of the region with the integration of the watershed algorithm with seed region growing algorithm which based on the conventional seed region growing algorithm.

Juyong Zhang *et al.*, [2] put forth a diffusion approach to seeded image segmentation. Seeded image segmentation is a most admired type of supervised image segmentation in computer visualization and image processing. The images were treated as a weighted graph in the earlier methods of seeded image segmentation technique and reduce an energy function on the graph to build segmentation. In this research, the author proposed to carry out the seeded image segmentation based on the result of a heat diffusion method in which the seeded pixels are regarded as the heat sources and the heat diffuses on the image beginning from the sources. Once the stable state is attained after the diffusion, the image is segmented according to the pixel temperatures.

Random Walk algorithm is comprised in this implemented framework for image segmentation which diffuses only along the two coordinate axes. For the purpose of controlling the diffusion this paper incorporates the features of the image into the diffusion method, providing an anisotropic diffusion technique for image segmentation. The experimentation result shows that the implemented anisotropic diffusion process yields better segmentation results. When this technique is tested by using the ground truth dataset of Microsoft Research Cambridge (MSRC), achieves an error rate of 4.42%, which is very lesser than the other modern algorithms.

A novel image segmentation method based on random walk is given by Yi-hua Lan *et al.*, [3]. To reduce the difficulties, image segmentation depended on the random walk model in graph hypothesis can be altered into large-scale sparse linear equations. The absolute result of the equation and the iteration convergence rate is according to the selection of the initial value. If the initial value is chosen randomly to segment the large scale image then it ends in a significant disadvantage. In this research, the author proposed a novel image segmentation technique depended on random walk model. Initially the author down-samples the original large image to the small image, then the small image segmentation guides to sparse linear equations in very small scale. Once the solution is obtained, the possible results will be up-sampling to the up layer and then the sparse linear equations in this layer are solved. On repetition of this up-sampling process until to the top layer where the original image is obtained. The final probability image is segmented at the end with a pre-set threshold. The algorithm is tested by taking two images and the segmentation results are compared with the original random walk algorithm. The segmentation result ensures that this method is far better from other techniques. This algorithm works by taking the low-scale image probability output as the initial value of the high-scale image segmentation process. The author concluded that under the same computation time the segmentation result by our algorithm is much better than that by the original random walk segmentation algorithm.

Ye Hou *et al.*, [4] suggested image segmentation based on GC-CV. This is an integrated method which was formulated and applied to image segmentation. The proposed technique combines the advantages of both the graph cut method and the simplified Mumford-Shah model (C-V model). The proposed approach is tested under its three different operational modes. The first mode directly segments the binary images using GC-CV. The second mode uses recursive GC-CV to segment the multi-region images. In the last mode, both the color images and gray images are segmented with the combination of GC-CV technique and EM algorithm and also uses the YCbCr color space for the segmentation of color image. With the use of many serious experiments, the feasibility and effectiveness of the proposed GC-CV method is verified. From the experimental results it is concluded that the proposed GC-CV method significantly improves the speed of segmentation and considerably reduces the number of iterations when compared with C-V model.

PROPOSED IDEA

The image pixels are grouped based on the intensity, color and texture.

Algorithm:

Step 1: Read the image.

Step 2: Select the image region with mouse.

Step 3: Reduce the colors using L*a*b

Step 4: Normalize the color features

Step 5: Calculate texture feature by Algorithm 2 and obtain Mean, standard deviation and average .

Step 6: Execute algorithm 3 with the initial fitness function provided the values for a=2, b=4, c=6.

Step 7: Obtain global best points. Compare next pixels. If s(i.j) –s(i+1,j+1)>=50 then assign those pixel values as ‘0’ and change the color of those pixel values as ‘Yellow’ to represent global best points as in the diagram (c).

Step 8: Merge the cells .

Algorithm 2:

(To get the Statistical Geometrical features)

Step 1: Convert the whole image based on the threshold value so that the set of all connected ‘1’ valued pixels and ‘o’ values are obtained using the following equation.

$$Im(X,Y,T) = \begin{cases} 1, & Im(X,Y) \geq T \\ 0, & otherwise \end{cases} \quad \text{- - (7)}$$

Where T lies between 1 and max gray level(threshold value)
Step 3: For every region find irregularity using below equation.

$$Irr = 1 + \sqrt{\pi}\max_{i \in I}\sqrt{(xi - \overline{x})^2 + (yi - \bar{y})^2} \quad \text{-----(8)}$$

Step 4: The variation of the pixel irregularity from 0 to 1 Represents change in shape. That means connected region changes from circle to Line.
Step 5: find average irregularity for 1 valued pixels of the binary image. Find average, simple mean and standard deviation.

RESULTS AND DISCUSSION

This section gives the overview of the conducted experiment and presents the obtained result to evaluate the performance of the semi supervised hierarchical algorithm with the human segmented images. The performance metrics in this experiment is based on comparing the pixel between the resulting semi-supervised image and the human segmented image. And the RAND INDEX is also used for comparing the result between the proposed algorithm and the human segmented image. Clearly, the proposed algorithm gives remarkable improvement in natural image segmentation better than the human perception. The proposed semi-supervised algorithm efficiently reducing the time required for performing SGF and hierarchical segmentation.

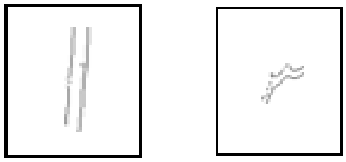
The experimental result shows that the proposed semi-supervised hierarchical image segmentation algorithm performs the segmentation of natural image and reducing time during merging process comparing to the existing image segmentation algorithm.

First the segmentation conducted on the natural images with the size of 300×300. Firstly the hierarchical segmentation is performed to show the variation between the images. Then the user can select the required region of the image for segmenting. Then the Hierarchical fuzzy based PSO is executed on the selected region of the image. Finally the result shows that the proposed algorithm works effectively on the user selected region.

Figure.2.Result of Segmentation
(a) Original image (b) optimal texture Portion

c) selected region

d) segmented portion of the image c.



e 1. e 2.
e. some other selected portion of the image c.

Performance chart.
Performance of the proposed system is compared using Pixel comparison measurement.

CONCLUSION

This paper discussed about semi supervised image segmentation using optimal hierarchical clustering algorithm. The results shows that any part of the image can be segmented easily and the results are compared with the method not using PSO algorithm. The proposed idea gives better results.

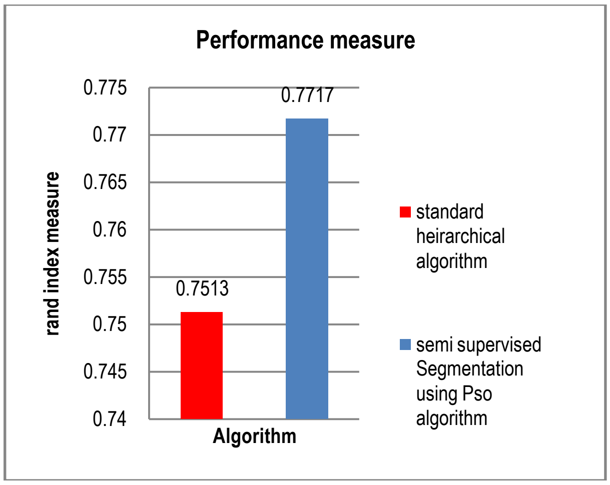


Figure 3. Performance chart

CONCLUSION AND FUTURE ENHANCEMENTS

This paper shows the performance of semi supervised image segmentation using optimal hierarchical clustering procedure for a selected region. The results are compared with standard hierarchical clustering procedure and the results shows better performance than standard one. So this will be useful for feature extraction, object detection, medical image segmentation etc.

REFERENCES

[1]. Yuntao Qian, Wenwu Si "A Semi-supervised Color Image Segmentation Method" IEEE proceedinangs, vol:5, 2005

[2]. Matthieu Guillaumin, Jakob Verbeek et al., "Multimodal semi-supervised learning for image classification", 2009.

[3]. Russell C.Ebethart, Yuhui shi, "particle Swarm Optimization: Development, applications and Resources" proceeding of IEEE, vol:01, 2001

[4]. Kaiping Wei1, Tao Zhang et al., "An Improved Threshold Selection Algorithm Based on Particle Swarm Optimization for Image Segmentation" proceeding of IEEE, vol:07, 2007.

[5]. SHI Zhen-gang, "A New Image Segmentation Algorithm Based on Fuzzy Logical" proceeding of IEEE, vol:10, 2010.

[6]. Leo grady,"Random Walks for Image segementations", IEEE Transactions on pattern analysis and Machine Intelligence, Vol 28, No 11, Nov 2006

[7]. Wladys law Skarbek, Andreas Koschan October 1994," Colour Image Segmentation - A surve

[8]. Nizar Grira, Michel Crucianu, Nozha Boujemaa," Unsupervised and Semi-supervised Clustering:a Brief Survey " ,October 26, 2004

[9]. A.K. Jain, M.N. Murty, P.J. Flynn. Dataclustering: A review. ACM Computing Survey, Vol.13, pp.264- 323, 1999.

[10]. R.O. Duda, P.E. Hart, D.G. Stork. Pattern Classification, 2001.

[11]. K. Wagstaff. Intelligent clustering with instance-level constraints. PhD thesis, Cornell University, 2002

[12]. D. Cohn, R. Caruana, A. McCallum. Semi-supervised clustering with user feedback. Technical Report TR2003-1892, Cornell University, 2003.

[13]. S. Basu. Semi-supervised Clustering: Probabilistic models, algorithms and experiments. PhD Thesis, Department of Science, the University of Texas at Austin, 2005.