A Survey On Comparison of Fast Nonparametric clustering algorithm

Chaitali Choudhury[1]
Mtech in DECC
School of Computing and Information Technology
REVA University
Bangalore, INDIA.
Chaitalichoudhury3@gmail.com.

Nirmala Guptha[2]
Associate Professor
School of Computing and Information Technology
REVA University
Bangalore, INDIA
nirmalaguptha@reva.edu.in

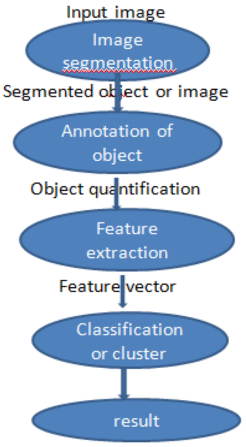
***Abstract*— The Gaussian Blurring Mean Shift (GBMS) is a clustering algorithm, a procedure that iteratively sharpens a dataset my moving each data point according to Gaussian Mean Shift (GMS). (1) Here a criterion has been used, that stops the procedure as soon as the clusters are formed which shows reliable image segmentation. (2) The results being compared with GMS which shows that GBMS (connecting spectral clustering) produces better result. (3) Further GBMS is accelerated by interleaving connection component and blurring step.**

Keywords—GBMS, GMS, spectral clustering

INTRODUCTION

Image Segmentation is the process of subdividing a digital image into different segments. The target of segmentation is to simplify the delineation of the image into something that is easier to analyse and gives more meaningful results. It is typically used to detect objects and boundaries in the image. More precisely , image segmentation is the procedure where each pixel is labelled in an image such that pixels which share same label have certain characteristics.

The output of image segmentation is a group of segments that collectively covers the full image , or a group of contours extracted from the image. Every pixel in a particular region are same based on some characteristic or the property computed such as color,intensity,or texture.



Segmentation can be achieved by the following methods:

- Edge Based
- Region Based
- Feature Based Clustering
- Model Clustering

I. ALGORITHM USED

A. Gaussian Mean Shift

Mean shift is a non-parametric feature space analysis technique for finding the maxima of a density function. It is also called as mode seeking algorithm. It operates on given discrete data sample of a function to find maximum of a density function. However, GMS algorithm proved tobe slow , since its time complexity is $O(kN^2)$, where N is the number of pixels and k is the average number of iteration per pixel.

Strengths:

- It is suitable for real data analysis as it is application independent.
- No predefined shape is assumed .It is based on current pixel position with certain characteristics.
- Able to handle arbitrary feature space.
- It depends on choice of bandwidth which has some physical meaning

Weakness:

- The selection of bandwidth size is not trivial.
- Inappropriate bandwidth may result in modes merging, or may generate shallow modes.
- Often needs adaptive bandwidth size.

B. Gaussian Blurring Mean Shift

GBMS uses the technique where the moment one cluster collapses , it is been replaced by a single point with a weighted proportional to the cluster's number of points. This gets effected when the cluster breakout at different speeds, which happens when the clusters have different sizes. In accelerated GBMS, it takes same number of iteration but each iteration uses dataset with lesser points and hence is faster.

Accelerated GBMS reduces steps by replacing coincident points with a single point. It finds the coincident point by considering points that are closer to min-diff, where min-diff takes same value as in the final connected components step.

Thus, min-diff is the main solution of the method .GBMS applies it after iteration stopped completely but accelerated GBMS applies it at each and every iteration. Thus this algorithm is a sequence of alternating connected component reduction and involves blurring mean shift steps.

LITERATURE REVIEW

A. Literature Survey

[4] The estimation of the gradient of a density function, with application in pattern recognition:

The main idea of this paper is to use the result of previous non-parametric density estimation. By doing so a general class of kernel density gradient estimate was obtained. The conditions on kernel function lead to a derivation that guaranteed asymptotic unbiasedness and also guaranteed consistency of the estimates.

Moreover the paper shows the application of gradient estimates and how it can be applied to pattern recognition. Gradient clustering algorithm uses two technique by which it can cluster different points. One way of clustering is to assign a point to the nearest mode along the direction of the gradient. To get this, one could make small step at every observation made in the direction of the gradient and the process continues iteratively. The other approach is to shift each point according to some weighted proportion. One reason to use this algorithm is that Gaussian density function helps to condense to single points at every iteration. Also by using this normalized gradient convergence can be shown as mean square. Also we can estimate directly using the mean shift estimate.

[6]Mean Shift, Mode seeking, and clustering:

Mean shift is been generalized and analyzed in this paper. The generalized mean shift make use of some k-mean like clustering algorithms. The generalized mean shift in this paper first defines the kernel with its notation and operations. Then this paper also illustrates the functioning using generalized sample mean and mean shift algorithm.

B. Literature Review

This paper illustrates that GBMS uses the technique where the moment one cluster collapses , it is been replaced by a single point with a weighted proportional to the cluster's number of points. This gets effected when the cluster breakout at different speeds, which happens when the clusters have different sizes. In accelerated GBMS, it takes same number of iteration but each iteration uses dataset with lesser points and hence is faster.

Accelerated GBMS reduces steps by replacing coincident points with a single point. It finds the coincident point by considering points that are closer to min-diff, where min-diff takes same value as in the final connected components step. Thus, min-diff is the main solution of the method .GBMS applies it after iteration stopped completely but accelerated GBMS applies it at each and every iteration. Thus this algorithm is a sequence of alternating connected component reduction and involves blurring mean shift steps.

II. OBJECTIVES

Mean Shift is a clustering algorithm that produces a good image segmentation . It operates by defining a gaussian kernel dentisty estimate for the data and iteratively clusters together the points with same characteristics . Further GMS is accelerated to get better result which also proves tobe a faster method.

The general goal is to provide an efficient method of Image segmentation which clusters the image iteratively and computes the time taken by specific algorithms. The input image need to be greyscale image and can be of different resolution. Our objective is to get the algorithm result on basis of certain parameters (such as clusters, time etc).

The objectives of this research work are:

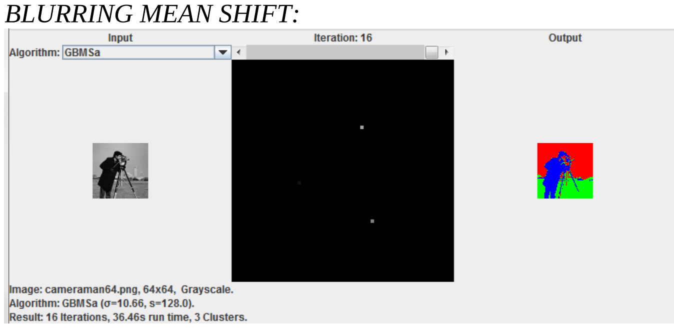
- To implement the fast non parametric clustering with the Gaussian blurring mean shift.
- To improve the drawback of GMS algorithm and comparing it with GBMS which gives much better results.
- Comparing and displaying result of the algorithm.

III. RESULTS

The experimental result has been shown which shows that accelerated GBMS is much faster algorithm.

OUTPUT OF GBMS ALGORITHM:

OUTPUT OF ACCELERATED GAUSSIAN BLURRING MEAN SHIFT:



Due to the nonparametric nature, both GMS and GBMS are particularly well suited algorithm for image segmentation. The member of cluster can be controlled by user through the single parameter, the bandwidth σ, which contains pixel units. Also no post processing is needed for the resultant cluster. Although accelerated GBMS is much faster compared to other nonparametric methods still clustering large datasets computationally costs more as the time complexity of GBMS is $O(kN^2)$. The cost is almost entirely due to first 4-6 iterations, therefore further acceleration try to reduce the cost per iteration rather than iterations. The experiment proves that GBMSa is 5-60 times faster than GBMS.

TABLE SHOWING COMPARISON BETWEEN GBMS AND ACCELERATED GBMS ON BASIS OF TIME:

OUTPUT COMPARISON:

IMAGE	GBMS	GBMSa
Cameraman64.png	153.48s	46.05s
Cameraman128.png	3226.61s	608.45s
Cameraman256.png	55188s	9198.98s

IV. CONCLUSION

The proposed algorithm, i.e accelerated GBMS proves to give better image segmentation with result been produced much faster. Moreover post processing is not need though one may use post processing to remove small clusters, which may further reduce the computional time. The above comparison of output also shows that accelerated GBMS is 5-60 times faster than GBMS.

References:

[1] Fast nonparametric clustering algorithm with Gaussian blurring mean shift .Dept of Computer Science and Engineering,Ogi school of Science and Engineering, Oregon Health and Science university,20000NW walker Road ,Seaverton , USA.

[2] K. Elissa, “Title of paper if known,” unpublished. Carreira-Perpin˜a´n, M. A´ . (2005). Gaussian mean shift is an EM algorithm.

[3] Carreira-Perpin˜a´n, M. A´ . (2006). Acceleration strategies for Gaussian mean-shift image segmentation. *CVPR’06*, to appear.

[4] Fukunaga, K., & Hostetler, L. D. (1975). The estimation of the gradient of a density function, with application in pattern recognition. *IEEE Trans. Inf. Theory*, *21*, 32–40.

[5] Horn, R. A., & Johnson, C. R. (1986). *Matrix analysis*. Cambridge, U.K.: Cambridge University Press.

[6] Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Trans. PAMI*, *17*, 790–799.

[7] C. Yang, R. Duraiswami, N. A. Gumerov, and L. Davis. Improved fast Gauss transform and ef_cient kernel density estimation. In *ICCV*, pages 464.471, 2003.
K. Roeder, “Density estimation with confidence sets exemplified bysuperclusters and voids in the galaxies,’’ ***J. Amer. Srutisticul Assoc.,*** vol. *85,* **pp.** 617-624 .

[9] **Y.** Cheng and K.S. Fu, “Conceptual clustering in knowledge organization,” *IEEE Truns. Puttern Analysis und Muchine Intelligence*, vol. **7,** pp. 592-598