Improved K-Mean Clustering with Steepest Ascent (Gradient) Method for Image Retrieval

Vaishali Ughade
SOIT, UTD, RGPV
BHOPAL
(M.P) India

Nishchol Mishra
Asst. Prof SOIT, UTD, RGPV
BHOPAL
(M.P) India

Dr. Sanjeev Sharma
Head of SOIT, UTD, RGPV
BHOPAL
(M.P) India

ABSTRACT
In this study we present a new frame work for clustering that uses an Improved K-Mean with Steepest ascent (Gradient) Technique. The basic idea of this paper is to use a Color Descriptors which work on RGB and HSV color space after that this result is used by Improved K Mean with Steepest ascent (Gradient) algorithm. In which it used a heuristic local search algorithm that provide additional information about the solution. In this direction it gives the effective result of clustering that provide stability and performs better in global searching.

Keywords
Color Descriptors, Improved K-Mean, RGB and HSV color Space, Steepest Ascent (Gradient) Technique.

1. INTRODUCTION
In image fusion main task is to give purity for image retrieval because over the last few days' documents have been changed from text documents to multimedia documents containing text, images, audio and video and in computing world a large amount of data is search and retrieved so for that Clustering algorithm is a vital field for research area, different clustering algorithm have been represent but there is a problem for clustering purity, algorithm stability and global searching so a method is introduced which used some other heuristic information with the help of it increase the searching probability.

K-means [9] was proposed by Macqueen in 1967. It uses the Heuristic information to make the search more objective in order that the searching efficiency is improved. Its basic idea is that the clustering number K is assigned, firstly creating an initial partition stochastically, then using iteration method to improve the partition through moving the clustering centroid continually until the best partition is obtained [5].

In conventional clustering method centroid value is depending on Cluster value so it's changed according to changing in cluster value so it's a main problem. And it not provide a best centroid value, so to solve this problem an improved K means Algorithm is developed with heuristic local search algorithm. This Provides additional information. This gives the best centroid value and stability in clustering.

In content based image retrieval (CBIR) Visual information plays an important role, this information describe the image feature like color texture and shape which is used for image retrieval. Color is the visual feature most used in CBIR system and also the most explored in literature [3].Color descriptors originating from histogram analysis have played a central role in the development of visual descriptors [7].In CBIR system in last research mainly used different Color spaces for image retrieval like RGB, HSV, YCrCb [7].The HSV and RGB color space is mostly used by researcher in last decade. The RGB and HSV are nonlinear but it is reversible.

This paper Present six sections section 2 describes K Mean Algorithm, section 3 Suggest the concept of Improved K mean With Steepest Ascent (Gradient) Technique, in section 4 represent the Flow chart of proposed method section 5 represent the experiment result and the last one section 6 describe the conclusion and future work.

2. K MEANS ALGORITHM
CLUSTERING problems arise in many different applications, Such as data mining and knowledge discovery[6].In clustering algorithm different approaches is used hierarchical clustering is one of them but it is not efficiently work on large data set so that's way a K mean clustering is used to perform a faster and accurately on large data set.

In K Mean clustering defined the value of k and initializes the K cluster center. K-Means is a popular clustering algorithm which adopts an iterative refinement procedure to determine data partitions and to compute their associated centre of mass, called centroids [1]**,** partition the objects to the nearest clustering centroid to form a cluster according to the Nearest-Neighbor rule[5], then compute the mean value of each cluster and make it the new clustering centroid [8] and in K mean work in iterative manner and used a error rule Function E .the error function is as follows:

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} (p - m_i)^2 \qquad (1)$$

Where *p* denotes the object in space, mi is the centroid of cluster C_i. The smaller *E is*, the more similar within group data is [4]

The time complexity of the k-means algorithm is O(nkl), where n is the number of objects and k is the number of clusters, and l is the number of iterations [2].K Means Clustering is used a datasets which has collection of n objects. Then partitioning a set of data (n) into a set of meaningful sub classes (K) called clusters and where the summation of Error is minimal.

1. Select the value for K.

2. Decide *K* initial clustering centroids randomly;

3 Partition all the objects to the corresponding clusters according to the nearest-Neighbor cluster center;

4. Calculate the sum value of each cluster and re-estimate the cluster centroids.

5. Check the error function *E*, if it is smaller then finish condition exit, otherwise go to3, repeat the process.

6. Output.

3. I-K MEAN WITH STEEPEST ASCENT METHOD

Improved K Mean algorithm is used to improve the cluster accuracy and perform a better global search it used the Heuristic local search algorithm steepest Ascent (Gradient) method to improve the clustering result.

3.1 Improved I-K mean with steepest Ascent Algorithm

Input: Data set includes n objects and selects the value of K.

Output: K cluster which contain best set of objects.

1. Decide the value for K.

2. Stochastically select *K* initial clustering centroids.

3. Decide the class memberships of the *N* objects by assigning them to the nearest cluster center using the nearest neighbor rule.

4. Compute the mean value of each cluster and renew the cluster centroids;

5. Applying steepest ascent Algorithm.

5.1. ***Initial state*** **(q)** = set of renew clusters which has selected n objects.

5.2 For loop until a goal images found and complete loop not change the initial state.

5.3 Generate a new result **R** by adding the most accurate result object in **initial state (q)** and remove the previous objects

5.4 Compare all the set elements in initial state which is nearest to goal.

5.5 Select ***n*** element of **R** which is more minimal value form goal.

5.5.1 if value **R(r)** < value **q(n)** then
Retrieve the **R(r)**, initial state **(q) = r.**

Else if value **R(r)** = value **q (n)** then **initial state (q) = r.**

Update the value of **Initial state (q)**

Else

Exit

4. FLOW CHART OF PROPOSED METHOD

In the proposed methodology I have taken a color image and following computations are carried on it which is divided into four main steps:

1. Take the color image as an input and load the image using MATLAB, then RGB values are converted into HSV color Space and get the Histogram value of Query Image.
2. Similar Process is applied on the Data Base images and indexed the images according to Histogram value.
3. On the bases of Histogram value Euclidean Distance method is applied on Query Image and Data base images. And calculate the similarity with the Query image.

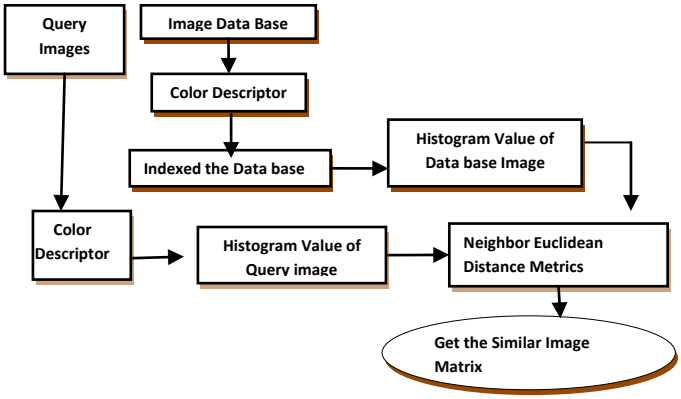

Fig 1: frame work of method

4. 4. Appling the Improved K means with Steepest Ascent Method to sort or indexed the similarity vectors and show images that are similar to the query image

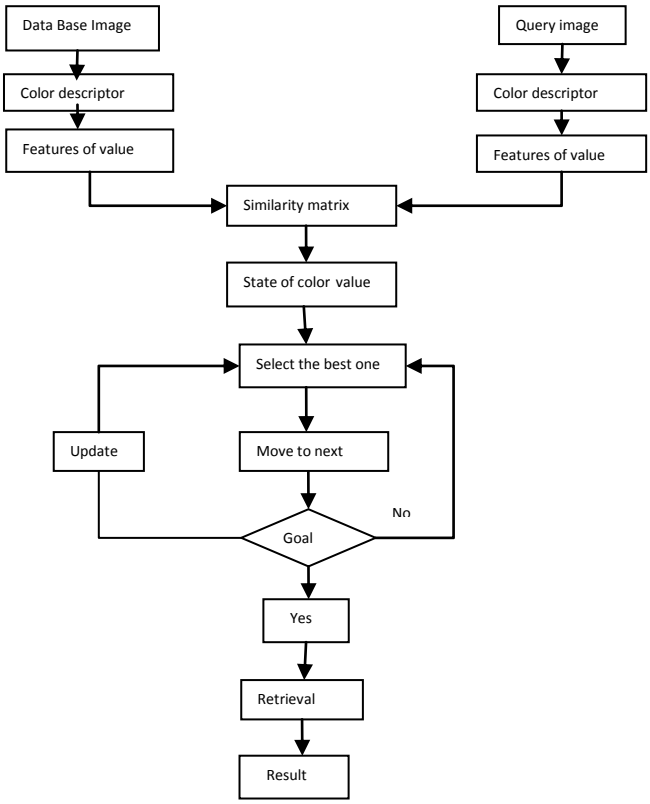

Fig 2: Frame work for I-K mean with Steepest Ascent method

5. EXPERIMENT AND RESULT

To show that our approach is better than the Previous K mean algorithm we perform different experiment on the real data base set in we check the cluster accuracy , running time and error

rate. The experiment was executed on Pentium IV CPU 2.58GHZ machine and software used is MATLAB (2007). So the image retrieval results of using different algorithm on same data base, to represent the usefulness of our algorithm we perform the clustering on Corel dataset. It consist 6 clusters which contain 137 images of horses, roses, architecture, sea, nature and buses

5.1 Clustering Accuracy

In Clustering accuracy we can simply compute the correctly classified objects for each cluster *c* and then summing up the number and then dividing by the dataset cardinality.

$$\text{Cluster Accuracy} = \frac{\text{Number of images correctly retrieve}}{\text{Total number of images in data set}}$$

5.2 Error Rate

Normally error rate is computed by to check the correct clustering information which would not be available and calculating this factor which represent the quality of both algorithms.

5.3 Running Time

In running time factor is computed by checking that how much time it take to generate the result so it is basically computed by wall clock in second and with the help of it check the running time efficiency of both algorithms

Table 1 Comparative Performance Results of both Algorithms

Algorithm	K	Cluster Accurecy	WCT(sec)	(E)
K Mean	6	3.16666	33.40	1.6
I K Mean with Steepest Accent (Gradient)Technique	6	5.33333	28.56	0.8

1. Number of clusters = K
2. Cluster Accuracy = P
3. Wall clock Time In seconds = WCT (sec)
4. Error Rate = E

In fig 3 shows all the factors in comparative form of both algorithms.

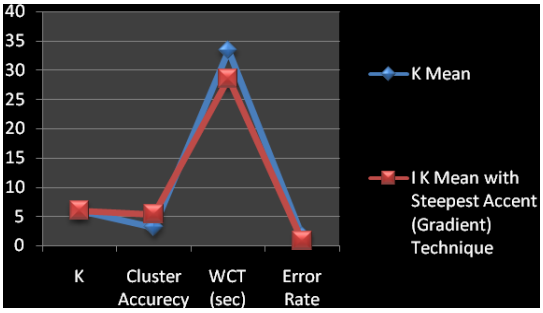

Fig 3: show all these factors in comparative form

6. CONCLUSIONS AND FUTURE WORK

The main objective of this paper; a method for here could improve the effective and efficient image retrieval using improved K means algorithm with steepest ascent (Gradient) method. Firstly we have converted the RGB images into a HSV color spaces and get the histogram value of the query image and database images after that we applying improved K means algorithm with steepest ascent (Gradient) method using the local heuristic search method to improve the cluster purity and give efficient image retrieval in the form of cluster.

This work is based on static image database and animated images used broadly in multimedia database. So in future work it improved the static data base. This work based on Descriptor concept and used a color Descriptor, so in future it used the Different types of Descriptors like Texture and Shape Descriptor. And it also used the combination of all these three Descriptors.

7. REFERENCES

[1] David Pettinger and Giuseppe Di Fatta "Space Partitioning for Scalable K-Means " School of Systems Engineering, University of Reading, Reading, UK- Ninth International Conference on Machine Learning and Applications-2010, pp .319-324

[2] Manning, C.D., Raghavan, P., and Schutze, H., "Scoring, Term Weighting, and the Vector Space Model", Introduction to Information Retrieval, Cambridge University Press,Cambridge, England, 2008, pp. 109-133.

[3] Otavio Augusto, Bizetto Penatti and Ricardo da Silva Torres Otavio" Color descriptors for Web image retrieval" Institute of Computing, University of Campinas – UNICAMP 13084-970, Campinas, SP, Brazil API.2008,pp.163-170.

[4] Zhe Zhang, Junxi Zhang, Huifeng Xue" Improved K-means Clustering Algorithm" (College of Automation, Northwestern Polytechnic University, Xi'an 710072) 2008, pp.169-172

[5] WANG Jiayao, ZHANG Xueping, ZHOU Haiyan."A Genetic K-means Algorithm for Spatial Clustering" Computer Engineering.(3),2006,pp. 188-190

[6] Tapas Kanungo, David M. Mount,Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu "An Efficient k-Means Clustering Algorithm: Analysis and Implementation " IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 7, JULY 2002,pp.881-891

[7] B. S. Manjunath , Jens-Rainer Ohm, Vinod V. Vasudevan, and Akio Yamada "Color and Texture Descriptors" VOL. 11, NO. 6, JUNE 2001,pp.703-715

[8] Mali U, Bandyopadhyay S." Genetic algorithm-based clustering technique. Pattern Recognition" , 33(9): 1455–1465 2000

[9] MacQueen J. Some methods for classification and analysis of multi-variate observations In: Proceedings of the 5th Berkeley Symposium on Mathematics Statistic Problem, 1967,pp- 281~297