Self-Organizing Map-Based Color Image Segmentation with Fuzzy C -Means Clustering and Saliency Map

D.Napoleon[#1], M.SivaSubramanian [#2], M.Praneesh[#2], S.Sathya[#2]

#1 Assistant Professor, Department of Computer Science, Bharathiar University.
#2 Research Scholars, Department of Computer Science, Bharathiar University.

ABSTRACT

Segmentation refers to the process of partitioning a digital image into multiple segments known as super-pixels. Image segmentation is typically used to locate objects and boundaries in images. SOM-K, a new unsupervised natural image segmentation method based on SOM and k-means. Intensity and L*, U*, V* values of a color image are taken as features to be trained by a SOM network. The output prototype vectors are filtered by the hit map at first and clustered by the k-means method. The method is proved to be robust to natural color image segmentation through experiments. In this paper we are comparing SOM-Based Fuzzy C-Means Clustering algorithm to taken different types of Remote Sensing data to test the performance. We compare the quality measures standard deviation, variance and also we can detect the images based on various Edge Detection techniques for efficient segmentation.

Corresponding Author:

1. INTRODUCTION

Data mining is the extraction of the hidden predictive information from large databases; it is the powerful new technology with great potential to analyze important information in the data warehouse. Since there is an increasing desire to use new technology in the domain, there is a need for such mining concepts. A collection of data objects that object is similar to one another and thus can be treated collectively as one group. Similar to Classification, Clustering is the organization of data in classes. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields, including machine learning, Data Mining, pattern recognition, image analysis and bioinformatics. Neural networks have a large appeal to many researchers due to their great closeness to the structure of the brain, a characteristic are not shared by more traditional systems. In an analogy to the brain, an entity made up of interconnected neurons, neural networks are made up of interconnected processing elements called units, which respond in parallel to a set of input signals given to each. The unit is the equivalent of its brain counterpart, the neuron.

A Self-Organizing Map (SOM) or Self-Organizing Feature Map (SOFM) is a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map. Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space. M.L Goncalves et al [1] in agglomerative clustering, the SOM neighborhood relation can be used to constrain the possible merges in the construction of the dendrogram. In addition, knowledge of interpolating units can be utilized both in agglomerative and partition clustering by excluding them from the analysis. If this is used together with the neighborhood constraint in agglomerative clustering, the interpolative units form borders on the map that the construction of the dendrogram must obey. It may even be that the interpolating units completely separate some areas from the rest of the map.

D. M. Ristic et al [2] A Growing number of digital media applications has required the development of tools for the representation, access and retrieval of visual information. Image segmentation is the first and one of most important steps in image processing. The goal of segmentation is to find and isolate desired objects from the image background. Image segmentation produces a corresponding segmentation mask: a gray-scale image in which different gray levels denote different regions identified by the algorithm. The results of image segmentation are important in further analysis, especially in retrieval multimedia data, extraction of region-specific indexing features, in identifying regions of interest or in processing of medical images. Unfortunately, universal solution for extracting objects has not yet been found. Depending on the problem, specific segmentation method is used. Intensity, color, edge, and texture are some of the low-level features that can be used in image segmentation algorithms. *D. E. Ilea et al [3]* Segmentation algorithm consists of several steps: extraction of the feature vectors corresponding to each image pixel, estimation of initial values for region centers using self-organizing map, and final pixel classification using a variant of K-Means algorithm. In our segmentation method, the self-organizing map is used for detecting main image features and generating potential candidates for region centers. Edge detection is widely used in image processing as it is a quick and easy way of extracting most of the important features in an image. Edge detection must be efficient and reliable because the validity, efficiency and possibility of the completion of subsequent processing stages rely on it.

2. Edge Detection Techniques

Different Edge detection techniques are available which are most frequently used for edge detection namely Sobel, Canny.

Sobel Edge Detection Technique

The kernels are designed to respond maximally to edges running vertically and horizontally relative to the pixel grid, one kernel for each of the two perpendicular orientations. The kernels can be applied separately to the input image, to produce separate measurements of the gradient component in each orientation (call these *Gx* and *Gy*). These can then be combined together to find the absolute magnitude of the gradient at each point and the orientation of that gradient. The gradient magnitude is given by:

$$|G|=\sqrt{Gx2 + Gy2}$$

Canny edge detection techniques

The canny edge detector first smoothes the image to eliminate and noise. It then finds the image gradient to highlight regions with high spatial derivatives. The algorithm then tracks along these regions and suppresses any pixel that is not at the maximum (no maximum suppression). The gradient array is now further reduced by hysteresis.

3. METHODOLOGIES

3.1 CLUSTERING

Cluster analysis is one of the major data analysis methods widely used for many practical applications in emerging areas[13].Clustering is the process of finding groups of objects such that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups. A good clustering method will produce high quality clusters with high intra-cluster similarity and low inter-cluster similarity [12]. The quality of a clustering result depends on both the similarity measure used by the method and its implementation and also by its ability to discover some or all of the hidden patterns [8].

3.2 K-MEANS BASED ALGORITHM

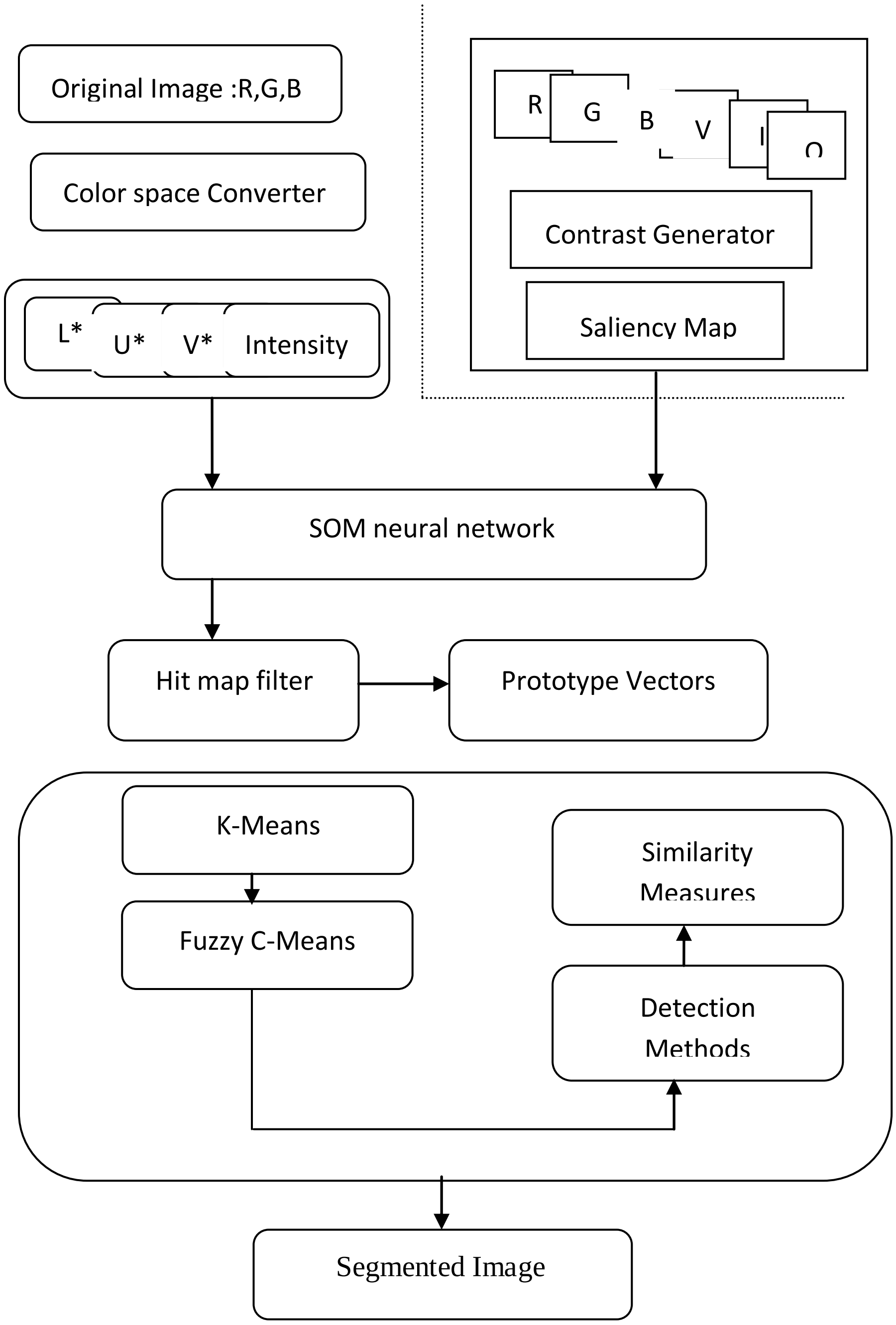
Step 1: The region number and the region centers are initialized, using the output of the max min algorithm.

Step 2: For every pixel p, the distance D (p,ri) between p and all region's intensity and spatial centers is calculated. The pixel is then assigned to the region for which the distance is minimal. Normalization of the spatial distance with the area of each region is necessary in order to enable the creation of large connected regions. Otherwise, pixels would tend to be assigned to smaller rather than larger regions due to greater spatial proximity to their centers. In this case, large objects would be broken down to more than one neighboring smaller regions instead of forming one single, larger region.

Step 3: Region centers are recalculated. Regions with areas below the threshold th size are dropped.

Step 4: The number of regions is also recalculated, taking into account only the remaining regions.

Step 5: If the number of regions K is equal to the one calculated in the previous iteration and the difference between the new centers and those in the previous iteration is below the corresponding threshold for all centers, then stop, else go to Step 2. Since there is no certainty that the K suppressed mean algorithm will converge for any given image, the maximum allowed number of iterations here was chosen to be 20.



System Architecture

3.3 FUZZY C-MEANS CLUSTERING ALGORITHM

Fuzzy C-Means algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Let $X = \{x_1, x_2, x_3 \ldots, x_n\}$ be the set of data points and $V = \{v_1, v_2, v_3 \ldots, v_c\}$ be the set of centers.

Step 1: Choose primary centroids Ci (prototypes)

Step 2: Compute the degree of membership of all feature vectors in all the clusters:

$$u_{ij} = \frac{\left[\frac{1}{d^2(X_j, C_k)}\right]^{\frac{1}{(q-1)}}}{\sum_{k=1}^{K}\left[\frac{1}{d^2(X_j, C_k)}\right]^{\frac{1}{(q-1)}}}$$

Step 3: Compute new centroids $C^{\wedge}{}_{i}$

$$C^{\wedge} i = \frac{\sum_{j=1}^{M}(u_{ij})^{q} X_j}{\sum_{j=1}^{M}(u_{ij})^{q}}$$

and update the memberships, u_{ij} to $u^{\wedge}_{ij}$ according to step 2.

Step 4: If $\max_{ij} ||u_{ij}-u^{\wedge}_{ij}||<tol$ stop, otherwise go to step 3.

It is important to note that the computation of the degree of membership u_{ij} depends on the definition of the distance measure $d^2(X_j,C_k)$.

4. QUALITY MEASURES

STANDARD DEVIATION

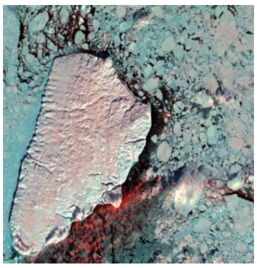
Standard deviation is a widely used measure of variability or diversity used in statistics and probability theory. A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data points are spread out over a large range of values.

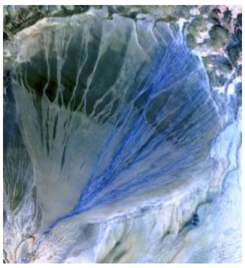
VARIANCE

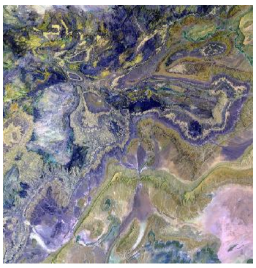
Variance is the limit up to which we can have the experimental numerical data which means all the possible, meaningful and effective values come under this limit. In this process we are apply the measure the quality of the data.

5. DATASET DESCRIPTION

We conduct our experiments on a satellite photos of earth. Data set which data is gathered from Web designer depot website. Three data sets are Akpatok Island- Canada, Alluvial Fan- China, Atlas Mountains. **Akpatok Island-Canada**: Akpatok Island lies in Ungavabay in Northern Quebec, Canada. Accessible only by air, Akpatok island rises out of the water as sheer clifts that soar 500 to 800 feet above the sea surface. The island is an important sanctuary for clift-nesting Sea birds. **Alluvial Fan, China:** A vast alluvial fan blossoms across the desolate landscape between the kulan and altun mountain ranges that from the southern border of the Taklimakan degert in china's Xinsiang province. **Atlas Mountains:** These are Anti-Alay mountains, part of the Atlas mountains range in Southern morocco, Africa. The region contains some of the world's largest and most diverse miheral resources. Most of which are still untouched.







Datasets: Akpatok Island-Canada **Alluvial Fan, China** **Atlas Mountains**

6. EXPERIMANTAL RESULTS AND ANALYSIS

The proposed system is a Self-Organizing Map tool for the segmentation of Color images. We are taken different types of Satellite photos of earth are considered to performance. Segmentation results derived from the proposed Clustering algorithm is compared with the K-Means clustering algorithm based on Standard deviation and Variance. In our work SOM-based Fuzzy C-Means Clustering algorithm gives better results than SOM based K-Means algorithm. Test results for Datasets: (a) L*U*V* images (b) K-Means Segmentation and Edge Detection methods (c) Fuzzy C-Means Segmentation and Edge detection methods.

(a)

(b)

(c)

TABLE- I

SHOW RESULTS OF STANDARD DEVIATION FOR VARIOUS CLUSTERING ALGORITHMS

Image name	K-Means	Fuzzy C-Means
Akpatok Island	34.07	41.52
Alluvin Fan	33.68	40.96
Atlas Mountains	34.03	41.80

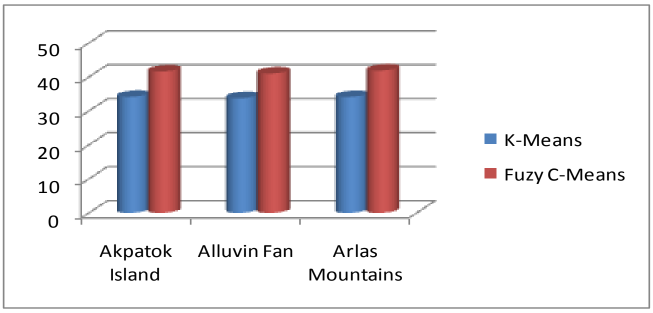


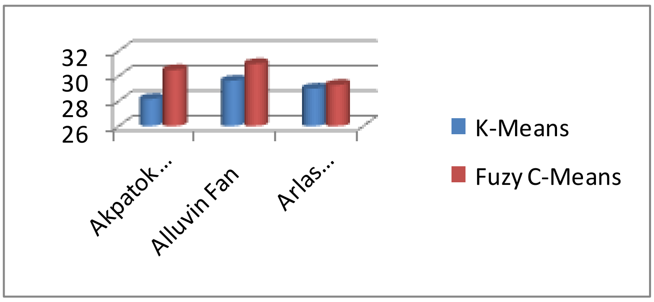
Fig 1Shows Standard Deviation

The above results show that the SOM based K-Means algorithm and SOM based Fuzzy C-Means Algorithm provides Standard Deviation for Image Segmentation . Figure I shows graph of Standard Deviation.

TABLE- II

SHOW RESULTS OF VARIANCE FOR VARIOUS CLUSTERING ALGORITHMS

Image name	K-Means	Fuzzy C-Means
Akpatok Island	28.16	30.48
Alluvin Fan	29.62	30.96
Atlas Mountains	28.99	29.28



The above results show that the SOM based K-Means algorithm and SOM based Fuzzy C-Means Algorithm provides Variance for Image Segmentation . Figure II shows graph of Variance.

7. CONCLUSION

In this paper Satellite photos of Earth Color Image Segmentation method based on Self-Organizing Map with K-Means algorithm, SOM-Based Fuzzy C-Means Algorithm have been implemented. The method trains features of intensity and L*U*V* color space with SOM neural network, the output prototype vectors are clustered with SOM-Based K-Means and SOM-Based Fuzzy C-Means algorithm. In our proposed algorithm SOM-Based Fuzzy C-Means Clustering algorithm gets better segmentation results. In future include other Edge detection methods and various Quality measures to be calculated for better results.

8. REFERENCES

[1] M. L. Gonc¸alves, M. L. A. Netto, J. A. F. Costa, and J. Zullo Junior, “An unsupervisedmethod of classifying remotely sensed images using Kohonen self-organizing maps and agglomerative hierarchical clustering methods,” *International Journal of Remote Sensing*, vol. 29, no. 11, pp. 3171–3207, 2008.
[2] D. M. Risti´c, M. Pavlovi´c, and I. Reljin, “Image segmentation method based on self-organizing maps and K-means algorithm,” in *Proceedings of the 9th Symposium on Neural Network Applications in Electrical Engineering (NEUREL ’08)*, pp. 27– 30, Belgrade, Serbia, September 2008.
[3] D. E. Ilea and P. F. Whelan, “CTex—an adaptive unsupervised segmentation algorithm based on color-texture coherence,” *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1926–1939, 2008.
[4] J. Vesanto and E. Alhoniemi, “Clustering of the self-organizing map,” *IEEE Transactions on Neural Networks*, vol. 11, no. 3, pp. 586–600, 2000.
[5] Z. Yu, O. C. Au, R. Zou, W. Yu, and J. Tian, “An adaptive unsupervised approach toward pixel clustering and color image segmentation,” *Pattern Recognition*, vol. 43, no. 5, pp. 1889–1906, 2010.
[6] K. T. Park and Y. S. Moon, “Automatic extraction of salient objects using feature maps,” in *Proceedings of the 32nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’07)*, vol. 1, pp. 617–620, Honolulu, Hawaii, USA, April 2007.

[7] V. V. Srinivas, S. Tripathi, A. R. Rao, and R. S. Govindaraju, "Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering," *Journal ofHydrology*, vol. 348, no. 1-2, pp. 148–166, 2008.

[8] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Transactions on Neural Networks*, vol. 11, no. 3, pp. 586–600, 2000.

[9] K. M. Shaaban and N. M. Omar, "Region-based deformable net for automatic color image segmentation," *Image and Vision Computing*, vol. 27, no. 10, pp. 1504–1514, 2009.

[10] Jianbo Shi & Jitendra Malik (1997) Normalized Cuts and Image Segmentation, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 731-737.

[11] T. Kanungo, D. M. Mount, N. Netanyahu, C. Piatko, R. Silverman, & A. Y.Wu (2002) An efficient k-means clustering algorithm: Analysis and implementation Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp.881-892.

[12] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp.1254–1259, 1998.

[13] K. Tas¸demir and E. Mer´enyi, "Exploiting data topology in visualization and clustering of self-organizing maps," *IEEE Transactions on Neural Networks*, vol. 20, no. 4, pp. 549–562, 2009.