Clustering of Digital Images using Shape Features with SOM

Mayank Singh, Viranjay M. Srivastava

Abstract - In these days people are interested in using digital images. So the size of image databases is increasing rapidly. It leads retrieval problem of images from large databases. Machine learning algorithms are applying in recent research to simplify the task of image retrieval and make it automatic. Thus the concept of content based image retrieval system came into existence. In this system the images are extracted based on similar content. Content means features of the images and it is formed by feature extraction of the images in databases. Contents can be edges, color, shape, gradient, orientation, histogram gradient etc. These contents are clustered into various groups of similar feature vectors. So for any input image the selected feature is searched for and image is retrieved from the database. This reduces the time complexity. There have been many algorithms for implementing the content based image retrieval system. In this research work we propose a novel paradigm where in shape features are extracted from the database images and are used to train the self-organizing map to cluster the shape features. These clusters are then used for image retrieval.

Keyword: Digital image clustering; Clustering techniques;content based image retrieval; shape features; self-organizing map.

I. INTRODUCTION

Image processing is a domain of digital image processing where large number of images undergoes certain type of transformation based on the transformation function. The transformation function can be image enhancement, image restoration and so on.Over the years image processing has seen many developments. It is widely used in information retrieval system where in a particular image is retrieved based on the relevant requested query from the database. There are many techniques which support the image retrieval from the image database.Researches focused on automatic image retrieval system based on machine learning concepts. Many techniques like histogram analysis of gradient, bag of features etc. have been implemented for image retrieval. To simplify the process similar images are grouped into clusters. Images are selected from the clusters based on some features this is content based image retrieval system. Content based image retrieval system thus helps in resolving the image retrieval problem. There are many techniques for implementing content based image retrieval system. These techniques utilize the concept of machine learning and image processing toperform the required task. It uses a special technique called feature extraction where in the image undergoes certain typeof linear or nonlinear transformation and unique feature like shape, histogram gradient, edges etc. are extracted from the input image and stored in the image database. This feature database is used with machine learning tools like K- Means or more recently support vector machines to cluster similar feature vectors in the data base. Now this feature vector database is used for similarity measure based on the input query image.

In this research work, we propose a method of feature extraction using edge extraction and use it to train a self organizing map which shall make feature vector clusters. These clusters are further used to retrieve the desired image from the database. This work is thus to make the number of clusters of images using Kohonen Map on the basis of Shape feature of the images. Content Based Image Retrieval is the application of computer techniques to resolve the digital image retrieval problem. In Content Based Image Retrieval, images are retrieved based on color, texture and shape. The CBIR system uses these options for retrieval of images and therefore the technique for obtaining these options is understood as Feature Extraction. For feature extraction we are actually working on the feature Shape. Further, for clustering of images we use Self-organizing Maps (SOM) approach.

II. RELATED WORK

Shanmugam et al. [1], ordering war scene from the characteristic scene by separating wavelet features. By utilizing in the wake of extricating wavelet features they are ordered by utilizing Artificial Neural Network and after that Support Vector Machines (SVM).ANN characterize the Image utilizing in reverse proliferation calculation and SVM order the Image utilizing spiral premise part work with p=5. On account of SVM, it gives just 59% characterization rate and on account of ANN, it gives just 72.5% order rate. Therefore ANN gives great arrangements bring about grouping war scene by separating wavelet features when contrasted and SVM.Hua Zhang et al. [2], propose a novel fluffy topology coordinated Support vector machines vector machine (SVM) (FTSVM) order system for remotely sensed Images in light of the standard SVM. It make limit fluffy convention is embodied into the standard SVM.

Jun Chen et al. [3] proposed another coordinated model is advanced for selecting 3PL suppliers in view of Support vector machines vector machine (SVM) and fuzzy analytic hierarchy process (FAHP).SVM is utilized as a part of the first stage to arrange all the ventures to be chosen. At that point fluffy AHP is embraced to gauge the fabulous endeavors which were chosen in the first stage.

Revised Manuscript Received on October31, 2019.

Mayank Singh *,Department of Electrical, Electronic & Computer Engineering, Howard College, University of KwaZulu Natal, Durban 4041, South Africa. Email: dr.mayank.singh@ieee.org
Viranjay M. Srivastava, Research Scholar, BITS, Pilani, India. Email: viranjay@ieee.org

Contrasted and the customary technique, the model in light of SVM-FAHP can enhance the choice proficiency and lessen the computational expense amid choice making procedure and the expense of data accumulation. The FAHP model can take care of the vulnerability issue viably when changing over the subjective case to quantitative ones. The case study demonstrates that the SVM-FAHP model is practical and compelling. The exploration can give choice making for undertakings to choose 3PL suppliers.

G. Mountrakis et al. [4] concentrated on an extensive variety of strategies for examination of airborne- and satellite-determined symbolism keeps on being proposed and surveyed.Zhibin Liu et al. [5] beat the deficiency of customary direct SCDA evaluation technique, proposes an improved Support vector machine (SVM) assessing strategy taking into account the multistage dynamic fluffy choice, takes the multistage dynamic fluffy judgment as the examining foundation, utilizes the SVM guideline to start assessment model.

Qiu Zhen Ge et al. [6] introduced the Image order issue as an Image surface learning issue by survey an Image as a gathering of locales, every got from Image division. A methodology performs a supportive quality mapping through a picked metric separation capacity.Farid Melgani et al. [7] considered the issue of the arrangement of hyper otherworldly remote sensing Images by Support vector machines (SVMs).F. Emecki et al. [8] consider a situation where numerous information sources are willing to run information mining calculations over the union of their information the length of every information source is ensured that its data that does not relate to another information source won't be uncovered.Bart Kuijpers et al. [9] consider security safeguarding choice tree instigation by means of ID3 for the situation where the preparation information is evenly or vertically disseminated.

III. BACKGROUND

Images may be two-dimensional, for example, a photo, screen show, and a three-dimensional, for example, a statue or visualization.An image as characterized on, say a photographic film is a persistent capacity of shine qualities. Every point on the created film can be connected with a dim level worth speaking to how brilliant that specific point is. To store Images in a PC we need to test and quantize (digitize) the Image capacity. Examining alludes to considering the Image just at a limited number of focuses [10]. Also, quantization alludes to the representation of the dark level quality at the sampling point utilizing limited number of bits. Every Image test is known as a pixel.A computerized Image is basically a gathering of pixel qualities. There are different organizations of putting away an image in a record. Basically they all include putting away the pixels values alongside other pertinent data about the image [11].

Content Based Image Retrieval also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision to the image retrieval problem, that is, the problem of searching for digital images in large databases.Following diagram shows how we can retrieve images from the data base with the help of features of the images.

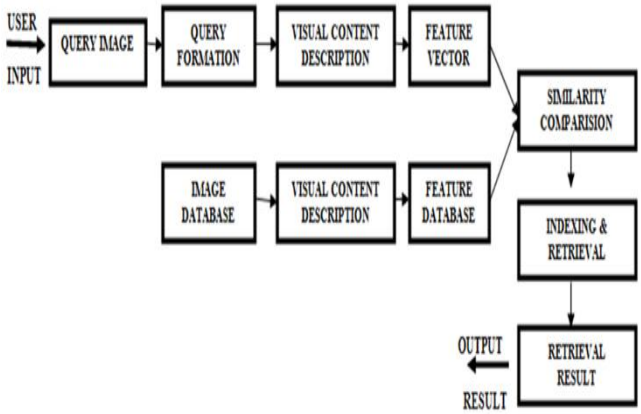

Fig. 1. Content base image retrieval system process

IV. DATASETS

The input data set is formed by considering the features of the input images. Sample input images have been downloaded from the internet. For this research work we have collected sample of images of geometrical shapes. About 50 images of three types of flowers rose, chrysanthemum and tulip are selected [12]. Features are measured for each flower. The features selected are: Length of sepal, Width of sepal, Length of petal, Width of petal. For SOM there is no need to specify the output, hence the data set has attributes i.e. ID, Sepal length, Sepal width, Petal length, Petal width and type.

Example of the data is

5.1 3.5 1.4 0.2 Tulip
3.9 2.0 1.4 0.2 Rose
6.7 3.2 1.3 0.2 chrysanthemum
3.6 2.1 1.5 0.2 Rose
5.0 3.6 1.4 0.2 chrysanthemum
5.4 3.9 1.7 0.4 Tulip
4.6 3.4 1.4 0.3 Rose

Fig. 2. Chrysanthemum, Rose and Tulip

V. METHODOLOGY

The proposed algorithm for clustering of digital images by extracting shape features from the image using SOM (CIFSOM) is an attempt for training SOM over the input data set of flowers. The CIFSOM is trained over the data and validated over the data set to give accurate clusters for the input dataset.

The procedure followed in the whole process of the CIFSOM is depicted in the figure 3.

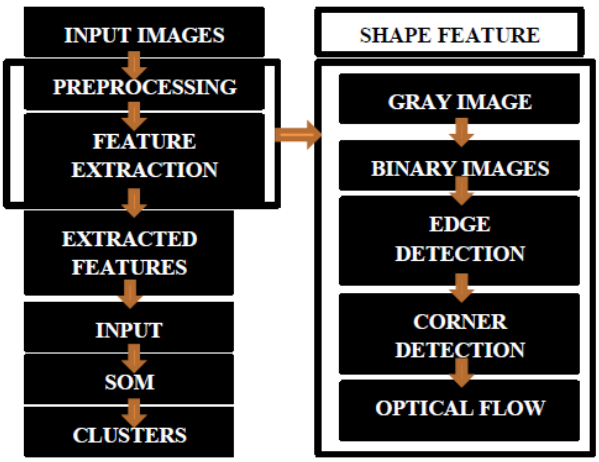

Fig. 3. Process diagram for working

Algorithm 1: CIFSOM algorithm

Step1: Collect large number of images of 3 flowers.
Step 2: Pre-process the images and extract their feature vectors to prepare the data set.
Step 3: Split the data set into training and testing data set. Use 90 -10 ration. Use the extracted feature of the images created at step 2 to train the SOM.
Step 4: SOM creates the Cluster. Repeat step 3 for testing the data set.
Step 5: Use the extracted features of the input images for matching the images
Step 6: If image is found then display similar images else display not similar images
Step 7: Stop

In the pre-processing, all images are downloaded and they are converted to gray level images. Their sizes are reduced to reduce the computational complexity. Furthermore in feature extraction step, the digital images feature are extracted based on the selected attributes and stored in the data set as specified by the SOM tool box format. The next step is to train the SOM. Use 90% of the dataset for the training purpose. The feature vector dataset is given to SOM for training. Training is performed using SOM algorithm for clustering. Furthermore, Using the features extracted from the dataset, edges are now used as feature matching techniques. Match points for the feature matched are founded. If there is a match point between two images, that display a message that images are similar and if there is no match point display images are not similar. At last, validation of the CIFSOM is done using different images of the flowers to measure the performance. Plot the cluster map of the SOM and the results obtained.

VI. Results and Discussions

A GUI for the proposed CIFSOM was developed by integrating the SOM Toolbox and MATLAB R2013b. The GUI has following functionalities that is shown in the figure 4.

- SOM Feature extraction push button for interface to SOM TOOLBOX.
- MATCH FEATURE push button for matching extracted features.
- Exit push button.

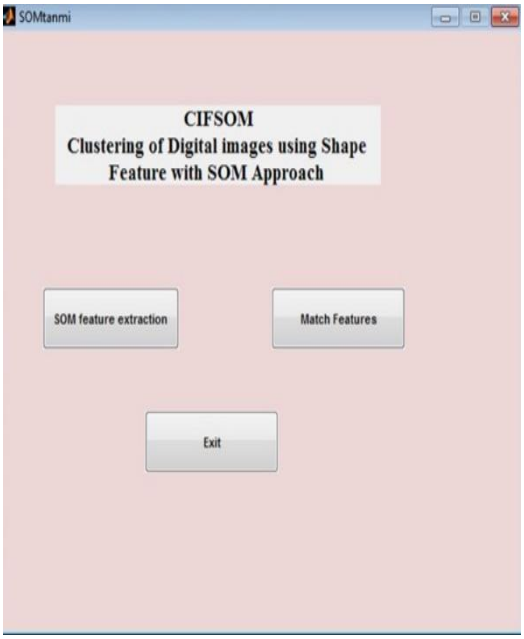

Fig. 4. CIFSOM GUI

The training and intialization in Self-organizing map toolkit with certain parameters is shown in figure 5.

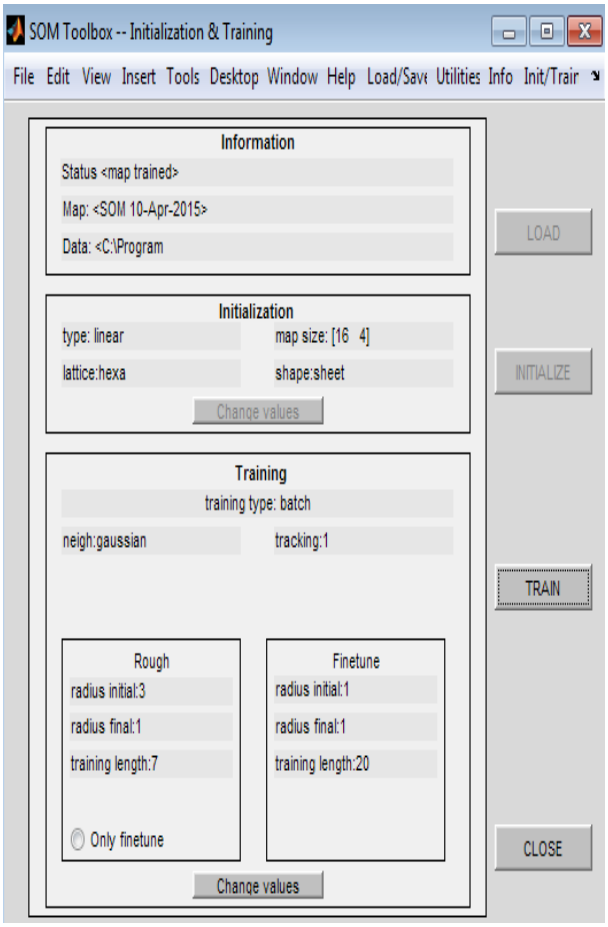

Fig. 5. Interface of SOM Toolbox

```
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Finetuning phase...

Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Training:    0/    0 s
Final quantization error: 0.303
Final topographic error:  0.067
```

Fig. 6. Training output

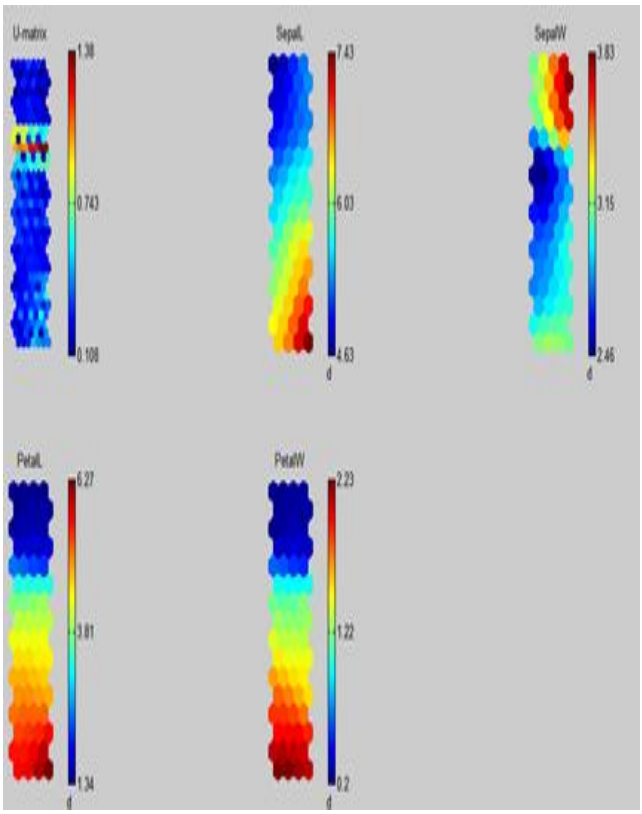

Fig. 7. Cluster

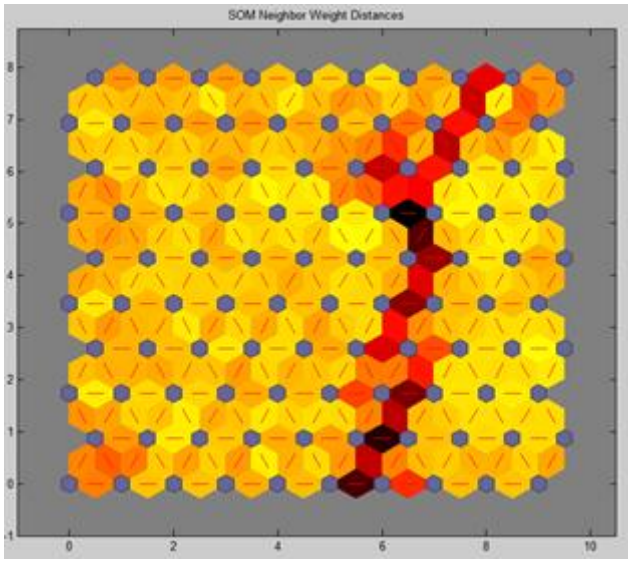

Fig. 8. Neighbour weight distance

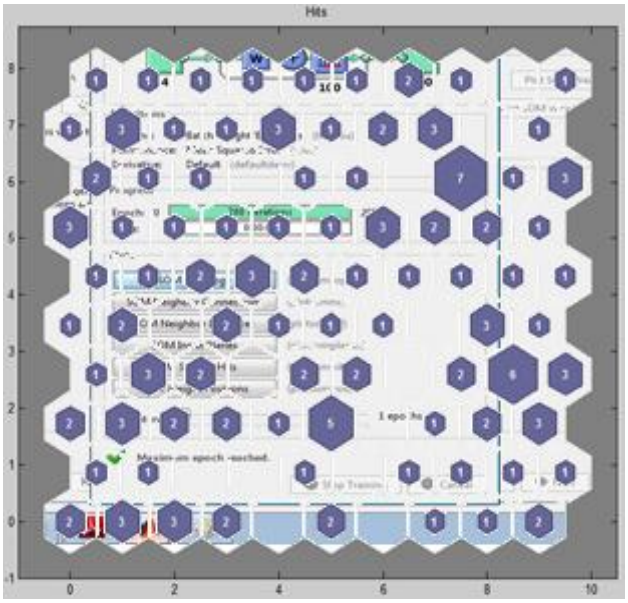

Fig. 9. SOM sample hits

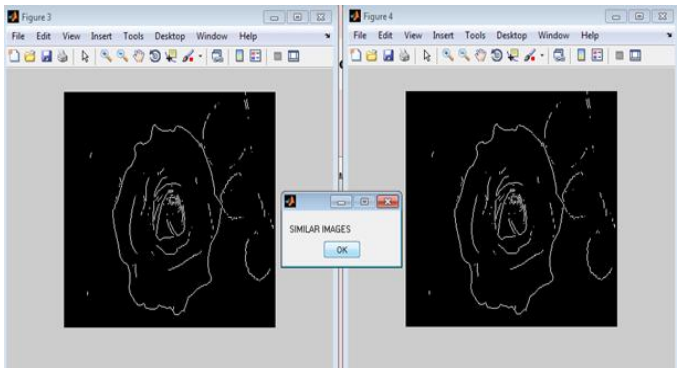

Fig. 10. Output of similar images

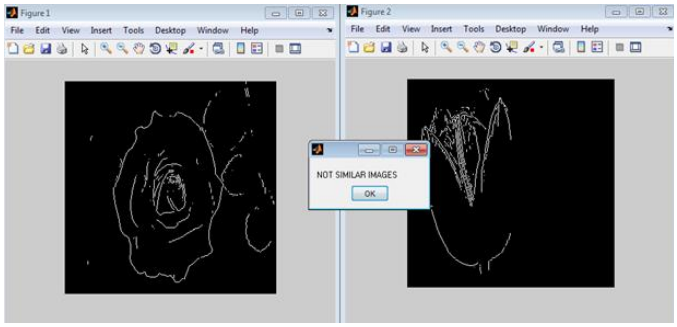

Fig. 10. Output of non-similar images

VII. CONCLUSIONS AND FUTURE DIRECTIVES

The proposed technique of CIFSOM was successfully implemented here. It was proven that SOM can be used for image clustering of digital images based on feature extraction. The experimental results show the performance of the proposed technique. The proposed technique provides high classification ratio, hence the accuracy is more. The training time is also less as compared to MATLAB based somformap function. It was experimentally proven that shape features can be extracted from the image and they can be used to train a self organizing map for clustering the dataset based on the feature vectors formed. The feature shapes generated can also be used to generate edges of the digital image and thus can be used to find out matching points between two images. If there are common matching points then one can use this technique for image matching or retrieval from data bases. The promising results using the proposed method can be further used for developing automatic image retrieval system from collection of large databases. This can be considered as a challenge and continued as a future work.

REFERENCES:

1. S. Daniel Madan Raja, Dr. A Shanmugam, "ANN and SVM Based War Scene Classification using Wavelet Features: A Comparative Study," Journal of Computational Information Systems, pp. 1402-1411–86, 2011.
2. Hua Zhang, Wenz hong Shi, and Kim fung Liu "Fuzzy-Topology-Integrated Support Vector Machine for Remotely Sensed Image Classification", IEEE Transactions on Geosciences And Remote Sensing, Vol. 50, No. 3, March 2012.
3. Jun fei Chen, Jiameng Zhong "An Integrated SVM and Fuzzy AHP Approach for Selecting Third Party Logistics Providers", Przeglad Elektrotechniczny (Electrical Review), ISSN 0033-2097, R. 88 NR 9b/2012.
4. Giorgos Mountrakis, Jungho Im, Caesar Ogole "Support vector machines in remote sensing", ISPRS Journal of Photogrammetry and Remote Sensing, 2010.
5. Zhibin Liu, Haifen Yang, Shaomei Yang "Integration of Multi-layer SVM Classifier and Multistage Dynamic Fuzzy Judgment and Its Application in SCDA Measurement", Journal of Computers, Vol. 4, No. 11, November 2009.
6. Qiu Zhen Ge, Zhang Chun Ling ,Li. Qiong, Xin Xian Hui , Guo Zhang "High Efficient Classification on Remote Sensing Images Based on SVM", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B2. Beijing 2008.
7. Farid Melgani, Lorenzo Bruzzone "Classification of Hyperspectral Remote Sensing Images With Support Vector Machines", IEEE Transactions on Geosciences And Remote Sensing, Vol. 42, No. 8, August 2004.
8. F. Emeki et al "Privacy preserving decision tree learning over multiple parties", Data & Knowledge Engineering 63, February 2007.
9. Bart Kuijpers, Vanessa Lemmens, Bart Moelans, "Privacy Preserving ID3 over Horizontally, Vertically and Grid Partitioned Data", Karl Tuyls Department of Industrial Design, Eindhoven University of Technology, The Netherlands, March 2008.
10. Shalini Soni, Punit Kumar Johari, "An Approach of K-means Clustering Based SVM ensemble for ImageRetrieval,"International Research Journal of Engineering and Technology, Volume: 05 Issue: 07 | July 2018, pp. 2486 – 2496.
11. T. Balaji, "An Automatic Color Feature Vector ClassificationBased on Clustering Method," International Journal of Research and Scientific Innovation (IJRSI) | Volume IV, Issue IV, April 2017, pp. 13-18.
12. Anand R, Veni S, Aravinth J, "An Application of image processing techniques forDetection of Diseases on Brinjal Leaves UsingK-Means Clustering Method,"2016 Fifth International Conference On Recent Trends In Information Technology, 2016.