Class Based Semantic Feature Similarity for Efficient Image Mining using BC Patterns

T. Rajendran, M. Baskar, T. Gnanasekaran, N. Mohamed Imtiaz

Abstract: The problem of image mining has been well studied. Image mining has been approached with several key features like color, texture and shapes. However, the efficiency of image retrieval is still a questioning fact. To improve the performance of image mining, an efficient Class Based Semantic Feature Similarity (CSFS) measure based image mining algorithm is presented in this paper. The method classifies the images under various semantic classes and according to the meanings of semantic class. First, the input image has been applied for noise removal and quality enhancement with Gabor filter and histogram equalization. Second, the input image has been extracted for Binary and Contrast (BC) patterns. Third, the method estimates semantic feature similarity for different classes. Finally a single one has been identified and the images of the class have been returned as result. The proposed algorithm improves the performance of image mining and reduces false classification ratio.

***Index Terms*: Image Mining, Semantics, BC Pattern, CSFS.**

I. INTRODUCTION

The way of presenting information has been changed from text to images and videos now a day. Number of organization keeps their documents in form of images and videos instead of keeping them in original text or in papers. The organizations maintain their information in form of images and retrieve them whenever required. When the numbers of images are less, identifying them from few lists is quite easier, but when the size and number or volume of images are increasing, then identifying the related similar images is not possible. This requires the automated systems to find the relevant similar images from large pool.

Image mining is the process of identifying and retrieving the similar set of images from large pool of image database. In an image data base, there will be images indexed belongs to various categories or various classes. Similarly, each class would have number of sub classes under which the images are classified. The images are classified or indexed based on various features. The basic feature being used is color, shape and textures. Among these features, any one or multiple features can be used. The accuracy of classification and retrieval is highly depending on the measure of similarity. If the algorithm uses multiple features in estimating the similarity, then the accuracy can be improved.

In the modern trend, the images has been given various meanings and they can be classified under the meanings also. The semantics is the way of representing the images with set of meanings and features. For example, the cricket images can be classified based on various concepts like ball, bat, stump, and so on. However, the images are belongs to the objects but they can be classified in to cricket as the meanings has been shared. Similarly, various images can be grouped under specific class names. Similarly, when you specify football, the images of different players playing football can be classified. By using semantic meanings, the set of images of similar meanings can be grouped.

Considering, the color feature for image mining, would produce higher false results with higher irrelevancy. Similarly, number of algorithms available for image mining, but produces poor results. To improve the performance, a class based semantic feature similarity based algorithm is presented in this paper. The method generates binary patterns and contrast pattern for the retrieval of images. Based on the BC patterns, the method estimates the similarity measure. The detailed approach is discussed in the next section.

II. RELATED WORKS

There are many algorithms discussed for the problem of image mining. This section presents a short review on methods of image mining.

In [1], an object based automatic crawling system has been presented. The method is presented towards user centric web solution in image retrieval.

In [2], a content based retrieval based on edge and inner features is presented. The method split the image into number of windows and for each of them, the edges and inner features are identified. Based on the EI features, the method estimates cluster similarity to identify the similar images.

In [3], an SIFT based CBIR algorithm is presented. The method extract the SIFT features from the input image and applies Bacteria foraging optimization algorithm (BFOA) to identify the efficient results.

In [4], a global and local feature based image retrieval algorithm is presented. The method combines both SURF (Speed Up Robust Feature) and HOG (Histogram Orientation Gradients) features to extract local and global features. Using both, the method estimates similarity to retrieve the images.

In [5], a texture based image mining algorithm is presented. The method first applies the gabor filter to remove the noise. Then, the

Revised Manuscript Received on April 06, 2019.

T. Rajendran, Assistant Professor, Department of CSE, GRT Institute of Engineering and Technology, Tamilnadu, India.

Dr. M. Baskar, Professor, Department of CSE, MLR Institute Technology, Dundigal, Hyderabad, Telangana, India.

Dr. T. Gnanasekaran, Professor, R.M.K. Engineering College, Kavarapetai, Tamilnadu, India.

N. Mohamed Imtiaz ,Assistant Professor, Department of CSE, GRT Institute of Engineering and Technology, Tamilnadu, India.

image has been split into number of windows and generates location features which are non-overlapping. Using the features extracted, the bag of words has been generated. Generated features have been used to measure similarity to perform image mining.

In [6], an low level image feature based mining algorithm is presented which extracts low level features from textual images by generating the GLCM feature. Based on the feature extracted, the distance between each feature of GLCM has been measured. According to the value of GLCM similarity, the image retrieval has been performed.

In [7], the concept of semantic based image retrieval has been detailed and the author presented a survey on various methods of semantic based image retrieval. The paper details how the objects are identified based on concepts, how the low level features are related with query and the way of considering the feedback to produce results to the user.

In [8], an image mining algorithm with textual and image feature has been presented. The method extract text features from input query where the local feature has been extracted. The mutual information filter has been used to select the features and converted into visual descriptors. Using these, the similarity measure has been estimated to perform image retrieval.

A combination of visual and texture feature based image mining algorithm is presented in [9]. The method generates bag of words from textual features and bag of visual words from images. Using both the set of words, the method identifies the list of images related to the query.

A relevancy feedback based image retrieval algorithm is presented in [10]. The method works iteratively to get the feedback from the users and generate results accordingly. The support vector machine has been used for classification. In [11], various algorithms and descriptors for image retrieval have been analyzed. The analysis performed by combining various image descriptors in image retrieval.

All the above discussed algorithms produce poor results in image retrieval with higher irrelevancy.

III. CLASS BASED SEMANTIC IMAGE RETRIEVAL

The proposed Class based semantic feature similarity based algorithm reads both input query and image. From input query, the method extracts features by applying natural language processing. Second, the method extracts the low level and high level features. Using all these, the method estimates the semantic feature similarity on different type of features to perform image retrieval. The detailed approach is presented below:

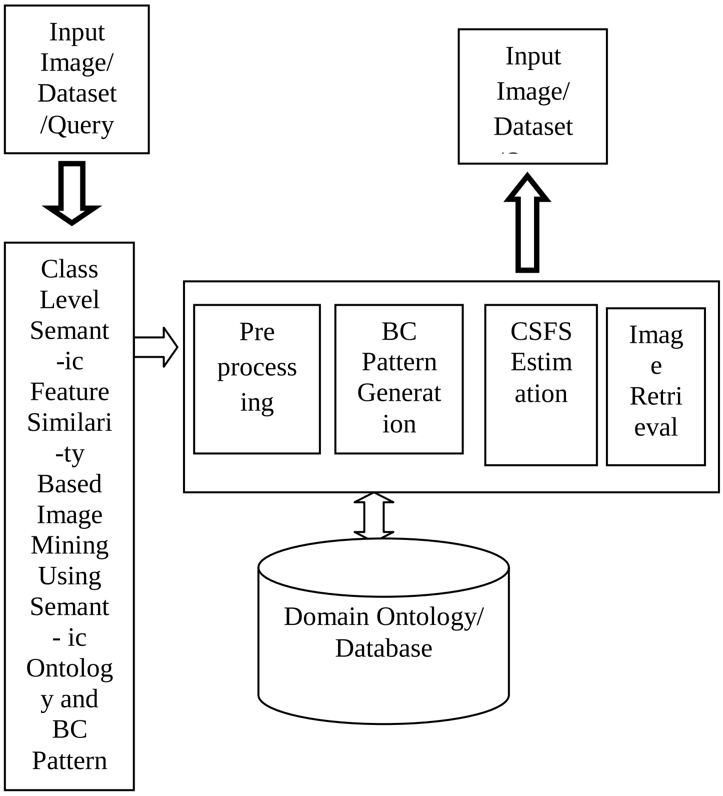

Figure 1: Architecture of proposed CSFS Image Retrieval System

The Figure 1, present the architecture of proposed CSFS image retrieval algorithm and shows various components.

A. Preprocessing

The input image has been read and the presence of noise has been removed initially. The Gabor filter has been applied to remove noise at different levels. The noise removed image has been used to extract the high level and low features. On the other side, the method reads the input query. From the input query, the method extracts the list of key words by applying the NLP technique. The stop word from the input query has been removed and stemming is performed. Finally, POS Tagger has been applied to produce the key terms.

Algorithm:

Input: Image img, Query q

Return: Preprocessed Image Pimg, Query term Set QTs.

Start

Read input image img and query Q.

Initialize Gabor filter Gf = $\int_{i=1}^{no\ of\ level} Initialize(Gf(i, Coordinates))$

Pimg = Gf(img)

Term set Ts = Split(QTs,' ')

Apply stop word removal.

Ts = $\int_{i=1}^{size\,(Ts)} if\ Ts(i) \in Sl : Ts = Ts \cap Ts(i)$

For each term Ti

Ti = Stemming (Ti)

Token = PoSTag (Ti)

If Token==Noun then

QTS = $\int \sum (Terms \in QTS) \cup Ti$

End

End

Stop

The above discussed algorithm performs noise removal from input image and extracts the terms from the input query which are used to perform image retrieval.

B. Pattern Generation

The preprocessed image has various features. From the preprocessed image, the method convert the image into binary image. From the binary image, the method generates binary pattern. Also, from the original preprocessed image, the method extracts the contrast features. Both contrast and binary features are converted into feature vector to perform image retrieval.

Algorithm:

Input: preprocessed image Pimg
Output: Vectors of Binary, Contrast
Start
Read image pimg
Extract contrast features contrast = $\int_{i=1}^{size(pimg)} \sum Contrast(pimg(i))$
Pimg = Binarize(pimg)
Extract Binary feature Binary = $\int_{i=1}^{size(Pimg)} Binary(pimg(i))$
Stop

The above discussed algorithm represents the working principle of BC pattern generation. The BC pattern has been used to perform image retrieval.

CSFS Estimation:

The class based semantic feature similarity has been measured based on both textual feature and the image features. Using the text feature, the method estimate the conceptual semantic similarity (CSS) where using the image features, the method estimates conceptual image semantic feature similarity (CISFS). Using both of them the method estimates the CSFS measure towards image retrieval.

Algorithm:

Input : Binary vector Bv, Contrast vector Cv, QTs, Ontology So
Output: Class C
Start
Read Bv, Cv and QTs
For each class c

Compute Css = $\frac{\sum_{i=1}^{size(So(c))} So(c(i)) \in QTs}{size(QTs)}$

For each image feature im

Compute CISFS = $\frac{\sum_{i=1}^{size(c)} im(i)(BV) \equiv BV}{size(c)} \times \frac{\sum_{i=1}^{size(c)} im(i)(CV) \equiv CV}{size(c)}$

Compute CSFS = $Css \times CISFS$

End

Compute cumulative CSFS = $\frac{\sum CSFS}{size(c)}$

End

Choose the class with maximum c = Max (CSFS)

Stop

The above discussed algorithm estimates the class based semantic feature similarity measure on various classes. Based on the value of CSFS, a single class has been selected as result.

C. CSFS Image Retrieval

In this stage, the method reads the query and input image. The input query has been extracted with term features using natural language processing techniques. The input image has been extracted with binary and contrast features. Using these features, the method estimates the class based semantic feature similarity measure. The class with maximum value of CSFS has been identified as target class. The images from the target class have been retrieved as result.

Algorithm:

Input: Query Q, Image Im, Ontology O
Output: Resultant Image Set Ris
Start
Read Q, Im, O
[Term Set Ts, Preprocessed Image Pimg] = Preprocessing(Q, Im)
[Binary vector Bv, Contrast vector Cv] = BC Pattern Generation(pimg)
Class C = CSFS-Estimation(BV,CV,Ts)
Ris = $\sum Images \in C$
Stop

The above discussed algorithm retrieves the image based on the class level semantic feature similarity measure.

IV. RESULTS AND DISCUSSION

The proposed algorithm has been implemented using matlab. The method has been evaluated for its efficiency using different data sets. The performance of the method has been measured under various factors and presented in this section. The result produced has been compared with the results of other methods. The results produced is presented below:

Parameter	Value
Tool Used	Matlab
Dataset Used	UCI
Total Semantics	100
Total no of Images	1 Million

Table 1: Details of data set

The details of data set being used for the evaluation of proposed algorithm have been presented in Table 1. Using the data set, the method has been measured for its performance in various factors as follows:

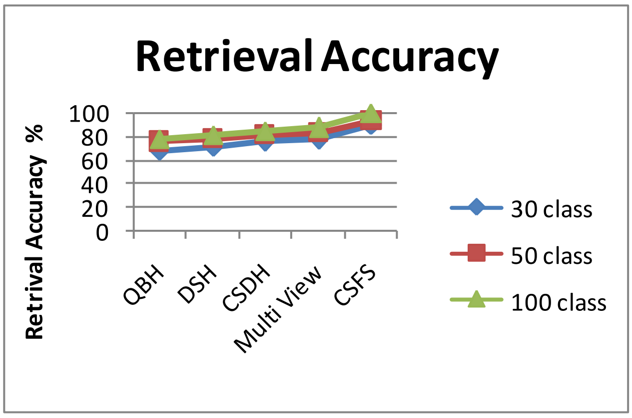

Figure 2: Performance on Retrieval Accuracy

The performance of the algorithm in retrieval accuracy has been measured and compared with the result of other methods. The proposed CSFS algorithm has produced higher retrieval accuracy than any other method.

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

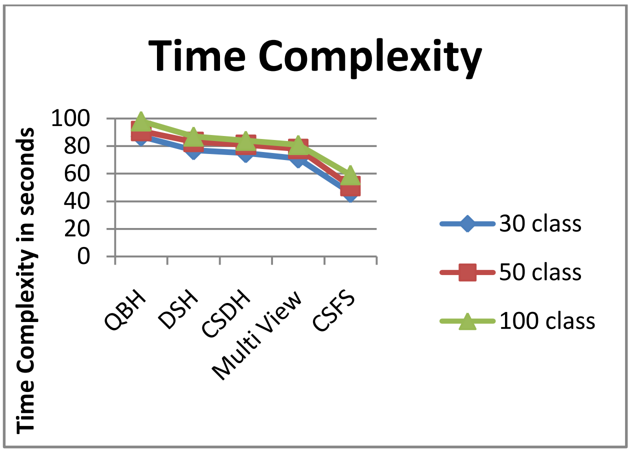

Figure 3: performance on time complexity

The performance of the methods in terms of time complexity has been measured and compared with the result of other methods. The proposed CSFS algorithm has produced less time complexity than any other method.
The performance on irrelevant ratio has been measured and compared with the result of other methods. The proposed CSFS algorithm has produced less irrelevancy than any other method.

V. CONCLUSION

In this paper, an efficient class based semantic feature similarity based image retrieval algorithm is presented. The method reads the input query and image to extract the term feature and high/low image features. Extracted features are converted to feature descriptors. Using the descriptors, the method estimates conceptual semantic similarity (CSS) on text features, and conceptual image semantic feature similarity (CISFS). Using both, the method estimates CSFS to identify the class with maximum value. The image of identified class has been retrieved as result to the user. The method improves the performance of image retrieval and reduces the false classification ratio with less time complexity.

REFERENCES

1. ShashiShekhar, An object centric image retrieval framework using multi-agent model for retrieving non-redundant web images, International Journal of Image Mining, vol.1, No. 1, 2015.
2. M. Yasmin,An Efficient Content Based Image Retrieval using EI Classification and Color Features, Journal of Applied Research and Technology, vol.12, No.5, 2014.
3. AasiaAli , Content based image retrieval using feature extraction with machine learning, IEEE International Conference on Intelligent Computing and Control Systems, 2017.
4. ZahidMehmood, Content-Based Image Retrieval Based on Visual Words Fusion Versus Features Fusion of Local and Global Features, Springer, Arabian Journal for Science and Engineering, vol.43, issue. 12, 2018.
5. RushiLanSiZhongZhenbing Liu, A simple texture feature for retrieval of medical images, Multimedia Tools and Applications, Vol.77, Issue. 9, 2018.
6. AnkitaTripathi ; ShivamPandey ; Hitesh Jangir , An improved and efficient image mining technique for classification of textual images using low-level image features, IEE International conference on Inventive Computation Technologies (ICICT), 2016.
7. X. Huang, T. Zhao, C. Yu, M. Xiangming, T. Pierre, "Towards the improvement of textual anatomy image classification using image local features", Proceedings of the 2011 international ACM workshop on Medical multimedia analysis and retrieval, pp. 25-30, 2011.
8. O. Augereau, N. Journet, A. Vialard, J. Domenger, "Improving the classification of an industrial document image database by combining visual and textual features", Document Analysis Systems (DAS)llth IAPR International Workshop on, pp. 314-318, 2014.
9. J. Wang, H. Yonghao, K. Cuicui, X. Shiming, P. Chunhong, "Image-Text Cross-Modal Retrieval via Modality-Specific Feature Learning", Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, pp. 347-354, 2015.
10. Hoque, E., O. Hoeber, G. Strong, M. Gong. Combining Conceptual Query Expansion and Visual Search Results Exploration for Web Image Retrieval. – Journal of Ambient Intelligence and Humanized Computing, 2013, pp. 1-12.
11. ZhegaoPiao, Hyung-GeunAhn, SeongJoonYoo, Performance analysis of combined descriptors for similar crop disease image retrieval, Springer, Cluster Computing, vol 20,issue 4, pp. 3565-3577, 2017.

AUTHORS PROFILE

T. Rajendran received M.C.A. from Anna University, Chennai and M.E. Computer Science and Engineering from St.Peter's University, Chennai. His Area of research interest includes Data Mining, Image Mining, Semantic Ontology, Cloud Computing and Network Security.

Dr. M. Baskar received B.E. Computer Science and Engineering from Anna University, Chennai, M.Tech. Information Technology from Sathyabama University, Chennai and Ph.D., (Information and Communication Engineering) from Anna University, Chennai. His Area of research interest includes Computer Networks and Security, Parallel and Distributed Systems, Image Processing, Big Data, Machine Learning and IoT. He is published 19 Research Article in reputed International Journals and 10 Article in International Conferences. He is acting as a reviewer in Cluster Computing, Journal of Web Engineering, Multimedia Tools and Applications, Neural Processing Letters and Concurrency and Computation: Practice and Experience. He is a Life time Professional body member of CSI, ISTE, IET, ISRD, IRED, IACSET, IAENG, SDIWC and UACEE.

T. Gnanasekaran received Ph.D. from Anna University, Chennai. . His Area of research interest includes Computer networks and Security, Parallel and Distributed Systems, Cluster computing, Wireless Sensor Networks, Cloud Computing, Data Mining and WIMAX.

Mohamed Imtiaz.N received M.Tech. Information Technology from Sathyabama University, Chennai. His Area of research interest includes Big Data, Cloud Computing and Internet Security.