ISSN:0975-9646

A Review on Image Mining Frameworks and Techniques

T. Y. Gajjar[1], N. C. Chauhan[2]

[1]*Birla Vishwakarma Mahavidyalaya, V. V. Nagar, Gujarat, India*
[2]*A. D. Patel Institute of Technology, New V. V. Nagar, Gujarat, India*

***Abstract* - Image mining is used to discover the knowledge from the image dataset. Similar to data mining tasks, image mining task can also be categorized into classification, clustering, association rule mining, and characterization based summarization. This paper contains introduction to image mining and also elaborates image mining frameworks. The paper also reviews image mining techniques based on decision tree, neural network and association rule mining given by different researchers.**

I. INTRODUCTION

Use of the digital data is increased due to the digitalization of each sector. Digital data is available in the form of text, audio and video. Mostly knowledge discovery from the data stored in typical alphanumeric database, such as relational database, has been focus of work in data mining. Nowadays, non-standard databases like images are also used. Mining of these image datasets can be performed to discover the knowledge not explicitly stored in images. Image mining is the concept used to extract useful knowledge from images. It deals with the extracting inherent and embedded knowledge, image data relationship, or other pattern which is not explicitly found in the images [1, 9]. Mining large collection of images, and combined data mining of large collections of images with associated alphanumeric data are the two important themes of image mining [2]. Image mining is still at the experimental stage and growing filed for research. Lack of understanding the research issues of image mining is the obstacle to rapid progress. Image mining is not just the expansion of data mining to image domain. It can be considered to be a an efficient hybridization of image processing and data mining concepts to extract the useful knowledge. Various application domains of image mining includes natural scene recognition, remote sensing, egeria detection, weather forecasting, criminal investigation, image segmentation, etc.

II. REVIEW OF IMAGE MINING FRAMEWORK

There are two different frameworks of image mining have been proposed by researchers [10, 11]: function driven framework and information driven framework.

Most of the existing image mining system architectures fall under the function driven framework. However, function driven framework is not a generalized framework. It can be application oriented or organization oriented. Mihai, et al. [11] have given function driven framework for intelligent satellite mining system. Function driven framework for the MultiMediaMiner is given in [12]. The drawback of this framework is that it can organize and clarify the different tasks to be performed in image mining, but it is not able to differentiate levels of information representation which is necessary for image data to perform meaningful mining. This drawback is overcome in another information driven framework.

Zhang et al. [10] have given information driven framework for the image mining representing different levels of information. This framework has four levels of information pixel level, object level, semantic level and pattern and knowledge level. Pixel level is the lowest level in an image mining system. It deals with the raw image information such as image pixels and the primitive image features such as color, texture, and shape. It can answer the query like "Retrieve the image with red color". But it cannot answer the query like "Retrieve the image of girl". Next level is object level which can retrieve images for such queries. It deals with object information based on the primitive features in the pixel level. Object recognition assigns correct labels to a single region or set of regions. But still it cannot retrieve images for query like "Image with sad faces". The third Semantic concept level takes into consideration domain knowledge to generate high-level semantic concepts from the identified objects to answer such queries. These three levels are retrieving information for the image itself to mine it. If the user query is like "retrieve all the heart images from last one month" then it needs to consider alphanumeric data corresponding to images. It supports all the information needs within the image mining framework.

III. REVIEW OF IMAGE CLASSIFICATION BASED METHODS

In this section, a brief review of image classification as an image mining task is presented using two classification methods: decision tree based methods, and neural network based methods.

Sethi, et al. [3] has taken database of 2100 images. They have used Access database to store the feature vectors extracted from the images. From each image they have extracted

quantized global historical. For global histogram following quantization scheme was obtained based on the images in the database.

Hue: (H1, H2, H3, H4, H5, H6, H7, H8) = (0, 25, 41, 61, 138, 200, 213, 241, 359)

Saturation: (S1, S2, S3, S4, S5, S6, S7, S8) = (0, 9, 16, 24, 33, 42, 56, 80, 100)

Lightness: (L1, L2, L3, L4, L5, L6, L7, L8) = (0, 13, 27, 38, 47, 56, 65, 78, 100)

Four classes they have considered: Sunset, Arid, Marine and Noctune. For each image they have created a feature vector with 24 components: 1 to 8 related to hue, 9 to 16 for saturation, and 17 to 24 for lightness. The decision tree was induced using the CART option from the SIPINA software package [4]. All the decision trees induction methodologies allowed by this software package were applied to this problem. Association rules were extracted from a set of decision trees induced over the set of extracted low-level features from the images.

Kun-che Lu and Don-lin Yang [5] proposed a general framework based on the decision tree for mining and processing image data. The proposed model can be very efficient and effective for image processing and image mining.. Input data for the presented model was formatted as a set of equal sized raw and label image pairs. In image transformation part, they transformed them into database like table. In data reduction step redundant information, generated due to similar feature vector for neighboring area pixels, in the result table is removed. After that mining algorithms can be used on it. Authors have chosen the decision tree for this purpose. This decision tree classifier can be converted to if-else statements which can provide useful information about the training image

Mausoko, et al. [6] have taken 25 images of size 1024 × 768 pixels representing microscopic wax structure of leaves from trees. They have used unsupervised texture image classification algorithm using competitive neural network. Their goal is to study major problems of texture analysis, including the classification of texture and propose solution based on wavelet transformation. Neural network design algorithm implemented by MATLAB Neural Network toolbox and algorithms based on mathematical computation and empirical computation for determining the class boundary are given by the authors in their work. They have concluded that unsupervised classification of textured images has successfully adopted the neural network approach. But there are some problems which are associated with the choice of wavelets, suitable level of decompositions.

IV. Association Rule Mining Based On Image Content

Ordonez and Omiecinski discussed an algorithm for mining association rules for images [2]. This algorithm reduces I/O and CPU overhead. They have built their data mining system on the top of content-based image retrieval (CBIR) system. An algorithm first segment images into blobs. Then identify and label objects in the images. Using similarity function, similarity between two blobs is found as follows.

$$similarity = e^{-\frac{distance(blob_1, blob_2)}{2}}$$

$$distance(blob_1, blob_2) = \sqrt{(blob_1, blob_2)^T \Sigma^{-1} (blob_1, blob_2)}$$

Similarity measure is one if perfect match on all desired features and zero if match becomes worse. Then generate auxiliary images with identified objects to interpret the association rules. Data mining algorithm is applied to produce object association rules.

Yi, et al. [7] had given the description about how the association rule can be used for image retrieval. For each query image, find all association rules which use the query image as the antecedent (A). The consequent (B) are the candidate images for retrieval. Rank those candidate images by their confidence value. If the confidence value is higher than the rank will be higher. Support value of rule A $\Longrightarrow$ B is greater than A $\Longrightarrow$ C if B is a subset of C. If the candidate image set is empty or having images less than number of images to be presented then the system randomly picked several images from the database and that would give every image the chance to establish the association rules.

Deshpande has presented a data mining approach to find association rules based on image content [8]. Her purpose of experiment is to explore the feasibility of data mining approach in image mining based on content. Traditional association rule algorithms adopt an iterative method to discovery frequent item set, which requires very large calculations. In the overall approach, the image mining was divided into four important phases: image preprocessing, feature extraction, conversion of image database to transaction database, and applying association rule mining to this transaction database. She proposed a new association rule algorithm in order to reduce the number of scans of Apriori algorithm. This algorithm is described in four important steps. In the first step the transaction database was transformed into Boolean matrix. In the second step, frequent 1 itemset L1 is generated. The Boolean matrix is pruned by deleting some rows and columns. In the last step, frequent k-item sets Lk (k>1) are generated.

V. CONCLUSION

The image mining task on image datasets majorly deals with classification, clustering, and/or mining of knowledge from images using association rules. It can be used to group the images on web, efficient retrieval of images, or to extract hidden meaningful information from image datasets which is not explicitly available from image sources. In the review, a study of image mining techniques for classification have been carried out using neural network and decision tree based methods. A review of association rule mining methods based on image contents has also been carried out and included in this paper. However, researchers have explored these areas by developing different methods for image mining, still there many more chances for the research in this field. New techniques for the image mining are being generated and also there are scopes for improvement of the existing frameworks/techniques for large number of applications.

REFERENCES

[1] C. Lakshmi Devasena, T.Sumathi, Dr. M. Hemalatha, *"An Experiential Survey on Image Mining Tools, Techniques and Applications"* International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 3 Mar 2011.

[2] C. Ordonez and E. Omiecinski, *"Image Mining: A New Approach for Data Mining"*, 1998.

[3] I. Sethi, I. Coman, D. Stan, *"Mining Association Rules between Low-level Image Features and High-level Concepts"* Proc. of AeroSense 2001, Data Mining and Knowledge Discovery: Theory, Tools and Technology III, Orlando, Florida, USA, pp. 279-290, 16-20 April 2001.

[4] SIPINA for Windows. ERIC Laboratory, University of Lyon 2, http://eric.univ-lyon2.fr/ ricco/sipina.html.

[5] K. Lu and D. Yang, *"Image Processing and Image Mining using Decision Trees"* Journal of Information Science and Engineering 25, 989-1003, 2009.

[6] V. Musoko, M. Kol´ınov´a, A. Proch´azka, *"Image Classification Using Competitive Neural Networks"* 12th conference MATLAB 2004, Prague, Czech Republic, 25 November, 2004.

[7] H. Yi, D. Rajan and L. Chia, *"ARIRS: Association Rule Based Image Retrieval System"*, International Workshop for Advanced Imaging Technology (IWAIT'05), 2005

[8] D. Deshpande, *"Association Rule Mining Based on Image Content"*, International Journal of Information Technology and Knowledge Management, Vol. 4, No. 1, pp. 143-146, Jan-Jun 2011.

[9] J. Han and M. Kamber, *"Data Mining: Concepts and Techniques"*, 2/E, Morgan Kaufmann Publication, 2006.

[10] J. Zhang, W. Hsu and M. L. Lee, *"An Information driven Framework for Image Mining"*, in Proceedings 12th International Conference on Database and Expert Systems Applications (DEXA), Munich, Germany, September 2001.

[11] M. Datcu and K. Seidel. *"Image information mining: exploration of image content in large archives"*, IEEE Conference on Aerospace, Vol.3, 2000.

[12] O. R. Zaiane, J. W. Han et al., *"Mining MultiMedia Data"*, CASCON'98: Meeting of Minds, pp 83-96, Toronto, Canada, November 1998.