



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

A Review of Image Mining Concepts, Techniques and Challenges

Dr. Rina, Shailja Agnihotri,

Assistant Professor, Department of IT, GGSDS College, Chandigarh, Punjab, India

Assistant Professor, Department of IT, GGSDS College, Chandigarh, Punjab, India

ABSTRACT: Along with the advancement in the technology, the data volume is also increasing and becoming huge day by day. And if the database is the image database, then the problem is more complex than expected and needs more of attention. In almost every field, the image databases are required, may be medicine, geographical systems, robotics, health sciences etc. So it plays a vital role in research and development areas. The main idea is extraction and discovery of new information or knowledge from the images present in the database. This extraction is known as Image Mining. It is more advanced field in data mining. It is different from data mining as it is focused on images and extraction of information from the images only. The relationship between the image sets and other patterns are mined based on the users requirements. Various algorithms have been developed and used for mining images but more work needs to be done, for the results to be more specific, precise, accurate and effective. This paper focuses on the current techniques and approaches for mining data from images and identifies the challenges and the future of the research in this area.

KEYWORDS: data mining, image mining, data sets, clustering, segmentation, relational databases

I. INTRODUCTION

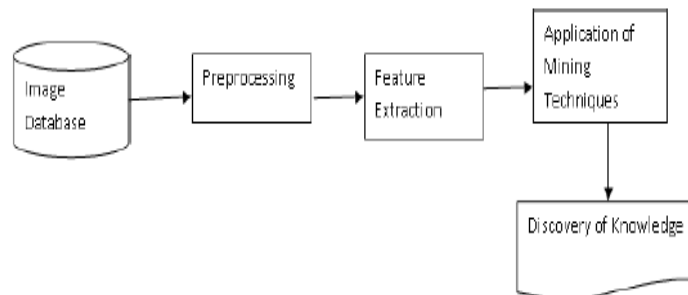
Image mining is the discovery of knowledge from the images. It also takes into consideration the image data association and further patterns which are implicitly defined with the images. The data is typically stored in alphanumeric databases for e.g. relational databases but more focus should be laid down on the image datasets. Extraction of relevant, meaningful information from these image databases is far too difficult. The issues related to image mining also involves content based image retrieval, data mining, image retrieval and also the databases. The research done in this area is still in its initial stages; much more have to be done to produce meaningful extraction. Image mining systems which can automatically extract semantically meaningful knowledge are in demand. The area of image mining should not be confused with the computer vision and image processing as the later only focus on single image and only extracts the specific features. Also sometimes the image mining overlaps the CBIR i.e. Content Based Image Retrieval where the query generated by the user is matched from the image database and the resultant matching image is produced but in contrast with all these discussed, the image mining deals with large volume of data as well as extracts or inferences new knowledge from the images available in the database.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

The image mining involves the following steps:



- a) Image analysis which includes preprocessing, object recognition, and feature extraction
- b) Image classification
- c) Image indexing
- d) Image retrieval

The fundamental issue is to retrieve or discover the information from the images without the loss of the present information. These can be stated as Firstly, issue of the representation of image pattern, which elaborates on that the current spatial or contextual information, should be retained after the knowledge extraction. Secondly, image feature should be selected in such a way that discovered knowledge remains meaningful. Thirdly, the mined patterns should be presented to the user in a way which fulfills the users requirement.

II. APPLICATIONS

Image databases are much of use in every research area. It has proved its importance as the critical component for knowledge discovery. The storage and processing of images is required in military reconnaissance, medical field, weather forecasting etc. But the huge volume is the main limitation. Like data mining association rules are also applied on the images. For eg: for a grocery store, problem can be to find the number of female customers who are buying cosmetic products as well as diet products. But in case of images, the data is different from what was available for data mining. So image mining is the extension of the data mining area and should not be confused with it. It has the following differences in comparison with data mining:

The image databases focuses on the relative values, like the values of the neighboring pixels whereas the relational databases focus on the absolute values. Spatial information is more important in image databases for processing. Multiple interpretations can be there for the same patterns whereas in relational databases we may have unique interpretations.

III. TECHNIQUES USED IN IMAGE MINING

After the preprocessing of the images, the mining techniques can be applied. The preprocessing lays the foundation for the whole procedure. It improves the quality of the images and also the noise is reduced to normalize the images. Algorithms may be applied to improve the resolution also. For applying the various techniques, the collection of images should be gathered first. The techniques which can be used for the mining of knowledge from the image databases are as follows:

1. Object Recognition: It includes the supervised labeling issue of the objects in an image. This is one of the important



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

tasks in image mining. The correct labels are assigned to the region or to the set of regions identified in the image. Various algorithms can be applied for the identification of the objects. Segmentation techniques can be applied like:

Marker based-in this the object is identified by the area covered by the marker. Contour based-in this contours need to be matched with the pixels.

Object based-in this objects are identified without determining the exact boundaries. Feature Extraction: It focuses on compressing of information which is derived from the objects into set of attributes. The descriptors can be used for the image representation. The descriptors which can be used are colors, shape or texture etc of the objects.

2. Image Classification: It includes the classification or categorization of the objects identified in an image. Supervised classification is main approach or the original approach for the classification of images. Machine learning methods are applied during classification. In contrast to this, image clustering is the unsupervised classification technique. The objects are clustered on the basis of similarity and not on the basis of the labels. 3. Image Indexing and Retrieval: The method of using indexing is not same as that of which is applied on relational databases or for the data mining purposes. Instead of using indexes on the primary key or the secondary keys, the indexing methods used in image mining can be K-D-B tree, R-tree, R+-tree or R*-tree etc. There are two types of indexing which can be followed:

a) Textual or manual indexing which is based on keywords and includes caption indexing, content titles, classification etc. This indexing is more labor intensive and faces inter indexer consistency difficulties.

b) Content-based or automated indexing is based on the attributes of the images. The retrieval technique using this technique is known as CBIR i.e. Content Based Image Retrieval. This technique efficiently accesses the stored images based on the content. The indexing is done with the help of algorithms which distinguish various attributes of the image. The retrieval techniques may include the retrieval based on the metadata, or on the basis of keywords or on the basis of the content which is visually available. The requirements can be classified into the levels. The levels are defined according to the increasing complexity.

a) Level 1 includes the retrieval of images on the basis of features like color, shape or texture.

b) Level 2 includes the retrieval of images on the basis of logical features.

c) Level 3 includes the retrieval on the basis of attributes supporting high level attributes.

There may be text based retrieval which includes the input from the user and the keywords are matched with database. In Query based image retrieval the image is matched on the basis of the visual features with the image in the database.

4. Data Management: Several ways can be there for storing images. Many of the compression formats are also available. Due to these various options, it leads to some problems of maintaining databases for image mining. This area also requires some research attention as the need is to store the images in efficient manner so that retrieval should also not affect the performance levels. Some solutions are provided like object oriented databases or so but still more options are required to be explored.

IV. LITERATURE REVIEW

Wynne Hsu, Mong Li Lee and Ji Zhang[1] explained the image mining and the paper also listed the differences between the relational databases and the image databases. They proposed the two frameworks also for the image mining purpose. The function-driven framework lays stress on the functionalities of different components of the image mining system and the information driven framework focuses on information needs required at various levels of hierarchy. After listing the techniques for image mining, the paper also highlighted the work which needs to be done in this area.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

Peter Stanchev[2] described the method for image retrieval. The paper mainly focused on high level semantic features for image mining purposes. Data mining is through the extraction of low level features like color, shape etc. and then back to their conversion to high level features using fuzzy rules.

P. Rajendran, M.Madheswaran[3] focused on health care or medical image databases. They discussed the technique for data mining which combines the low level features and high level information from the images. The pruned association rule with MARI algorithm is introduced which has positively affected the performance measure.

Hemlatha and Devasana[4] proposed a novel approach using LIM based image matching technique with neural networks for extraction of images in image mining.

Nishchol Mishra et al.[5] Their paper talks about the emergence and the proliferation of social networking sites such as Facebook etc. Analysing the huge amount of data leads to the difficulty in knowledge discovery. The concept of Multimedia Data Mining(MDM) is highlighted which is the process of extracting meaningful information from not only images but also from audio,video etc. It tells that the future of CBIR is related with the advancements made in the image retrieval techniques.

Shaikh Nikhat Fatima[6] discovers the importance of image patterns available in the collection of given images. The main goal is to identify the underprocessed low level features to convert those into processed high level features. The paper focuses on the present of image mining and as well as future of research in image mining area.

Md. Farooque[7] identifies the huge amount of data which is available online. The paper focuses on the content based image retrieval techniques which try to manipulate on the features like shape, color and texture.

Mahip M. Bartere et al.[8] They examined that the performance of the result is related with the users point of view. In the paper the influence of distance function is compared the basis of image color and also mining through the name of the image. Various experiments have been carried out for finding the color similarity between the sample and the resultant image.

V. CONCLUSION

The image mining area is the extension of the data mining area of research. As the performance measure is related with the users requirements, the main aim of image processing should be no loss of inherent information of images while processing. Analysis of different sized and resolution images is of much importance. The higher resolution images are slower to process, as the numbers of objects are more, the extraction process becomes slow. In contrast, objects from lower resolution images are faster to extract but there would be less of discovered knowledge. The paper focuses on various aspects and various techniques of image mining. But still are various issues which need to be resolved. Image mining is the recent research area which is now mainly working on efficient discovery of knowledge from the image databases without loss of the patterns available inherently with the images. Many new algorithms are needed to be developed to improve the performance of the existing algorithms available for data mining. The further research may be carried out for developing the same. And also more work needs to be done to efficiently identify the objects, low level features and high level features in the images for the knowledge discovery. More comprehensive semantically powerful query languages need to be designed for mining data from the images. There should be more efficient indexing and retrieval techniques for mining the relevant information from the image databases.

REFERENCES

1. Wynne Hsu, Mong Li Lee and Ji Zhang , Image Mining: Trends and Developments,Journal Of Intelligent information Systems, Vol. 19, no.1, pp.7-2, Jan 2002.
2. Peter Stanchev Using Image Mining For Image Retrieval IASTED Conf. Computer Science and Technology, Cancun,Mexico,214-218, May 19-21,2003
3. P. Rajendran and M. Madheswaran , An Improved Image Mining Technique For Brain TumourClassification Using Efficient classifier, International Journal Of Computer Science and Information Security, Vol. 6, no. 3, pp. 107-116, Jan 2009.
4. M. Hemalatha and C. Lakshmi Devasena , A Hybrid Image Mining Technique using LIM based Data Mining Algorithm, International Journal Of Computer Applications, Vol. 25, no.2, pp. 1-5,July 2011.



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

5. Nishchol Mishra and Dr. Sanjay Silakari A Perspective Image Mining in the Context of Content Based Image Retrieval IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012.
6. Shaikh Nikhat Fatma , Image Mining Method and Frameworks, International Journal of Computational Engineering Research (ijceronline.com) Vol. Issue. 8, 2012.
7. Md. Farooque, Image Indexing and Retrieval, DRTC Workshop on Digital Libraries: Theory and Practice March DRTC, 2003
8. Mahip M. Bartere and Dr. Prashant R. Deshmukh An Efficient Technique using Text and Content Base Image Mining Technique for Image Retrieval, International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 1, pp. 734-739, Jan-Feb 2012.
9. Shah, V.P. Younan, N.H. Durbha, S.S. King, R.L. Feature Identification via a Combined ICA Wavelet Method for Image Information Mining, Geoscience and Remote Sensing Letters, IEEE issue date: Jan. 2010 Volume: 7 issue: 1 pp: 18-22 issn: 1545-598x
- 10.

BIOGRAPHY

Dr. Rina is an Assistant Professor in the Information Technology Department, GGSD College, Sector 32, Chandigarh, India. She received her PhD degree from Singhania University, Rajasthan in 2013. Her research interests include data mining and e-commerce.

Shailja Agnihotri is an Assistant Professor in the Information Technology Department, GGSD College, Sector 32, Chandigarh, India. She received Master of Science in Information Technology (MSc (IT)) degree in 2005 from Panjab University, Chandigarh, India. Her research interests include data mining, image mining and networks.