**SURVEY PAPER ON IMAGE MINING TECHNIQUES AND CHALLENGES ASSOCIATED**

Mrs Nirmala JSa, Rahul Banerjeeb, Rajath S Bharadwajc

aDepartment of CSE, Assitant Professor, Nitte Meenakshi Institute of Technology,Bengaluru,India

bDepartment of CSE, Research Scholar, Nitte Meenakshi Institute of Technology,Bengaluru,India

cDepartment of CSE, Research Scholar, Nitte Meenakshi Institute of Technology,Bengaluru,India

**ABSTRACT**

In today’s world, multimedia plays a very vital role. Images among them are the most widely used form of media. Image mining is strategically the most complicated yet interesting field of Data Mining. Image mining is an important technique in which we use the images directly to mine data in the form of knowledge. Image segmentation is the first step in image mining. Since Image Mining is more or less an extension of Data Mining, all the concepts of Data Mining work well with Image Mining as well. Image mining works with hidden knowledge extraction, image data association and additional patterns which are dispersed locally in the images. It is an interdisciplinary field that incorporates techniques like computer vision, image processing, data mining, machine learning, data base and artificial intelligence, etc. A crucial task of the mining process is to generate all significant and frequently occurring patterns without prior knowledge of the patterns. This paper presents a survey on various image mining techniques and challenges related to them. Also, this paper provides a marginal overview for future research and improvements.

**Keywords**— Data Mining, Image Mining, Knowledge Discovery, Segmentation, Machine Learning, Artificial Intelligence, Rule Mining, Datasets.

1. **INTRODUCTION**

Images are by far the most widely used Data form [1]. Image Mining comes from a diverse field of Data Mining. Data mining means mining useful or even not useful data/information or knowledge from large databases and information depositories. This has become a field which has caught the attention of many researchers and scholars and a lot of work has been done in this field which is reflected by the research work being carried out in Image mining. It is a highly demanding task but has made good progress in the past years. Image Mining is one of the most used forms of Data Mining. The process of reclaiming images form depositories is called image mining [8]. The researches show advancements in technologies like image digitization, storage and transmission [3]. The number of digital images has become multifold since the last decade. Therefore, content-based image classification and recoup systems have been the major subjects of many multimedia data mining research works in the past few years [2]. The most used features for image description are:

* Color
* Texture
* Shape
* Spatial Features.

Many of the existing image repositories allow users or operators to formulate their queries by complying an example image. The system then identifies those stored pre-existing images in the dataset whose feature values match those of the query most approximately, and displays [4] [5] [6]. Image mining deals with the extraction of steadfast/constant knowledge (hidden knowledge), image relationships or other patterns not explicitly stored in the image. [4]

The main motive of image mining is to outturn all substantial patterns without any knowledge of the image content, the patterns types are different. These can be classification patterns, description patterns, correlation patterns, temporal patterns and spatial patterns. Image mining deals with all physiognomy of huge image databases which includes indexing methods, image storages, and image retrieval, all related to an image mining system.

Image pre-processing is the operation on images at the lowest level of entrancement whose aim is an improvement of the image data that undermine undesired distortions or enhances some image features which might be important for further processing.

This paper presents a survey in the next section on various image mining techniques and challenges to them that were proposed earlier. Also, this paper provides a marginal capsulation for future research and improvements.

1. **LITERATURE REVIEW**

**2.1 IMAGE CHARACTERISTICS/ FEATURES USED FOR MINING**

There are primarily three features of an image that are utilized to extract and store into a database for coordinating with query-request. These are color-shading, shape and surface texture.

1. Color Feature: A computer image is a matrix of pixels which are represented in form of 0s/1s. Each pixel’s value corresponds to the brightness of each analogous point in the scene. Color images are illustrated by three-intensity components. These components generally correspond to red, green and blue (RGB) or sometimes Yellow, Cyan and Magenta (YCM). An integer value can be incorporated in the image with each pixel that can be used as an index to a table that holds the intensity of each color component. The histogram is used to plot the number of pixels with an accurate brightness level against the predefined brightness level. For 8-bit pixels, the brightness ranges from zero (black) to 255 (white). The operations based on color characteristics include histogram normalization, histogram equalization and thresholding.
2. Shape or edge feature: Edge is simply a huge change in frequency. Many techniques of image analysis are based on edges since interpretation based on edge detection is unresponsive to change in the overall radiance level. Edge detection highlights image contrast. Detecting contrast, which is the difference in intensity, can highlight the boundaries of features within an image.
3. Texture feature. Texture is defined as the proximity characteristics of a region or a block. The variation of each pixel with respect to its adjacent pixels defines texture. Texture feature is an important subsided feature in the image, it can be used to describe the contents of an image or a region in addition to color features as just these features are not sufficient to identify the image since different images may have similar histograms.

**2.2 IMAGE MINING FRAMEWORK**

Image processing includes a domain-associated application where we focus on extricating the most important image attributes from information knowledge bank. This is useful in getting to know the collaborative properties of high-level state human impression of images and low-level highlights [3].

Image mining generally refers to the extrication of image data relationships or other blocks of patterns that are not easily visible or recognizable to the user in the images. The dire need to preprocess images in image mining is to upscale the quality of images. These pictures experience different changes so as to grow essential attributes from the pictures. With these created highlights mining is done utilizing data mining methods to investigate noteworthy examples. The then coming about examples are approved and translated to get the last learning, which is connected to applications [7].

As of today, two of the most functional frameworks are as follows: a) Function–Driven structure b) Information-Driven Frameworks. It can be noted that most of the image mining process architectures fall under the category of Function-Driven Frameworks. The goal of Function-Driven Framework is to build and clarify the roles and duties that ought to be carried out throughout image mining. Information-Driven Framework is an effective way of establishing low-level and pixel representation that is contained in a raw image which can be processed to identify high-level objects.

Content-Based Image Retrieval (CBIR) systems are being used in many commercial sites. The primary work of CBIR system is to search for images in an image dataset according to the query passed by the user or the operator. They also aim at searching and going through large image libraries based on few automatically derived image features [15] [18].

There are several ways to approach the image mining framework but the process discussed in the paper is one of the most efficient ways to go by. We can see from the results that advances in image acquisition and storage technology have resulted in tremendous growth in the image mining and processing industry.

The need of tools with proper analysis capabilities is very much important. Systems which can automatically retrieve data are the one with the highest demands. Here in this paper, we are discussing one of such systems which has the capability to do the same.

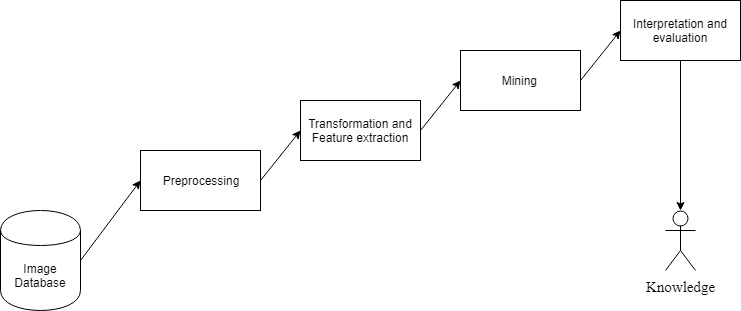
****

Fig: Image Mining Framework

**2.3 IMAGE MINING TECHNIQUES**

Image mining techniques can be broadly divided into object recognition, image indexing and retrieval, image classification and clustering, association rules mining and neural networks [14] [17] [19] [20].

Object Recognition

Object recognition has been a dynamic field of center in regards to picture processing. The process of Object recognition programming identifies the different items from the test picture. This is a stage that prompts the fragmentation of pictures. The key concentration here is to recognize objects in a picture and part a picture into smaller zones which would then be able to dissect and process the pictures independently. This likewise may be alluded to as a controlled labeling issue as per the models of known things.

Image Retrieval

Keeping in mind the end goal to recover pictures from a database, we have to file them legitimately. For that we require a legitimate ordering/indexing programming which does the activity for us. The motivation behind image mining is to recover pictures according to the particular prerequisite of the client. Necessities determination of a client can be named: I) low-level highlights - shading, shape or the spatial area of the picture component ii) high-level state highlights- recognize objects and bigger shapes in the picture iii) Logical highlights – singular items or individual.

Image Indexing

Utilizing an effective ordering plan is fundamental to enhance the image recovery rate and characterization. A portion of the regularly utilized indexing plans are decreasing dimensions or ordering high dimensional data. This may likewise incorporate naming/indexing of pictures as indicated by a specific plan.

Image classification

The motivation behind image clustering is to order pictures in view of key properties that are related with the picture. It might center on the distinguishing purposes of a picture which influences it to emerge from the others. Today image classification is the most looked into area. The most part utilized techniques for image characterizations are a) Supervised classification b) Image grouping

Association Rules Mining

Association mining rules are produced just when the support and confidence are more noteworthy than preset limits. Similar works in two sections: i) In the initial step, all significant item sets that matches the support threshold are distinguished by the rule mining calculation. ii) The second step creates decides that match the confidence edge.

Neural network

Neural systems imply correspondence of a few neurons. These neurons are characterized as straightforward handling units which are isolated into different separated layers with either full or fractional associations set up. The essential undertaking of a neuron is to get the yield that is passed on from the previous neuron and process and produce a yield that will be transmitted to the last neuron. The neurons speak with a few different neighbors keeping in mind the end goal to gather data required for additional preparation.

1. **APPLICATIONS**

Explicitly analyzing images can reveal significant information to the user. Today, the process of image mining is widely used in various fields. Some of the most popular real-world applications are:

1. Diagnosis and categorization of medical conditions like tumors: Systematic diagnosis is important for successful treatment of a tumor. Content based medical image retrieval (CBMIR) can assist the radiologist in the investigation of brain tumor by recovering analogous images from medical image dataset. Magnetic resonance imaging (MRI) is the most frequently used procedure for imaging of tumors [15] [18].
2. Satellite cloud imagery: A Satellite Cloud Image is an image captured from meteorological satellite instruments. It may capture the dissemination of clouds in the outer atmosphere, or maybe to find the weather system and verify the definiteness of surface weather maps that have been drawn/pixelated. It can also be used for inspection/investigation of sea ice distribution to determine the sea surface temperature and the medium-and long-term weather forecast oceanographic data. The satellites revolving around the earth are regularly collecting data in form of images. There are a huge amount images being clicked every second therefore an efficient methodology is required to mine and store the images. Image mining comes of great help when it comes to Satellite Cloud Imagery [25].
3. Natural Scene recognition: Natural scene recognition refers to the procedure where an agent (such as a human being) visually intakes and interprets scenes (images) that it typically encounters in natural modes of operation (e.g. busy streets, meadows, living rooms).This process has been described in several differentiated ways that are guided by different concepts based on preset rules. [23][24]
4. Agriculture field: The image mining concepts can be used in order to detect the diseased stem, leaf, fruit or flowers. It can also be used to define the degree of severity of the diseased part. It can also be used for classification of different kinds of agricultural produce [21].
5. Industrial Work: The mining process can be carried out in the fields of Space Imagery or in the coal industry, or say remote sensing which includes working with different kinds of sensors, mainly the camera sensor which is used for capturing images to be compared with the already existing images in the image repository [22][25].
6. **CHALLENGES**

Although image mining is a boon for human users, it possesses a few challenges considering the fact that it is a relatively new and emerging concept. They are as follows [7] [14] [20]:

1. The automatic classification and clustering of images are still not a cent percent accurate and a lot of work still has to be done in the field of Image Mining.

2. There is a need for unification of a clustering query language which can work with both textual and pictorial data format.

3. World Wide Web (WWW) is a huge database of unlabeled images. Whenever a need of images arises, the retrieval and classification of images become a tedious task for the processor.

4. The need of the hour is to come up with a self-sufficient indexing method which standardizes procedures to index and retrieve knowledge from the indexed images.

5. Image mining has problems related to interfaces which often requires the involvement of both domain and technical experts. This might lead to wasteful time consumption and may also put extra load on the operators/experts [14].

6. Real-life images are not clean; there might be a lot of added external noise, or maybe the image can be over-exposed or under-exposed, or maybe the image is defocused [7].

7. There is a very rare scope of the software for Image Mining being reused as the software is mainly designed for specific use or utility [3].

8. Bottleneck maintenance is not such a feasible process therefore work has to be put in, in order to make advancements in the same [22].

9. The researchers might solve the theoretical problem aspects but the industry demands issues to be solved related to the practical aspects [22].

10. There is still a gap in the client-server technology when it comes to Image Mining and this can be solved using other techniques like NLP, Machine Learning, and AI etc. [3][4]

11. To distinguish influenced locale from woods fire, fire blazes, fire conduct from wind course, an expectation of flame spread[21].

**5. FUTURE SCOPES**

Summarizing, Image mining is an exciting field for examination and introspection. Image mining research is still in its early stages and numerous issues stay inexplicable. In particular, image mining research in order to advance to another stature, the accompanying issues should be explored at a future scope [7][14]:

(a) Propose new portrayal plans for visual examples that can encode adequate relevant data to take into account significant extraction of helpful visual qualities;

(b) Devise productive substance based image ordering and recovery methods to encourage quick and successful access to vast image databases;

(c) Design semantically capable question dialects for image databases;

(d) Explore new revelation systems that consider the exceptional attributes of picture information;

(e) Incorporate new representation systems for the perception of image patterns.

1. **CONCLUSION**

Image mining is an advanced field that falls into the broader field of data mining techniques. The ulterior motive of image mining is to retrieve meaningful information hidden in images in accordance with the needs of the user. The intent of this paper is to understand the concept of image mining and its process, techniques, applications and challenges that are explained in the paper. Important information or knowledge can be hidden implicitly in the images; contrariwise, few researchers talk about data mining on them. Image segmentation is the fundamental phase of image mining. In other words, image mining is simply an amplification of data mining in the field of image processing. Image mining handles with the hidden knowledge extraction which is the detection of implicit characteristics in an image, image data association which is the accumulation of similar data attributes and added patterns which are not clearly concentrated in the images. The notion of image mining is beneficial to human users. However, there is huge scope for further research to be carried out in this emerging field of data science.

**6. REFERENCES**

[1] Dr. S. Vijayarani and Ms. A. Sakila, "Multimedia Mining Research – An Overview," in International Journal of Computer Graphics & Animation (IJCGA) Vol.5, No.1, January 2015.

[2] A.Hema, E. Annasaro, "A Survey In Need Of Image Mining Techniques," in International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 2, February 2013.

[3] A. A. Khodaskar, S .A. Ladhake “Image Mining: An Overview of Current Research,” in Fourth International Conference on Communication Systems and Network Technologies, IEEE 2014

[4] T.Karthikeyan, P.Manikandaprabhu, "Function and Information Driven Frameworks for Image Mining - A Review" in International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2013.

[5] Deepika Kishor Nagthane, “Image Mining Techniques and Applications”, in International Journal Of Engineering Sciences & Research Technology July, 2013.

[6] K. R. Yasodha, K.S. Yuvaraj, “A Study on Image Mining Techniques”, In International Journal of Applied, December 2013.

[7] Ji Zhang, Wynne Hsu, Mong Li Lee, “Image Mining: Trends and Developments”, International Workshop on Multimedia Data Mining, 2001.

[8] C. Ordonez and E. Omiecinski (1999), “Discovering association rules based on image content,” Proceedings of the IEEE Advances in Digital Libraries Conference(ADL’99)

[9] Itten, J., Kunst der Farbe. Ravensburg, Otto Maier Verlag. 1961.

[10] Mira p., Jesse j., Laurence W., Fast Content-Based Image Retrieval Using Quasi-Gabor Filter and Reduction of Image Feature Dimension.

[11] Mori K., Wada K., Toraichi K., Function Approximated Shape Representation using Dynamic Programing with Multi-Resolution analysis ,ICSPAT’99,1999.

[12] Peter Stanchev, “Using Image Mining for Image Retrieval,” . [www.kettering.edu/~pstanche](http://www.kettering.edu/~pstanche).

[13] J. Fernandez, N. Miranda, R. Guerrero and F. Piccoli (2007), “Appling Parallelism in Image Mining,” www.ing.unp.edu.ar/wicc2007/trabajos/PDP/120.pdf.

[14] Wynne Hsu, Mong Li Lee and Ji Zhang (2002), “Image Mining: Trends and Developments,” Journal of Intelligent Information Systems, vol. 19, no. 1, pp. 7-23.

[15] Abhi Gholap, Gauri Naik, Aparna Joshi and CVK Rao (2005),“Content-Based Tissue Image Mining”, IEEE Computational Systems Bioinformatics Conference - (CSBW’05), pp.359-363.

[16] Sanjay T. Gandhe, K. T. Talele and Avinash G. Keskar (2007), “Image Mining Using Wavelet Transform”, Knowledge-Based Intelligent Information and Engineering Systems, Springer link book chapter, pp. 797-803.

[17] Sabyasachi Pattnaik, Pranab Kumar Das Gupta and Manojranjan Nayak (2008), “Mining images using clustering and data compressing techniques”, International Journal of Information and Communication Technology, vol. 1, no. 2, pp. 131-147.

[18] Petra Perner (2002), “Image mining: issues, framework, a generic tool and its application to medicalimage diagnosis”, Engineering Applications of Artificial Intelligence, vol. 15, no. 2, pp. 205-216.

[19] Lu Kun-Che and Yang Don-Lin (2009), “Image Processing and Image Mining using Decision Trees”, Journal of information science and engineering, vol. 25, no. 4, pp. 989-1003.

[20] L. Jaba Sheela and V. Shanthi (2007), “Image Mining Techniques for Classification and Segmentation of Brain MRI Data,” Journal of Theoretical and Applied Information Technology,” vol. 3, no. 4, pp. 115-121.

[21] Divya T.L, Vijayalakshmi M.N, “Analysis of Wild Fire Behaviour in Wild conservation Area using Image Data mining”,,IEEE,2015.

[22] K. Nienhaus, and F. Mavroudis: Automation and Process Monitoring in the Mining Industry by Infrared Sensor Technology and Industrial Image Processing. MPES 2009

[23] A. Yilmaz, O. Javed, and M. Shah, “Object tracking: A survey,” Acm computing surveys (CSUR), vol. 38, no. 4, p. 13, 2006.

[24] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06), vol. 2. IEEE, 2006, pp. 2169–2178.

[25] Liu. Y.. Hu. J.. Snell-Feikema . I.. VanBemmel. M. S.. Lamsal. A.. & Wimberly. M. C. (2015 ). Software to facilitate remote sensing data access. Environmental Modelling & Software. 74(December ). 247-257