IMAGE MINING TECHNIQUES: A LITERATURE SURVEY

Mahendra S.Makesar[1]
Asstt. Professor
Dept . of Computer Science and Engineering
Prof .Ram Meghe College Of Engineering,Badnera
mahen_mak@yahoo.co.in

Dr.Nitin A. Koli[2]
Head , Computer Centre
Sant Gadge Baba University,Amravati
nitinkoli@sgbau.ac.in

Ratnashil N.Khobragade[3]
Asstt.Professor,
Sant Gadge Baba University ,Amravati
rnkhobragade@gmail.com

Abstract:

Nowadays, a large portion of information is in visual form; it is essential and certainly pleasing to search for images. Image mining has a variety of applications in various sectors like medical diagnosis, biology, remote sensing, space research, etc.Image mining is a class of analytical techniques that examine a large amount of image data. It has proved highly effective in addressing many important business problems. Comparison among different mining by similarity systems is particularly challenging owing to the great variety of methods implemented to represent likeness and the dependence that the results present of the used image set.

This paper presents a survey on various image mining techniques.It also provides a improvements for future research.

Keywords: ***Classification, Pre processing, Segmentation, Feature Extraction, Image Mining,***

1.INTRODUCTION

Advancements in image acquirement and storage technology have led to implausible growth in very large and detailed image databases.These image if analyzed, can expose useful information to the human users. Image mining deals with the extracting inherent and embedded knowledge, image data relationship, or other patterns which is not explicitly found in the images. Image mining is an interdisciplinary challenge that draws upon proficiency in computer vision, digital image processing, image extraction, data mining, machine learning, databases and artificial intelligence.

The fundamental challenge in image mining is to reveal out how low level pixel representation enclosed in a raw image or image sequence can be processed to recognize high level image objects and relationship. Image mining normally deals with the study and developments of new technologies that allow accomplishing the object. Image mining technique deals with the extraction of implicit knowledge and image with data relationship or other patterns not explicitly stored in the images. The image mining is highly specific because the image databases are predominantly non relational. In addition many image attributes are not directly visible to the user. In the past, image mining have been used to solve various problems including target recognition, object recognition, face recognition,and face detection/verification. The objective of this paper is to review image mining techniques,so that the set of these techniques can be appreciated. It will help in image mining using genetic algorithm.

2. LITERATURE REVIEW

The literature survey carried out related to technology impact in the study of different image mining techniques. Z Sun et, al[1] argued that feature selection is an important problem in object detection and demonstrate that genetic algorithms (GAs) provide a simple, general, and powerful framework for selecting good subsets of features, leading to improved detection rates. but they have not generalize the encoding scheme to allow Eigen vector fusion.

Nicolay Y. Nikolaev and Hitoshi Iba[2] presented an approach to regularization of inductive genetic programming tuned for learning polynomials.The objective is to achieve optimal evolutionary performance

when searching high-order multivariate polynomials represented as tree structures.This approach will lose slightly in speed of computation.

Image mining is more than just an expansion of data mining to image domain.Wynne Hsu,mong Lee and Ji Zhang[3] examined the research issues in image mining,development in image mining. They proposed an information driven framework for image mining.In that they made out four levels of information: Pixel level, Object level, semantic concept level, pattern and knowledge level. This approach does not addressed noise redundancy.

Aura Conci.,Everest Mathias, M.Castro[4] proposed a framework for mining images by colour content.Their framework provides the possibility of use 5 distance function for evaluation of similarity among images and 2 types of quantization. The procedure here presented considers only retrieval aspects. Considerations like complexity or time performance are not treated here. The ideas presented are only a small step in a very rich research direction. Others visual features such as texture, shape, and use of compressed images can be identified for further extension of this problem.

Peter Stanchev[5] proposed a new method for image retrieval using high level semantic features.It is based on removal of low level color system which uses image mining approaches.In this approach full system realization is not obtained.

David A. Fay et al. [6] have developed a system for multisensor image fusion and interactive mining based on neural models of color vision processing,learning and pattern recognition. They included add-on modules for Image Conditioning, Image Fusion, extraction of Context Features, and interactive Image Mining. Together, these modules create a work flow enabling a user to create vector products of foundation features (e.g., roads, rivers,and forests) and highlighted target detections from raw multisensor or multispectral imagery. In this Image mining based on multispectral imagery modified by simulated environmental conditions not addressed.

Herbert Daschiel and Mihai Datcu[7] demonstrated the concepts of a prototype of a Knowledge-driven content-based information mining system produced to manage and explore large volumes of remote sensing image data. The system consists of a computationally intensive offline part and an online interface. The offline part aims at the extraction of primitive image features, their compression, and data reduction, the generation of a completely unsupervised image content-index, and the ingestion of the catalogue entry in the database management system. This approach does not addressed the use of knowledge and semantic information that is stored in the database system.

R.Brown ,B.Pham[8] described in detail a general hierarchical image classifier approach and illustrated with which it can trained to find objects using support vector machine concept. In this approach speed and time complexity of algorithm is not discussed.

Ross Brown et al.[9] designed a image mining system for digital forensic.They proposed to use a baysian networks approach to deal with information uncertainties which are inherent in forensic work.These inference network will be constructed to model probability interactions between beliefs,adapt to different users,retrieval patterns and mimic human judgement of semantic content of image patches.In this approach uncertainties occur in image characteristics,object description,co-occurrence of objects and human semantic interpretation of image content.

Sanjay T.Gandhe,Avinash G.Keskar[10] put forth an image mining technique using wavelet transform.The author proposed an image mining approach using wavelet transform.It uses common pattern identification and data mining models.They have constructed a protoype system for identification using DWT +PCA system. Handling large number of images at the source machine is a crucial task and hence gives rise to memory management issue.The database should be so efficient to handle images efficiently.

Lionel Gueguen and Mihai Datcu [11] addressed the problem of extracting relevant information from Satellite image time series (SITS) based on the information-bottleneck principle. The method depends on suitable model selection, coupled with a rate–distortion analysis for determining the optimal number of clusters. They presented how to use this method with the Gauss–Markov random fields and the auto binomial random fields model families in order to characterize the spatio-temporal structures contained in SITS. In this approach spectral or geometrical information was not taken into account.

Aksoy and Cinbis[12] proposed a new image mining technique using directional spatial constraints.Retrieval performance was evaluated by the author using precision and recall using a ground truth that was constructed by manually identifying the objects satisfying each query.It needs manual interpretation to improve processing time.

P.Rajendran,M.Madheswaran[13] discussed an image mining technique.It combines low level features extracted from images and high level knowledge from specialist.It does not addressed feature redundancy,image noise and time complexity.

Dr.Sanjay Silkari,Dr.Mahesh Motwani and Maheshwari[14] proposed a framework focuses on color as feature using color Moment Block Truncation Coding(BTC). To extract feature for image dataset.Then K-Means clustering algorithm is conducted to group the image dataset into various clusters. This approach does not addressed time and speed of the algorithm.

Tao Jiang and Tan[15] proposed two methods for discovering the underlying associations between text and images.The first method based on transformation measures the information similarity between visual features and textual features.Another method uses a neural network to learn direct mapping between visual and text features by incrementally summarizing associated features into a set of information template.It needs to perform batch learning on a fixed set of training data.

Maybin Muyeba, et al.[16] proposed a system for image mining using fuzzy rule. It relates the property of composite attributes.They partitioned the property value into fuzzy property sets. In this approach fuzzy measures and correlation association is not described.

Surya S. Durbha, et al.[17] proposed feature selection and feature transformation based on wrapper based approach.They adopted region based system..They have argued that selecting a relevant feature subset increases the rate of correctly identifying a semantic class.This system currently used imagery from only one sensor.

Shah,V.P et al.[18] presented a new feature set,obtained by integrating independent component analysis and wavelet transformation for image information mining in geospatial data.The recall accuracy of the water bodies is least because few of the smaller objects are missed during the coarse level segmentation. The objects that occupy areas less than 8*8 pixels may not be recalled correctly.

Dubey[19] illustrated about an image mining methods which is dependent on the color Histogram, texture of that image. The query image is considered,then the COLOR Histogram and Texture is created and in accordance with this the resultant image is found.In this approach computing time for RGB color space not considered.

Victor and Peter[20] put forth a new minimum spanning tree based clustering algorithm for image mining. The minimum spanning tree clustering algorithm is proficient of detecting clusters with irregular boundaries.The author presented a minimum spanning tree depending on the clustering technique using weighted Euclidean distance for edges,which is vital constituent in constructing the graph from image. The technique construct ‘k’ clusters with segments.In this approach time complexity of the algorithm is not discussed.

Jusong Yuan et al.[21] addresses the classification problem by mining informative features derived from image contents and spatio temporal traces of GPS coordinates that characterize the underlying movement patterns of various event types, both based on the entire collection as opposed to individual photos.How to apply the discovered visual collocations for image search and categorization is not discussed in this approach.

Hemlatha & Devasana[22] proposed a research to find out the accurate image while mining an multimedia image database and developed a technique for mining images by means of LIM dependent image matching method with neural network.In this approach time complexity of the algorithm is not discussed.

Andrea Julea et al.[23] proposed a method to extract, in an unsupervised way, temporal evolutions at the pixel level and select those covering at least a minimum surface and having a high connectivity measure. To manage the huge amount of data and the large number of potential temporal evolutions, a new approach based on data-mining techniques has been presented. The large number of pixels are covered by these patterns but the level of purity is not high.

Ja-Hwung Su et al.[24] proposed novel method,navigation pattern relevance feedback(NPRF) to achieve high efficiency and effectiveness CBIR. By using NPRF high quality of image retrieval on RF can be achieved in a small number of feedbacks.The negative examples existing at each feedback are all skipped.

Noorhaniza Wahid[25] proposed a swarm based algorithm for classification.He compared swarm based technique with support vector machine and obtained the result.PSO needs to allocate more memory than SSO for each particle to achieve better performance.

From the survey of the available literature, following open issues are identified that need further investigations.

i) Time complexity of algorithm.
ii) Redundancy of features .
iii) Application of optimizing technique for image mining.
iv) Full system realization.
v) Multispectral imagery modified by simulated environmental conditions.

3. METHODOLOGY:

The existing methodologies consists of following general steps.

\Step1: **Pre-processing**

The image data is highly non-trivial. reprocessing phase is applied to remove noise from the image. And it also consist of image segmentation. By applying noise removal filter,thinning,cleaning noise can be removed. Pre-processing phase also include object identification.Preprocessing determines the effectiveness of image mining application.

Step2: **Feature Extraction**.

Features,characteristics of the objects of interest, if selected carefully,are representative of the maximum relevant information that the image has to offer for a complete characterization of the lesion

Step 3 : **Feature selection.**

Feature selection helps to reduce the feature space that improves the prediction accuracy.

Step 4 : **Classification.**

Classification process involves two phases: training and testing phase.

4. Conclusion.

This paper presents a survey on various image mining techniques that was proposed earlier by researcher. This overview of image mining focuses on image mining implementations, usability and challenges. It also delivers conceptual overview of methodology. Image mining is an expansion of data mining in the field of image processing. Future investigations that are discussed may be implemented in the area of image mining.

5. REFERENCES.

[1] Z. Sun, G. Bebis, R. Miller, "Object detection using feature subset selection,"*Pattern Recognition 372165,March 2004.*

[2] Nikolay Y. Nikolaev, Hitoshi Iba," Regularization Approach to inductive genetic programming", *IEEE transactions on Evolutionary Computation, Vol.5,no.4,August 2001*

[3] 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.*

[4] Aura Conci and Everest Mathias M. M. Castro , "Image mining by content", *Expert Systems With Applications, Vol. 23, no. 4, pp. 377-383,April 2002.*

[5] Peter Stanchev "Using Image Mining For Image Retrieval*" IASTED Conf. Computer Science and Technology,Cancum,Mexico,214-218, May 19-21,2003*

[6] David A. Fay, Richard T. Ivey, Neil Bomberger, and Allen M. Waxm, "Image Fusion & Mining Tools for a COTS Environment", *ISIF pp. 606-613,2003.*

[7] Herbert Daschiel and Mihai Datcu, " Information Mining in Remote Sensing Image Archives: System Evaluation", *IEEE Transactions On Geoscience and Remote Sensing, Vol 43, no. 1, January 2005.*

[8] R. Brown, B. Pham "Image Mining and Retrieval Using Hierarchical Support Vector Machines" , IEEE *Proceedings Of the 11th International Multimedia Modelling Conference, 1550-5502,2005.*

[9] Ross Brown, Binh Pham, and Olivier de Vel **, "**Design of a Digital Forensics Image Mining System", *Springer-Verlag Berlin Heidelberg, pp. 395.404 Sept 2005.*

[10] Sanjay T.Gandhe, K. T. Talele and Avinash G. Keskar , "Image Mining Using WaveletTransform", *Knowledg-Based Intelligent Information and Engineering Systems, Springer link bookchapter, pp. 797-803,2007.*

[11] Lionel Gueguen and Mihai Datcu, " Image Time-Series Data Mining Based on the Information-Bottleneck Principle", *IEEE Transactions On Geoscience and Remote Sensing, Vol. 45, no. 4,827, April 2007.*

[12] Selim Aksoy and R. Gokberk Cinbi , "Image Mining Using Directional Spatial Constraints,"*IEEE Ggeosciences and Rremote Sensing letters, vol. 7, no. 1, pp. 33-37,Aug 2009.*

[13] 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.*

[14] Sanjay Silakari, Mahesh Motwani and Manish Maheshwari , "Color Image Clustering using Block Truncation Algorithm",*Ijcsi International Journal Of Computer Science issues, Vol. 4, no. 2,Aug 2009*

[15] Tao Jiang and Ah-Hwee Tan,Learning, "Image-Text Associations", *IEEE Transactions On Knowledge and Data Engineering, Vol. 21, no. 2, February 2009*

[16] Maybin Muyeba, M. Sulaiman Khan, and Frans Coene, "A Framework for Mining Fuzzy Association Rules from Composite Items", *Springer-Verlag Berlin Heidelberg, pp. 62–74, May 2009.*

[17] Surya S. Durbha, , Roger L. King, and Nicolas H. Younan, "Wrapper-Based Feature Subset Selection for Rapid Image Information Mining" , *IEEE GeoScience and Remote Sensing letters, vol. 7, no. 1, Janaury 2010.*

[18] 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*

[19] Rajshree S. Dubey , "Image Mining using Content Based Image Retrieval System", *(IJCSE)*International Journal On Computer Science and Engineering Vol. 02, no. 07, pp. 2353-2356, October 2010.

[20] S. P. Victor and S. John Peter , "A Novel Minimum Spanning Tree Based Clustering Algorithm for Image Mining," *European Journal Of Scientific Research, vol. 40, no. 4, pp. 540-546,March 2010.*

[21] Junsong Yuan, Jiebo Luo, and Ying Wu, " Mining Compositional Features From GPS and Visual Cues for Event Recognition in Photo Collections", *IEEE Transactions On Multimedia, vol. 12, no. 7, November 2010.*

[22] 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.*

[23] Andreea Julea, Nicolas Méger, Philippe Bolon, Christophe Rigotti,Marie-Pierre Doin, Cécile Lasserre, Emmanuel Trouvé, , and Vasile N. L˘az˘arescu, "Unsupervised Spatiotemporal Mining of Satellite Image Time Series Using Grouped Frequent Sequential Patterns", *IEEE Transactions On Geoscience and Remote Sensing, Vol. 49, no. 4, April 2011.*

[24] Ja-Hwung Su, Wei-Jyun Huang, Philip S. Yu, , and Vincent S. Tseng, " Efficient Relevance Feedback for Content-Based Image Retrievalby Mining User Navigation Patterns", *IEEE Transactions On Knowledge and Data Engineering, vol. 23, no. 3,March 2011.*

[25] Noorhaniza Wahid **, "**A Novel Image Classification Algorithm Using Swarm-Based Technique for Image Database", *Springer-Verla-Berlin Heidelberg, pp. 460–470, April,2011*