View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by MAT Journals

Journal of Image Processing and Artificial Intelligence
Volume 3 Issue 2

# Image Analysis and Image Mining Techniques: A Review

***Vikramsingh R. Parihar[1], Roshani S. Nage[2], Atul S. Dahane[3]***
[1,2,3] Department of Electrical and Electronics Engineering, Prof Ram Meghe College of Engineering and Management, Badnera- Amravati, India.
Email: vikramparihar05@gmail.com

***Abstract***

*This paper presents the analysis of existing literature which is relevant to mining and the mechanisms associated with weighted substructure. Though, the literature consists of a lot many research contributions, but, here, we have analysed around thirty-five research and review papers. The existing approaches are categorized based on the basic concepts involved in the mechanisms. The emphasis is on the concept used by the concerned authors, the database used for experimentations and the performance evaluation parameters. Their claims are also highlighted. Our findings from the exhaustive literature review are mentioned along with the identified problems. This paper is useful for comparative study of various approaches which is prerequisite for solving image mining problem.*

***Keywords:*** *Image Analysis, Image Mining, Weighted Substructure*

## INTRODUCTION

Image analysis is the process of retrieving data or useful information from digital images by using digital image processing techniques. Image analysis can include such tasks as finding shapes, detecting edges, counting objects, or measuring properties of an object. Image mining normally deals with the study and developments of new technologies that allows image analysis. Image mining is, thus, an extended branch of data mining that is concerned with the process of knowledge discovery concerning digital images. It is concerned with the extraction of implied knowledge and image with data relationship or other patterns which are not overtly stored in the images.

Fig. 1 shows the general image mining process. The quality of database images is first improved by suitable preprocessing. The important features of the images are generated by various transformations and feature extraction processes. Using data mining techniques mining can be carried out with the aid of generated features, to discover significant patterns. The resulting patterns are interpreted to obtain the final knowledge, which can be applied to further applications.

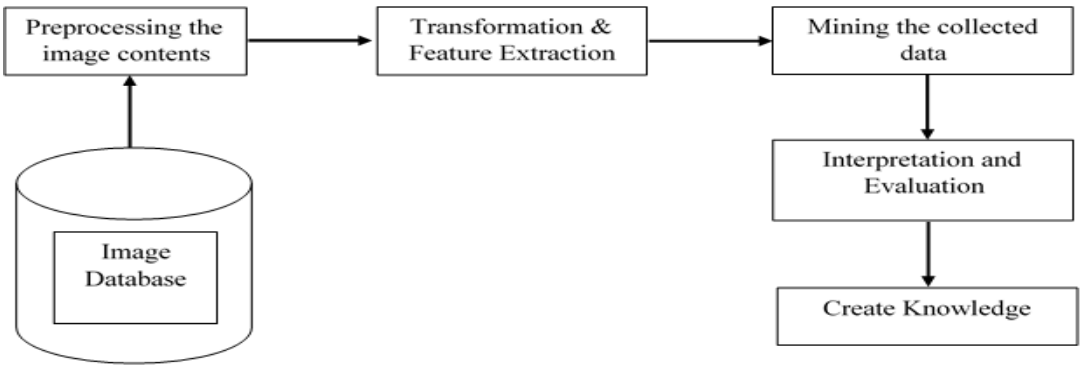


Fig. 1: General image mining process


Page 1-18 © MAT Journals 2017. All Rights Reserved

The problem, here, is modeled in terms of creating a dataset of images and extracting the features of each image. Then graphs are then generated for each image based on these features. The weight factor is used to measure the actual importance of each different substructure in a given graph dataset. On the basis of weighted substructure graphs the image mining process is done. For image mining, an external query image is to be provided by user. Its features are extracted and graph is generated. Later the substructure of query image is matched with the substructure of the dataset. The most closely matched substructures of images from the dataset are identified and it can be concluded that the identified images are close to the query image.

## LITERATURE REVIEW

### Image Analysis Related Work

In the year 2007, Hong Cheng, Xifeng Yan, Jiawei Han and Chih-Wei Hsu [1], introduced a feature selection algorithm i.e. Maximal Marginal Relevance Feature Selection (MMRFS) algorithm to select discriminative frequent patterns. The method is able to overcome two kinds of overfitting problems and is shown to be scalable. They claim that the frequent patterns are high quality features and have good model generalization ability. The framework is also applicable to more complex patterns, including sequences and graphs. They used performance evaluation parameters like accuracy and time and carried out experimentations on Standard database. It is identified that feature selection algorithm takes more time to execute.

Bryan C. Russell, Alexei A. Efros, Josef Sivic, William T. Freeman and Andrew Zisserman develop an algorithm using multiple segmentations to discover objects and their extent in image collections [2]. They used the performance evaluation parameters of average precisions. They carried out experimentations on Cal tech, MSRC and LabelMe database. It is identified that the visual quality of images is not good due to overlapping of segment.

In year 2003, J.C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang [3] introduced a approach which applies the EMD to texture extraction and image filtering. They developed an algorithm based on bidimensional empirical mode decomposition (BEMD) to extract features at multiple scales or spatial frequencies. They carry out the experimentation on synthetic and natural images and used the performance evaluation parameter of frequency components. They claim that results are good for both synthetic and natural images. EMD gives a new and promising way to decompose and extract texture features without parameter. However, their work does not concentrate on entire images at a time.

G. Dougherty, in year 2010, worked on image analysis algorithms and visualization [4]. He used the performance evaluation parameters of true positive fraction (TPF), false positive fraction (FPF) and showed that all mentioned applications0 which are useful for image analysis.

In the year 2003, Pew-Thian Yap, Raveendran Paramesran [5] proved that the Krawtchouk moments can be employed to extract local features of an images. Krawtchouk moments are a set of moments which formed by using Krawtchouk polynomials called as the basis function set. They carried out the experimentation on Uppercase letters and Lowercase letters. They used noise, density, Hu, Krawtchouk as the performance evaluation parameters for their approach. They claim that in object recognition experiment, the results show Krawtchouk moment invariants perform significantly better than Hu's moment


Page 1-18 © MAT Journals 2017. All Rights Reserved

invariant. And also perform better in terms of reconstruction error. However, computational time evaluation is not presented.

Xiaoguang Lu [6], introduced the model-based approaches including elastic bunch graph matching, active appearance model and 3D morphable model methods. He also provided appearance-based method and model-based face recognition method. Appearance-based face recognition method is applicable to only low resolution or poor quality images. The model-based method has intrinsic physical relationship with real faces. An explicit modeling of face variations, such as pose, illumination and expression, gives the possibility to handle these variabilities in practice. He used performance evaluation parameter like error rate, recognition accuracy, Computation (floating-point operations).

The systematic representation of above image analysis related work is given in Table 2.1.

***Table 2.1.*** *Image analysis related work*

| **Sr No** | **Ref. no. Concerned Author(s) and year** | **Concept used** | **Performance evaluation parameter** | **Database used** | **Claims by Concerned Author(s)** | **Our findings** |
|---|---|---|---|---|---|---|
| 01 | [1] Hong Cheng, Xifeng Yan et al. 2007 | A feature selection algorithm is introduced to select discriminative frequent patterns. | Accuracy, Time. | Standard | Frequent patterns are high quality features and have good model generalization ability. The framework is also applicable to more complex patterns, including sequences and graphs. | Feature selection algorithm takes more time to execute. |
| 02 | [2] Bryan Russell Alexei Efros et al. 2006. | Developed an algorithm using multiple segmentations to discover objects and their extent in image collections | Average precisions. | Cal tech, MSRC and LabelMe | Their results are computed from a small set of single object category images. | Visual quality of images is not good due to overlapping of segment. |
| 03 | [3] J.C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang. 2003 | The EMD is applied to texture extraction and image filtering, then an algorithm is developed based on bidimensional | Frequency components. | Synthetic and natural images | Result are good for both synthetic and natural images. EMD gives a new and promising way to decompose and extract texture features | The mentioned method does not concentrate on entire images at a time. |


Page 1-18 © MAT Journals 2017. All Rights Reserved

| | | | | | | |
|---|---|---|---|---|---|---|
| | | empirical mode decomposition to extract features at spatial frequencies | | | without parameter. | |
| 04 | [4] Dougherty G. 2010 | Image analysis algorithms and visualization. | True positive fraction, False positive fraction. | Not mentioned | All application which are mentioned in paper are useful for image analysis. | Requires more user interaction |
| 05 | [5] Pew-Thian Yap, Raveendran Paramesran. 2003 | Image analysis by Krawtchouk Moments. The Krawtchouk moments can be employed to extract local features of an images. | Noise, density. | Standard | In object recognition experiment, the results show Krawtchouk moment invariants perform significantly better than Hu's moment invariant. And also perform better in terms of reconstruction error. | Require more computational time as time evaluation is not presented. |

**Mining fundamental related work**

In the year 2013, A.Hema and E. Annasaro [7] carried out a survey in need of image mining and techniques. They provide various techniques used for image mining applications like image retrieval, matching, pattern recognition etc. However, they fail to mention the details regarding each technique separately and addressed the existing methods in a general manner.

Patricia G. Foschi, Deepak Kolippakkam, et al. introduced the concept of extract patterns and derive knowledge from large collections of images, and also extract unique features for the images [8]. They used the parameters of recall and precision for comparison with other approaches. They claim that the features used are sufficient to identify the patterns from the images and tested images required less human interaction for resolving the uncertain regions. However, the detection process is carried out manually, which is laborious and time consuming.

Piotr Dollar Zhuowen Tu et al. [9] used feature mining paradigm for image classification and examine several feature mining strategies. They used the performance evaluation parameters like True positive (TP), False positive (FP), computation time. They carried out experimentations on MIT CBCL face data-set. They claim that the proposed approach provides effective result for feature mining. However, computational complexity of the algorithm seems to be more as time estimation is not presented.

Till Quack, Vittorio Ferrari, and Luc Van


Page 1-18 © MAT Journals 2017. All Rights Reserved

Gool [10] presented a method for frequently occurring mining objects and scenes from videos. Video data mining is based on the detection of recurring spatial arrangements of local features; these features are represented in quantized codebooks. They claim that their approach is suitable and efficient tool for video mining. However, the mentioned approach fails in high resolution video mining technique and performance evaluation parameter is not given.

Josef Sivic and Andrew Zisserman worked on video data mining using configurations of viewpoint invariant regions [11] and also work on extract significant objects, characters and scenes in a video by determining the frequency of occurrence of spatial configurations. They used the parameters of precision and recall for comparison with other approaches. They claimed that textured region provides good results for video mining. However, they discussed only about textured regions and other regions are missed.

In the year 2010, A. Kannan, Dr. V. Mohan, Dr. N. Anbazhagan [12] worked on image retrieval technique based on optimum clusters for improving user interaction with image retrieval systems by fully exploiting the similarity information. In content based image retrieval (CBIR) image mining process can be done in following steps: The images from an image database are first preprocessed to improve their quality. These images then go through various transformations and feature extraction to generate the important features from the images. With the generated features, mining can be carried out using data mining techniques to discover significant patterns. The resulting patterns are evaluated and interpreted to obtain the final knowledge, which can be applied to applications. They used performance evaluation parameter as precision and recall. They claim that the the proposed method is very useful for image mining. By using this method one can easily find out the similar images related to query image and cbir provides good result. However, in CBIR the features of the query image alone are considered; database is also not given.

Ji Zhang, Wynne Hsu, Mong Li Lee [13] carried out a survey of image mining which provides the research issues in image mining, current developments in image mining, image mining frameworks, state of the art techniques and systems. However, they fail to mention the details regarding each approach separately and addressed the existing methods in a general manner.

In the year 2013, Gangin Lee and Unil Yun [14] show that mining of strongly correlated subgraph patterns is done by considering weight and support constraints. Weighted frequent graph mining is an approach for applying importance of objects in the real world to the graph mining. They used runtime, minimum threshold as performance evaluation parameter and carry out experimentations on PTE dataset and DTP dataset. They claim that MSCG algorithm not only mined meaningful patterns with strong correlation among the elements in sub-graphs efficiently but also have advanced performance as reducing search space by the mentioned measure. However, complexity is more and time required for the algorithm to execute seems to be more.

In [15], R. Brown, B. Pham in the year 2005, proposed a general hierarchical image classifier approach, and illustrates the ease with which it can be trained to find objects in a scene. This application area required a method which allowed the detection of objects trained from image patches, and also enabled arrangements of component detectors for image mining.


Page 1-18 © MAT Journals 2017. All Rights Reserved

The performance evaluation parameter used are Positive count, Negative count, False positive count (FPC), False negative count (FNC). They carried out their experimentations on Biology and Vietnamese art images. However, features of images are not mentioned.

In the year 2003, Mihai Datcu, Herbert Daschiel, Andrea Pelizzari, Marco Quartulli, et.al. [16] introduced a concept of a prototype of a knowledge driven content-based information mining system produced to manage and explore large volumes of remote sensing image data. They used performance evaluation parameter as size, resolution and signal models and carry out their experimentations on EU SC, NERSC database. They claimed that the presented approach is KIM, which is a prototype of a next-generation knowledge-driven image information mining system developed for the exploration of large image archives. However, computational complexity of the algorithm seems to be more as time estimation is not presented.

In the year 2011, C. Lakshmi Devasena, M. Hemalatha worked on new method for detecting images [17], and used the method Lorenz Information Measure (LIM) for representing features extracting from the images for retrieval. Their mentioned approach attempts to make the process mostly independent of any parameter setting to generate a robust solution. They used sample test, image database used, time as performance evaluation parameter. They claimed that the image mining system derived from the LIM based image matching technique provided promising results, which are comparable and significant. However, time required for the algorithm to execute seems to be more.

Basura Fernando, Elisa Fromont and Tinne Tuytelaars [18], introduced an effective method for image classification using itemset mining to discover a new set of mid-level features called Frequent Local Histograms (FLH). They used the parameters like Classification accuracy, equal error rate, mean average precision for performance evaluation. And carry out their experimentation on GRAZ-01, Oxford-Flowers, 15-Scenes and the PASCAL-VOC2007 dataset. However, the differences between the results of mentioned approach and other methods are not very obvious in case of the images which contain a large amount of edge details.

Gosta Grahne and Jianfei Zhu [19], proposed a novel array based technique that greatly reduces the need to traverse FP-trees and also present new algorithms for a number of common data mining problems. Also present an algorithm, FPclose, for mining closed frequent item sets. FPclose uses yet another variation of the FP-tree structure for checking the closeness of frequent item sets. They carry out the experimentation on sparse, synthetic and real datasets and used the performance evaluation parameter CPU Time(s), Minimum Support (%), Main Memory (M). They claim that all algorithms i.e. FPgrowth, FPmax FPclose need less main memory because of the compact FP-trees, MFI-trees, and CFI-trees.

In the year 2008, Vaibhav Kant Singh, Vijay Shah, Yogendra Kumar Jain, Anupam Shukla, A.S. Thoke, Vinay Kumar Singh, Chhaya Dule, Vivek Parganiha introduced an efficient method for frequent pattern mining [20] and showed how the different approaches achieve the objective of frequent mining. The process of data mining is helpful in generation of support systems that can help in various fildes. They used Time, Space as performance evaluation parameter and carried out the


Page 1-18 © MAT Journals 2017. All Rights Reserved

experimentation on Sparse and dense databases. They claimed that Apriori is the simplest and best algorithm for mining of frequent patterns from the transaction database. However, Prefix tree approach is better in performance as compared to Apriori.

Bingbing Ni, Zheng Song, Shuicheng Yan [21], proposed work in the following steps; A large size (~391k) human aging image database was crawled via a set of popular age related queries in first step. Then, after parallel detection and noise removal, a clean database with about 220k face instances is obtained. Finally, a robust multiple instance regressor learning method was developed for handling both noisy images and label. They used performance evaluation parameters like Mean absolute errors (MAEs), Age distribution statistics and carried out experimentations on Benchmark datasets, FG-NET UIUC etc. They claimed that automatic web image mining system towards building a universal human age estimator based on facial information, which is applicable to all ethnic groups and various image qualities, providing best result. However, before automatically obtained age information, a training set is required to be created which requires much amount of efforts and time.

In the year 2013, Harsh Mathur [22], worked for to find the feature extraction then finding the clusters of image with K-mean and after that finding the sequence of the image. Their approach is based on two categories i.e. abrupt change in intensity and portioning. They used time as performance evaluation parameter. They claimed that the mentioned approach gives better results in sequencing of images by using Spade Algorithm. However, the approach requires more user interaction.
The systematic representation of above mining fundamental related work is given in Table 2.2

***Table 2.2.*** *Mining fundamental related work*

| Sr. No. | Ref. no. Concerned Author(s) and year | Concept used | Performance evaluation parameter | Database used | Claims by Concerned Author(s) | Our findings |
|---|---|---|---|---|---|---|
| 01 | [8] Patricia G. Foschi, Deepak Kolippakkam, et al. 2002. | Extract patterns and derive knowledge from large collections of images, and also extract unique features for the images. | Precision, Recall | Not mentioned | The features used are sufficient to identify the patterns from the Images. and tested images required less human interaction for resolving the uncertain regions. | The detection process has been carried out manually, which is laborious and time consuming. |
| 02 | [9] Piotr Dollar Zhuowen Tu et al. 2007 | Feature mining paradigm for image classification and examine several feature mining strategies. | True positive False positive, computation time | MIT CBCL face data-set | This approach provides effective result for feature mining. | Systems use features of homogeneous type and com-putational complexity. |
| 03 | [10] Till Quack, Vittorio Ferrari, | Video Mining with Frequent | None | Not mentioned | Mining approach based on | The mentioned |


Page 1-18 © MAT Journals 2017. All Rights Reserved

| | | | | | | |
|---|---|---|---|---|---|---|
| | et al. 2006 | Itemset Configurations. | | | frequent item sets is a suitable and efficient tool for video mining. | approach fails in high resolution video mining technique. |
| 04 | [11] Josef Sivic and Andrew Zisserman. 2004. | Extract significant objects, characters and scenes in a video by determining the frequency of occurrence of spatial configurations | Precision, recall | None | Textured region provides good results for video mining. | They only discussed about textured regions and other regions are missed. |
| 05 | [12] A.Kannan , Dr.V.Mohan Dr.N.Anbazhagan. 2010 | Image retrieval based on optimum clusters is proposed for improving user interaction with image retrieval systems by fully exploiting the similarity information. | Precision and Recall | None | The mentioned method is very useful for image mining. By using this method we can easily find out the similar images related to query image. | In CBIR the features of the query image alone are considered . |
| 06 | [13] Ji Zhang Wynne Hsu Mong Li Lee. 2001 | The research issues in image mining, current developments in image mining, image mining frameworks, state-of-the-art techniques and systems. | None | Not mentioned | Gives two frameworks for image mining: function-driven and information-driven image mining frameworks. Both are useful for image mining technique. | Visual quality is not so good. So use of new visualization techniques for the visualization of image patterns is needed. |
| 07 | [14] Gangin Lee and Unil Yun 2013. | Mining strongly correlated sub-graph patterns by considering weight and support constraints | Runtime, Minimum threshold. | PTE dataset, DTP dataset. | MSCG algorithm not only mined meaningful patterns with strong correlation among the elements in sub-graphs efficiently but also advanced performance as reducing search space by the proposed measure. | Complexity is more and time required for the algorithm to execute seems to be more. |
| 08 | [15] R. Brown, B. Pham. 2005 | Used a general hierarchical image classifier approach, and | Positive count, Negative count, False | Biology and Vietnamese art images. | Capabilities of this approach is more as compare to other | Features of images are not mentioned. |


Page 1-18 © MAT Journals 2017. All Rights Reserved

| | | | | | | |
|---|---|---|---|---|---|---|
| | | illustrates the ease with which it can be trained to find objects in a scene. | positive count, False negative count. | | classifier. | |
| 09 | [16] Mihai Datcu, Herbert Daschiel, et al. 2003. | A concept of a prototype of a knowledge driven content-based information mining system produced to manage and explore large volumes of remote sensing image data. | Size, resolution, signal models. | EU SC, NERSC | The mentioned approach is KIM, which is a prototype of a next-generation knowledge-driven image information mining system developed for the exploration of large image archives. | computational complexity of the algorithm seems to be more as time estimation is not presented. |
| 10 | [17] C. Lakshmi Devasena, M. Hemalatha. 2011. | A hybrid image mining technique using LIM-based data mining algorithm. | Sample test, Image database used, Time. | Not mentioned | They claimed that the image mining system derived from the LIM based image matching technique provided promising results, which are comparable and significant. | Time required for the algorithm to execute seems to be more. |
| 11 | [18] Basura Fernando, Elisa Fromont and Tinne Tuytelaars. 2012. | An effective method for image classification using itemset mining is introduced to discover a new set of mid-level features called Frequent Local Histograms (FLH). | Classification accuracy, Equal error rate, Mean average precision. | GRAZ-01, Oxford-Flowers, 15-Scenes and the PASCAL-VOC2007 dataset. | They claim that their relevant pattern mining method and the chosen kernel all improve the classification results on various datasets. | The differences between the results of mentioned approach and other methods are not very obvious in case of the images which contain a large amount of edge details. |
| 12 | [19] Gosta Grahne and Jianfei Zhu, 2003. | A novel array based tech-nique that greatly reduces the need to traverse FP-trees and also present new algorithms for a number of common data mining problems. | CPU Time(s), Minimum Support (%), Main Memory (M). | Sparse , Synthetic and real datasets. | They claim that all algorithms i.e. FPgrowth, FPmax FPclose need less main memory because of the compact FP-trees, MFI-trees, and CFI-trees. | FPclose algorithm is good in to generates more non-closed frequent itemsets than the other algorithms. |


Page 1-18 © MAT Journals 2017. All Rights Reserved

| | | | | | | |
|---|---|---|---|---|---|---|
| 13 | [20] Vaibhav Kant Singh, Vijay Shah, et al. 2008. | An efficient method for frequent pattern mining is introduced. | Time, Space. | Sparse and dense. | Apriori is the simplest and best algorithm for mining of frequent patterns from the transaction database. | Prefix tree approach is better in performance as compared to Apriori. |
| 14 | [21] Bingbing Ni, Zheng Song, Shuicheng yan. 2009 | They carried out their work in the following steps. A large size (~391k) human aging image database was crawled via a set of popular age related queries in first step. Then, after parallel detection and noise removal, a clean database with about 220k face instances is obtained. Finally, a robust multiple instance regressor learning method was developed for handling both noisy images and label. | Mean absolute errors (MAEs), Age distribution statistics. | Benchmark datasets, (FG-NET UIUC etc) | An automatic web image mining system towards building a universal human age estimator based on facial information, which is applicable to all ethnic groups and various image qualities. And providing best result. | Before automatically obtained age information , a training set is required to be created which requires much amount of efforts and time |
| 15 | [22] Harsh mathur 2013. | Basic concept is to find the feature extraction then finding the clusters of image with K-mean and after that finding the sequence of the image. | Time | Not mentioned | Gives better results in sequencing of image by using Spade Algorithm. | Requires more user interaction. |

**Graph Related Work**

In the year 2006, Koji Tsuda, Taku Kudo [23], worked on clustering graphs. This work presented an approach to combine probabilistic inference and graph mining using simple mixture model. For general applications with insufficient domain knowledge, it is difficult to select the patterns manually in advance. They used frequent substructure mining methods to find the set of patterns that appear frequently in the database. They used computational time as a performance evaluation parameter. However, time


Page 1-18 © MAT Journals 2017. All Rights Reserved

required for the algorithm to execute seems to be more.

Taku Kudo, Eisaku Maeda, Yuji Matsumoto in year 2004, proposed an application of Boosting for classifying labeled graphs, general structures for modeling a number of real world data [24] and also discussed the relation between mentioned algorithm and SVMs with convolution kernels. The performance evaluation parameter used by them is REV, PTC and BOL (bag-of-lable). They claimed that the boosting algorithm is accurate and efficient for classification tasks involving discrete structural features. However, though the speed of boosting algorithm is fast but it seems that it is computationally complex.

In the year 2002, Xifeng Yan, Jiawei Han, introduce an algorithm called gSpan (graph-based Substructure pattern mining), which gives frequent substructures without candidate generation. gSpan builds a new lexicographic order among graphs, and maps each graph to a unique minimum DFS code as its canonical label [25]. Frequent substructure pattern mining has been a rising data mining problem with many scientific and commercial applications. They used Depth-First Search Tree as performance evaluation parameters and carried out their experimentations on Synthetic Datasets, Chemical Compound Dataset. However, computational complexity of the algorithm seems to be more as time estimation is not presented.

In the year 2012, Zhi-Yong Liu, Hong Qiao, Lei Xu, Fellow [26] used the path following algorithm to matching problems on directed graph models. Based on the concave and convex relaxations, they used PATH following algorithm to cover graphs. The performance evaluation parameter used by them are Mean, standard deviation, optimal matching. They carried out their experimentations on benchmark data set. They claimed that some experimental comparisons show the mentioned method is valid and good. However, Concave relaxation is better to solve matching problem than path following algorithm.

A.Sanfeliua, R.Alquezarb, J.Andradea, J.Climentc, F.Serratosad, J.Vergesa, in year 2000, presented a Graph-based representations and approaches for image processing and image analysis [27] and carried out their work in the following steps. In first step present a generalization of a graph partitioning greedy algorithm for colour image segmentation. Next describe a novel fusion of colour based segmentation and depth from stereo that yields a graph representing every object in the scene. Finally describe a new representation of a set of attributed graphs (AGs), denominated function described graphs (FDGs), a distance measure for matching AGs with FDGs and some applications for robot vision. They used Recognition ratio as a performance evaluation parameter and carried out their experimentations on Libor Spacek's database. However, computational time is not evaluated.

In the year 2005, Huaijun Qiu, Edwin R. Hancock [28], worked on a spectral method which used to partition graphs into non- overlapping subgraphs both for matching and clustering. They carried out their experimentations on Real word data. And used the performance evaluation parameters like House index, corners, correct, false, missed. They claim that the problem of matching the graphs into that of matching structural subunits can solve by using Fiedler vector of the Laplacian matrix.

Nan Hu, Raif M. Rustamov, Leonidas Guibas in year 2013, worked on the weighted graph matching problem with partially disclosed correspondences


Page 1-18 © MAT Journals 2017. All Rights Reserved

between a number of anchor nodes solve by using integer quadratic program (IQP) [29]. Their work provides max-margin formulation makes an effective use of the scarce training data even a small number of known correspondences leads to a large number of constraints on the proximity matrix. They used average, accuracy as the performance evaluation parameters. They claimed that they obtained superior performance as compared to the state of the art techniques based on adjacency matrices. However, they have not mentioned computational time for performance evaluaion.

In the year 2012, Xavier Desquesnes, Abderrahim Elmoataz, Olivier Lezoray [30], present PdEs-based morphology on graphs for cytological slides segmentation and clustering. They gave three methods for applications to cytology as first method - Papanicolaou coloration, Second method -Feulgen coloration and third method - Data Clustering. They claim that the approach is very general and has been illustrated through two methodologies, using different graph-based image representations.

In the year 2011, Xavier Desquesnes, Abderrahim Elmoataz, Olivier Lezoray [31], introduce a adaptation of PDEs level sets over weighted graphs of arbitrary structure, based on PdEs and using a frame- work of discrete operators. They claim experimental results have shown the potentiality of the proposed formulation of PDEs level sets and its adaptivity to graphs of arbitrary topology.

In 2006, Francesco M. Malvestuto, Mauro Mezzini, Marina Moscarini [32], provides minimal invariant sets in a vertex weighted graph. An invariant set is an edge set such that the sum of the weights of its edges is the same for every edge weighting. A complete axiomatisation of invariant sets is stated and give a polynomial algorithm to find a set of minimal invariant sets that completely characterise the set of all invariant sets. They carry out their experimentation on Statistical database. They claim that a graphical characterisation of minimal invariant sets and a complete axiomatisation of invariant sets both provides good result. However, Performance evaluation parameters are not given.

Vinh Thong Ta, Olivier Lezoray, Abderrahim Elmoataz [33], proposed a general formulation of discrete functional regularization on weighted graphs. For providing fast image segmentation, they had also proposed a graph based image simplification. Variational methods, based on regularization, provide a framework to handle image processing problems by designing and solving Partial Differential Equations (PDEs) in continuous domain. Then, PDEs are discretized in order to fit with the image domain. They claim that they provide two classification approaches. Both approaches can be applied in a wide range of applications in various domains, for instance data sets analysis or clustering. However, performance evaluation parameter and database are not mentioned.

The systematic representation of above graph related work is given in Table 2.3.

***Table 2.3.*** *Graph related work*

| Sr. No | Ref. no. Concerned Author(s) and year | Concept used | Performance evaluation parameter | Database used | Claims by Concerned Author(s) | Our findings |
|---|---|---|---|---|---|---|
| 1 | [23] Koji Tsuda and Taku Kudo, | The work present an approach to combine | Computation time | Not mentioned | It is possible to extend the learning method for more | Time required for the algorithm to |


Page 1-18 © MAT Journals 2017. All Rights Reserved

| | | | | | | |
|---|---|---|---|---|---|---|
| | 2006. | probabilistic inference and graph mining using simple mixture model. | | | advance probabilistic model. And if tree mining is used instead of graph mining, the time will be much shorter. | execute seems to be more. |
| 2 | [24] Taku Kudo, Eisaku Maeda, Yuji Matsumoto, 2004 | Decision stumps that use sub-graph as features, and a Boosting algorithm in which sub-graph-based decision stumps are used as weak learners. | REV , PTC and BOL | None | Boosting algorithm is accurate and efficient for classification tasks involving discrete structural features. | Though the speed of boosting algorithm is fast but it seems that it is complicated to understand. |
| 3 | [25] Xifeng Yan Jiawei Han, 2002 | Propose a novel algo- rithm called gSpan(graph-based Substructure pattern min- ing), which gives frequent substructures without can- didate generation. gSpan builds a new lexicographic order among graphs, and maps each graph to a unique mini- mum DFS code as its canonical label. | Depth-First Search Tree | Synthetic Datasets, Chemical Compound Dataset. | gSpan outperforms FSG by an order of magnitude and is capable to mine large frequent subgraphsin a bigger graph set with lower minimum supports | Requires a lot of user interaction, computational complexity of the algorithm seems to be more as time estimation is not presented. |
| 4 | [26] Zhi-Yong Liu, Hong Qiao, and Lei Xu, Fellow, 2012 | Used the path following algorithm to matching problems on directed graph models. Based on the concave and convex relaxations. | Mean, standard deviation, optimal matching. | Benchmark data set | Some experimental comparisons shows the proposed method is valid and good | Concave relaxation is better to solve matching problem than path following algorithm. |
| 5 | [27] A.Sanfeliua, R. Alquezarb et al 2000 | Graph-based representations and approaches for image processing and image analysis | Recognition ratio | Libor Spacek's database, University of Essex. | Graph approaches permit to represent objects in very natural way, without critical losses. | Computational time is not evaluated |
| 6 | [28] Huaijun Qiu and Edwin R. Hancock, 2005. | A spectral method which used to partition graphs into non-overlapping subgraphs both for | House index, corners, correct, false, missed. | Real word data | The problem of matching the graphs into that of matching structural subunits can solve by using Fiedler | Hierarchical matching method does not destroy cluster structure. |


Page 1-18 © MAT Journals 2017. All Rights Reserved

| | | | | | | |
|---|---|---|---|---|---|---|
| | | matching and clustering. | | | vector of the Laplacian matrix. | |
| 7 | [29] Nan Hu Raif M. Rustamov Leonidas Guibas, 2013. | The weighted graph matching problem with partially disclosed correspondences between a number of anchor nodes solve by using integer quadratic program (IQP). | Average, Accuracy. | Not mentioned | Obtained superior performance as compared to the state-of- the-art techniques based on adjacency matrices. | They claimed that they obtained superior performance but they does not mentioned computational time for performance. |
| 8 | [30] Xavier Desquesnes, Abderrahim Elmoataz and Olivier Lezoray 2012. | PdEs-based morphology on graphs for cytological slides segmentation and clustering. | None | None | The mentioned approach is very general and has been illustrated through two methods, using different graph-based image representations. In both cases, the proposed methodologies has given very good and promising results. | Feulgen coloration method requires more user interaction due to more complexity. |
| 9 | [31] Xavier Desquesnes, Abderrahim Elmoataz and Olivier Lezora. 2011 | A adaptation of PDEs level sets over weighted graphs of arbitrary structure , based on PdEs and using a frame- work of discrete operators | Not mentioned | None | Experimental results have shown the potentiality of the proposed formulation of PDEs level sets and its adaptively to graphs of arbitrary topology. | Sometime PDE based front propagation formulation enables the processing of a huge variety of discrete data. |
| 10 | [32] Francesco M. Malvestuto∗, Mauro Mezzini, Marina Moscarini. 2006 | Minimal invariant sets in a vertex-weighted graph. An invariant set is an edge set such that the sum of the weights of its edges is the same for every edge weighting, | Not mentioned | Statistical database | A graphical characterization of minimal invariant sets, and a complete axiomatization of invariant sets both are provides good result. | Performance evaluation parameters are not given. |


Page 1-18 © MAT Journals 2017. All Rights Reserved

| | | | | | | |
|---|---|---|---|---|---|---|
| 11 | [33] Vinh Thong Ta, Olivier Lezoray,and Abderrahim Elmoataz, 2007. | Author propose a general formulation of discrete functional regularization on weighted graphs. For providing a fast image segmentation they had also proposed a graph based image simplification. | None | Not mentioned | Provides two classification approaches. Both approach can be applied in a wide range of applications in various domains, for instance data sets analysis or clustering. | It requires more user interaction and also computational complexity seems to be more as no evaluation of time is presented. |

## RESULTS

There is a large literature on image mining, dating back over 20 years, with applications in Web Related Image Searching, Image Classification, and many areas other than computer vision as well. However, here, we have critically analyzed around thirty-five review and research papers. The findings from this exhausting literature review are summarized below:

- Mining approach has faster execution time and more compact memory usage due to weighted substructure.
- Graph approaches permit to represent objects in very natural way, without critical losses.
- Mining approach based on frequent item sets is a suitable and efficient tool for video mining.
- MSCG algorithm not only mined meaningful patterns with strong correlation among the elements in sub-graphs efficiently but also advanced performance as reducing search space by the proposed measure.
- gSpan outperforms FSG by an order of magnitude and is capable to mine large frequent subgraphsin a bigger graph set with lower minimum supports
- Boosting algorithm is accurate and efficient for classification tasks involving discrete structural features.

## IDENTIFIED PROBLEMS

- In any web search application when we try to search a particular image we get the irrelevant images compare to our search.
- It is very difficult task to sort image automatically in given database. Image can be model as the collection of parts.
- In image there is some pattern that can appear frequently. This frequent pattern is to be discriminative for telling apart samples of different classes.
- By identifying discriminative patterns in collection of image we mine the item (image).
- There is lot of working has been done for classification of images by using discriminative frequent pattern analysis.
- By using frequent pattern analysis, we can retrieve our relevant searched image from given set of images (database).
- It is possible to avoid outlier (irrelevant) images by using algorithms which can identify these images automatically.

## REFERENCES

1. Hong Cheng, Xifeng Yan, Jiawei Han and Chih-Wei Hsu, “Discriminative Frequent Pattern Analysis for Effective Classification”, 23rd International Conference on Data Engineering, (2007).
2. Bryan C. Russell, Alexei A. Efros, Josef Sivic, William T. Freeman and Andrew Zisserman, “Using Multiple Segmentations to Discover Objects and their Extent in Image Collections”,


Page 1-18 © MAT Journals 2017. All Rights Reserved

IEEE Computer Socity Conference on Computer Vision and Pattern Recognition, Vol. 2, Page No. 1605-1614, (2006).
3. J.C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang and Ph. Bunel "Image analysis by bidimensional empirical mode decomposition" Elsevier Image and Vision Computing Vol. 21, Page No 1019–1026, (2003).
4. Dougherty G, "Image analysis in medical imaging: recent advances in selected examples", Biomedical Imaging and Intervention Journal, Vol. 10, (2010).
5. Pew-Thian Yap and Raveendran Paramesran, "Image Analysis by Krawtchouk Moments", IEEE Transactions on image processing, Vol. 12, Page No. 11, (2003).
6. Xiaoguang Lu, "Image Analysis for Face Recognition", In Proceedings of personal notes, Department of Computer Science and Engineering USA Page No. 1-37, (2003).
7. A.Hema and E.Annasaro, "A survey in need of image mining techniques", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 2, ( 2013).
8. Patricia G. Foschi, Deepak Kolippakkam, Huan Liu and Amit Mandvikar, "Feature Extraction for Image Mining", Multimedia Information System, (2002).
9. Piotr Dollar, Zhuowen Tu, Hai Tao and Serge Belongie, "Feature Mining for Image Classification", IEEE Conference on Computer Vision and Pattern Recognition, Page No. 1-8, (2007).
10. Till Quack, Vittorio Ferrari and Luc Van Gool, "Video Mining with Frequent Itemset Configurations", Image and Video Retrieval Lecture Notes in Computer Science, Vol. 4071, Page No. 360-369, (2006).
11. [Josef Sivic and Andrew Zisserman, "Video Data Mining Using Configurations of Viewpoint Invariant Regions", IEEE Conference on Computer Vision and Pattern Recognition, Vol.1, Page No. 488-495, (2004).
12. A.Kannan , Dr.V.Mohan and Dr.N.Anbazhagan, "An Effective Method of Image Retrieval using Image Mining Techniques", International journal of Multimedia & Its Applications (IJMA) Vol.2, Page No.4, (2010).
13. Ji Zhang, Wynne Hsu, and Mong Li Lee, "Image Mining: Issues, Frameworks and Techniques", 2nd ACM SIGKDD International Workshop on Multimedia Data Mining, (2001).
14. Gangin Lee and Unil Yun, "Mining Strongly Correlated Sub-Graph Patterns by Considering Weight and Support Constraints", International Journal of Multimedia and Ubiquitous Engineering Vol. 8, Page No. 1, (2013).
15. [R. Brown and B. Pham, "Image Mining and Retrieval Using Hierarchical Support Vector Machines", IEEE Proceedings of the 11th International Multimedia Modelling Conference, (2005).
16. Mihai Datcu, Herbert Daschiel, Andrea Pelizzari, Marco Quartulli and Annalisa Galoppo et al. "Information Mining in Remote Sensing Image Archives: System Concepts", IEEE Transactions On Geoscience and Remote Sensing, VOL. 41, Page No. 12, (2003).
17. C. Lakshmi Devasena and M. Hemalatha, "A Hybrid Image Mining Technique using LIM-based Data Mining Algorithm", International Journal of Computer Applications (0975 – 8887) Volume 25, Page No.2, (2011).


Page 1-18 © MAT Journals 2017. All Rights Reserved

18. Basura Fernando, Elisa Fromont and Tinne Tuytelaars, "Effective Use of Frequent Itemset Mining for Image Classification", Computer Vision-ECCV, Vol.7572, Page No. 214-227, (2012).
19. Gosta Grahne and Jianfei Zhu, "Efficiently Using Prefix-trees in Mining Frequent Itemsets", IEEE ICDM Workshop on Frequent Itemset Mining Implementations, (2003).
20. Vaibhav Kant Singh, Vijay Shah, Yogendra Kumar Jain, Anupam Shukla, A.S. Thoke, Vinay Kumar Singh, Chhaya Dule and Vivek Parganiha, "Proposing an Efficient Method for Frequent Pattern Mining", World Academy of Science, Engineering and Technology 24, (2008).
21. Bingbing Ni, Zheng Song and Shuicheng Yan, "Web Image Mining Towards Universal Age Estimator", National University of Singapore 4 Engineering Drive 3 Singapore 117576, (2009).
22. Harsh mathur, "Sequential Mining of Multimedia Images by using SPADE Algorithm", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (6), Page No. 791-795, (2013).
23. Koji Tsuda and Taku Kudo, "Clustering Graphs by Weighted Substructure Mining", International Conference on Machine Learning, (2006).
24. Taku Kudo, Eisaku Maeda and Yuji Matsumoto, "An Application of Boosting to Graph Classification", 18th Annual Conference on Neural Information Processing Systems NIPS, Page no. 729-736, (2004).
25. [Xifeng Yan and Jiawei Han, "gSpan: Graph-Based Substructure Pattern Mining", International Conference on Data Mining (ICDM 02), Page No. 721–724, (2002).
26. Zhi-Yong Liu, Hong Qiao and Lei Xu, Fellow, "An Extended Path Following Algorithm for Graph-Matching Problem", IEEE Transations On Pattern Analysis And Machine Intelligence, Vol. 34, Page No. 7, (2012).
27. A.Sanfeliua, R.Alquezarb, J.Andradea, J.Climentc, F.Serratosad and J.Vergesa, "Graph-based representations and techniques for image processing and image analysis", Pattern Recognition Vol. 35 Page No. 639–650, (2000).
28. Huaijun Qiu and Edwin R. Hancock, "Graph matching and clustering usin g spectral partitions", Pattern Recognition Vol. 39, Page No.22–34, (2005).
29. Nan Hu, Raif M. Rustamov and Leonidas Guibas, "Graph Matching with Anchor Nodes: A Learning Approach", IEEE Page No. 2909-2913, (2013).
30. Xavier Desquesnes, Abderrahim Elmoataz and Olivier Lezoray, "Pdes-Based Morphology On Graphs for Cytological Slides Segmentation and Clustering", International Symposium on Biomedical Imaging (IEEE), Barcelone Spain, (2012).
31. Xavier Desquesnes, Abderrahim Elmoataz and Olivier Lezoray, "Pdes Level Sets On Weighted Graphs", International Conference on Image Processing (IEEE), Bruxelles Belgium, (2011).
32. Francesco M. Malvestuto, Mauro Mezzini and Marina Moscarini, "Minimal invariant sets in a vertex-weighted graph", Theoretical Computer Science Vol. 362 Page No. 140–161, (2006).
33. Vinh Thong Ta, Olivier Lezoray and Abderrahim Elmoataz, "Graph Based Semi and Unsupervised Classification and Segmentation of Microscopic Images",


Page 1-18 © MAT Journals 2017. All Rights Reserved

34. IEEE International Symposium on Signal Processing and Information Technology (2007).
35. Chuntao Jiang and Frans Coenen, "Graph-based Image Classification by Weighting Scheme", Applications and Innovations in Intelligent System , Page No. 63-76, (2009).


Page 1-18 © MAT Journals 2017. All Rights Reserved