International Journal of Modern Electronics and Communication Engineering (IJMECE) ISSN: 2321-2152
Volume No.-7, Issue No.-2, March, 2019

Literature Survey on Digital Forensic by Using Image Mining

CH. Kodanda Ramu
Department of CSE
MIRACLE EDUCATIONAL SOCIETY
Group of Institutions, Miracle City
Bhogapuram, Vizianagaram, Andhra Pradesh, India.
kodandachintu@gmail.com.

Dr. T. Sita Mahalakshmi
Professor, Department of CSE
GITAM (Deemed to be University)
Visakhapatnam, Andhra Pradesh, India
sita.tummala@gitam.edu.

Abstract: The mining of images is related to the development of mining techniques in the field of digital forensic science (DF). Image mining always works in the extraction of hidden data and additional information that is not clearly characterized in the images. Innovations in digital technology have new challenges today in the field of digital forensic imaging. In this field, some traditional approaches and frameworks are proposed to test the integrity and authenticity of digital images. Recovery-based tools are proposed to recognize unauthorized modified copies of the image. The authorization of a digital image can be tested by analyzing its acquisition. In the mining of images, the mining process depends on several techniques such as image management, information preparation, robotics and machine learning. The first imperative job of mining is to obtain highly vital elements of the hidden images. The mining of images contains diverse types of uses in diverse segments, like the investigation of the space, the remote detection, the finding of medicines, etc. Image mining can be a class of research systems that analyze a large-scale measurement of image learning. This document presents a review of several image mining procedures.

Keywords: Image handling, Image Mining, Data Extraction, Information Preparing, Digital Forensic, Digital image, Acquisition, Machine learning.

I. INTRODUCTION

Digital Forensics (DF) is the emerging field to identify, recover and analyze the evidence part of digital data. Process diagram for data extraction is shown in Fig1.1.

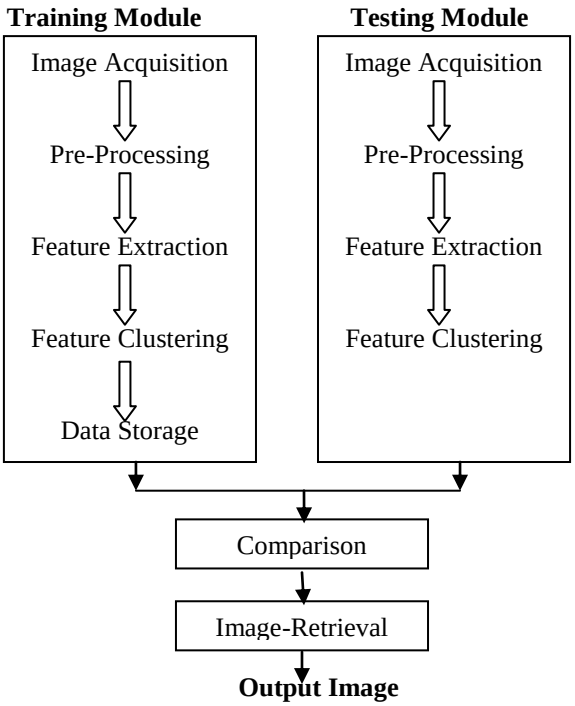

Figure 1.1: Data Extraction Flow Chart

This small part of the evidence is one of the crucial requirements in the cyber world. Image mining is part of the interdisciplinary field of knowledge discovery in database management. Digital forensic is the wing of forensic science merged with the recovery and investigation of material found in digital devices due to an incident of computer crimes. The digital forensic analysis can be classified into many that are shown in Fig: 1.2.

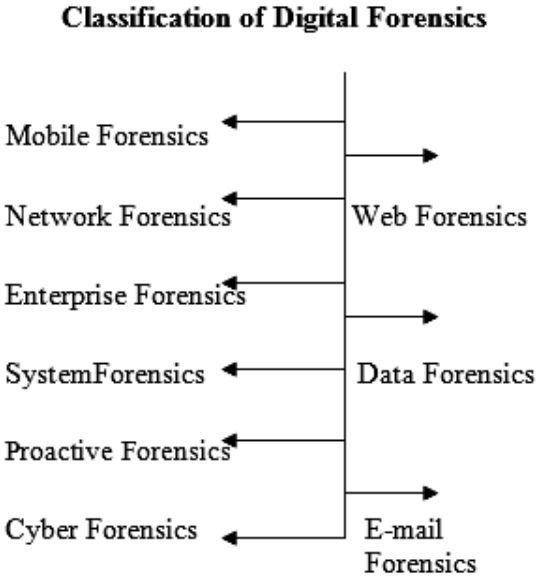

Figure 1.2: Classification Diagram of Digital Forensics

In last year's success in digital photography, the images are used in a variety of applications, including diagnostics, digital photography, industry, military photographs, robot, and more. Due to the positive effects, many new photo-

making tools are available in the market. Because of this, many deceased and cybercrime advances with the development of this tool, and many laws that are required to receive illegal or unauthorized disclosures from treatment the pictures. Almost all of these consultations focus on discovering the root or origin of a digital image with camera capture or other sources. The evidence identified from the forensic investigation will provide forensic information necessary for the use of cybersecurity, cloud security, intelligence agencies and law enforcement. Details on image acquisition techniques are useful for answering more criminal questions for confirmation of additional image processing. Image mining is a strategy to search for and search for useful information and data in large amounts of data. Image restoration, artificial intelligence, image preparation, and information extraction are techniques used to assemble learning. These strategies allow Image Mining to follow an unusual methodology for separating data from databases or accumulating images. In this way, the option is to extract a set of character set information. In the design architecture of image mining and image processing, recovery may be a rare type of spatial characteristics to minimize. Extraction of information involves improving the quality of the assets that are expected to accurately illustrate the large data organization. Some alternatives are used in image restoration. Those that prevail among them are properties in the light of shading, characteristics of the light surface and characteristics of the light model. The image mining strategy manages to extract some information and images with relevant data or several examples that are not clearly available in the images. Exploring images greatly, especially since the area unit in the image database is mostly non-proportional. In addition, some image credits do not appear to be particularly clear to the customer. Previously, photo exploration was conducted to solve various problems along with recognition of objective and visual observation, identification of confrontation and facial location verification. The aim of this document is to review the methods and procedures for the extraction of images that are evaluated periodically. Image mining is related to expanding data mining.

Thus, image mining is a field for image manipulation and pattern matching with data mining extensions.

II. LITERATURE REVIEW

The description of the writing in the investigation of different systems of image mining characterized some related innovations in the field of image management.

Sanjay T et al. [1] presents a method of photo mining that uses a wave change screen. The test of distinction and information mining models are used as part of the wavelet screen. The principal component analysis (PCA) framework of Discrete Wavelet Transformation (DWT) is used as a model for the recognizable testing framework. Caring for a considerable measure of the images on the supply machine can be a critical task and offers a memory management problem. The data should be in this sense prudent to treat images quickly.

Lionel Lionel Gueguen and Mihai Datcu [2] exhibited the rule of data bottlenecks to self-direct the issue of separating the applicable data from the Satellite Image Test System (SITS). The model depends on the analysis of property analysis for the safety of the groups. Irregular Gauss-Markov fields and arbitrary auto-binomial fields are used to describe the fleeting spatial structures contained in SITS. In the middle of this approach, no frightening or geometric data was thought of. This is the significant disadvantage here.

Selim Aksoy and R. Gokberk Cinbi [3] proposed new directional spatial imperatives, which is the new procedure for image mining. The creator used the accuracy and physically recognizes the elements that meet each query to calculate the execution of the recovery. He wants a manual elucidation to process the time.

P.Rajendran and M.Madheswaran [4] specified a photographic mining procedure. It joins low-level alternatives extracted from images and abnormal state data of authority. It does not deal with excess, the noisy image and the quality of time.

Ankita Tripathi et al. [5] proposed a system that classifies textual images (images that find text inside) using low-level image features. Image classification and content-based image retrieval is a growing field in the image classification area. The approach is based on several low-level image features that include Gray Level Coexistence Matrix (GLCM) characteristics as a mean, bias, energy, contrast, homogeneity. Using these various characteristics, the differences between the images are measured, and then they are used to classify the textual images by performing classification and grouping techniques in data sets. The proposed method experimented in 60 different textual images to obtain an improved result that was not obtained in previous systems together with the classification of images in three main categories: document, scene and title.

Ross Brown et al. [6] outlined an image mining framework for advanced measurable applications. Bayesian systems are utilized to change the data vulnerabilities that happen in scientific work. These are utilized to communicate between convictions, adjust to entirely unexpected clients, diverse example recovers and copy human judgment of semantic substance of image patches.

Sanjay Silkari and Dr. Mahesh Motwani [7] arranged a system which concentrates on Color Moment Block Truncation Coding (BTC) which utilized shade is an element to extract the element from image dataset. At that point, K-Means clustering equation was used to group the image dataset into various groups.

Tao Jiang and Tan [8] proposed two approaches for finding the basic relationship among the content and images. The First system finds the execution measures between the information closeness and visual choices. Another technique utilizes a neural system to discover coordinate mapping among visual and content highlights by incrementally related highlights into a gathering of information format. It must perform group learning on the settled arrangement of unrefined information.

Hemlatha and Devasana [9] proposed to search out the right image while mining a media framework and built up a path for mining images by recommending Lorenz Information Measure (LIM) subordinate image coordinating technique with the neural system.

B. Sathees Kumar and R. Anbu Selvi [10] focused on the comparison of three different intensity-based feature extraction method for the abnormal patterns in brain tumors. Physician's interpretation of brain tumors may lead to misclassification sometime. Hence an automated system is needed to solve our problem. The following major categories of brain tumor images are taken into our consideration. They are Metastatic bronchogenic carcinoma, Astrocytoma, Meningioma, sarcoma. The performance factor was evaluated against BRATS (Brain Tumor Segmentation) dataset. For the purpose of calculating and extracting various intensity related features MATLAB tool is used. The experimental results suggest that among the intensity-based feature extraction methods GLCM (Gray Level Co-occurrence) method is showing better results than the other methods. Waikato Environment for Knowledge Analysis (WEKA) tool classification algorithm J48 also shows close correlation with GLCM Features

Ja-Hwung Su et al. [11] presented a new approach named Navigation-Pattern-based Relevance Feedback (NPRF) by integrating the navigation pattern mining and a navigation-pattern-based search approach named NPSearch. The main feature of NPRF is to efficiently optimize the retrieval quality of interactive content-based image retrieval (CBIR). As a result, traditional problems such as visual diversity and exploration convergence are solved.

Noorhaniza Wahid [12] arranged a swarm-based recipe for order. He contrasted swarm-based procedure and Support vector machine(SVM) and acquired the outcome. Particle Swarm Optimization (PSO) must distribute extra memory than Simplified Swarm Optimization (SSO) for every molecule to acknowledge higher execution.

Satej Wagle et al. [13] proposed a medical image classification model for retinal fundus images. It uses a modified k-nearest neighbor (KNN) algorithm, which uses an instance-weighting scheme based on the distance measure. The proposed model was implemented on “Normal” and “Severe” category of images. The model is evaluated using the performance metrics such as classification accuracy and area under the ROC curve. The experimental results using the statistical texture-based features extracted from the images have shown that the performance of this improved KNN model is significantly better than some of the other existing classifiers.

Surya S. Durbha et al. [14] created highlight determination and highlight change in view of wrapper-based approach. It utilized locale-based framework. They have contended that picking a significant list of capabilities will expand the rate of semantic class. This framework directly utilized symbolism from only one finder.

Bambang Harjito and Heri Prasetyo [15] presents a new method on image forensics application using the Error Diffusion Block Truncation Coding (EDBTC) feature. The image forensics tries to detect the copy-move forgery image regions on the forged image. Firstly, an image is divided into several non-overlapping image blocks. The image feature is further derived for each image block. Herein, two image features, namely Color Feature (CF) and Bit Feature (BF), are composed from the EDBTC compressed data stream. The forged region is detected while the image feature of this region is similar to the image feature of another region separated far away. As documented in the experimental section, the proposed method gives a promising result on the image forensics takes, and, at the same time, outperforms the former existing scheme.

Rajshree S. Dubey [16] delineated the image mining ways that rely upon the histogram and surface of the image. The required image is considered, the shading bar diagram and Texture is made, the resultant image was found. In this technique registering time was not considered.

Akshara Ravi et al. [17] studied about Recovering deleted files. When a file is deleted, only pointers that link file's metadata to its content are deleted and metadata entry is marked as deleted. As long as data is not overwritten or wiped, deleted data will remain in unallocated space. One of the methods that can be used to recover these deleted files is file carving. File carving reconstructs files only based on their content unlike traditional data recovery methods that use metadata that points to the content. It is mainly done using headers and footers of file types. One of the primary challenges in file carving is to recover deleted files that were fragmented. When a file is fragmented, carving only based on header and footer will produce corrupted file. So, they discussed a method for carving fragmented document and image files from a Universal Serial Bus (USB) drive.

. David A. Fay et al. [18] have built up a model for multisensory image combination and intelligent image mining in light of neural models of shading vision picking up, controlling and design acknowledgment. The modules like Image securing, Image Fusion, extraction of Context highlights, and intuitive Image Mining produce a client to frame vector item in light of highlights like streets, streams, and woods and featured target discoveries from crude multisensory or multispectral symbolism. These yield images changed by recreated ecological conditions not self-tended to in the memory, for promoting identification.

Herbert Daschiel and Mihai Datcu [19] portray the thoughts of learning substance-based image mining framework made to oversee and investigate monstrous volumes of remote detecting image information. The framework works both on the web and disconnected web interface. The disconnected part goes for the extraction of crude image choices, their pressure, and data decrease, the age of a totally unsupervised image content-record, and along these lines the substantial procedure of the index passage inside the data framework. This approach doesn't self-tend to the learning and data that is put away in the data framework.

Wynne Hsu and mong Li Lee [20] inspected the investigation issues in image mining, advancement in image mining. They anticipated a data-driven system for image mining. In that, they made out four levels of data: component level, Object level, semantics thought level, and information level. This strategy does not focus on further extension.

Aura Conci and Everest MathiasM. M.Castro [21] proposed a framework for mining images by color content. The implementation was built using CGI, Java and C languages. The framework provides the possibility of use 5 distance function for evaluation of similarity among images and 2 types of quantization. For comparing the influence on system performance of these

parameters a simulating engine was implemented and all possible combinations have been tested considering sets of similar images. This framework can try online at html://www.caa.uff.br/~mathias/visual.htm.

Tanima Dutta [22] proposed a new technique for image compression. Image encoding and transformation of coefficients are the main steps in this process. Color space conversion from Red, Green and Blue (RGB) to Y′ is the luma component and CB and CR are the blue-difference and red-difference chroma components (YCBCR) is used for the transformation. The image is divided into non-overlapping blocks to decrease the number of operations. Then the image is transformed using integer discrete cosine transform (Int DCT). The transformed coefficients are then quantized using a quantization matrix. After transformation, the coefficients are 49 encoded. The coefficients of a block are partitioned into DC and AC coefficients and the DC coefficients are differentially encoded.

Deepak Kumar Jain et al. [23] proposed a technique for image compression using discrete cosine transform (DCT) and adaptive Huffman coding (AHC). The original image is divided into blocks. DCT is applied to each block by the pre-multiplying the modified block with DCT matrix and by the post multiplying by transposes of DCT matrix. Each block is then compressed through quantization. A quantization matrix is used in combination with a DCT coefficient matrix to carry out the transformation. The compressed image is reconstructed through reverse process i.e., by using inverse DCT. This technique has good performance as compared to other algorithms.

Rajkumar and Latte [24] analyzed renowned wavelet-based image encoding scheme (SPIHT). The Region of Interest (ROI) coding commences with the selection of ROI and its respective resolution by the user. The diverse ROIs are encoded with diverse resolution by applying lifting wavelet transform (LWT) and SPIHT. Their experimental results illustrate that using lifting wavelet transform and SPIHT, the proposed ROI encoding scheme provides high compression ratio and quality ROI.

Changlong Chen et al., [25] Proposed Geometric transformations, such as resizing and rotation, are almost always needed when two or more images are spliced together to create convincing image forgeries.

Anil Dada Warbhe et al. [26] presented a digital image forensic method which can detect one of such image tampering. As images can be tampered in a number of ways, they address a common case called as copy-paste tampering. Their proposed method was robust to affine transform; especially to rotation and scaling.

Sepideh Azarian-Pour et al. [27] proposed a new automatic method for discriminating original and tampered images based on "JPEG ghost detection" method, which is a subset of format-based image forensics approaches. The inconsistency of quality factors indicates that the photo is a composite one created from at least two different cameras and therefore it is a manipulated photo. classification algorithm first extracts the ghost border. Then the image is classified as original or tampered groups by thresholding a distance in feature space.

Neetu Singh and Rishabh Bansal [28] studied Benford law in Digital Image Forensics. Joint Photographic Experts Group (JPEG) and JPEG2000 approximately follow this law and JPEG2000 found closer to the given law. The non-following of this law in different forensics setups can be used as fingerprint. Further, amount of deviation from Benford law in compressed images can be used to find forgery.

Jocelin Rosales Corripio et al. [29] In this work we use the pattern noise of an imaging sensor to classify digital photographs according to the source smartphone from which they originated. This is timely work as new smartphone models large imaging sensors, afford significant improvements in classification rates using pattern noise. Our approach is to extract wavelet-based features which are then classified using a support vector machine. We show that this method generalizes well when the number of source cameras is increased.

Chao-Yung Hsu et al. [30] proposed an efficient cross-camera vehicle tracking technique via affine invariant object matching. To improve the accuracy of Affine and scale-invariant feature transform (ASIFT) feature matching between images from different viewpoints, they observed the spatially invariant property of ASIFT and exploited this property with the min-hash technique to achieve efficient vehicle object matching.

H. N. Gangavane et al. [31] presented a model using new methodology for evaluation of document clustering of criminal database by using k-means clustering technique. This model clusters the criminal data basing on the type crime.

Mandar Jadhav and K.K. Joshi [32] presented a methodology for acquisition of internal data from Firefox operating system based mobile device and its complete analysis to recover important evidence. It gives a way to start the process of acquiring data by use of Android Debug Bridge, followed by identification of important files which can be analyzed to obtain valid results. Further, it exhibits how use of Magnet axiom toolkit helped us to recover images, documents, audio files and web related data from the acquired evidence. They *also showed how tools like qtADB, FTK Imager and SQLite* Forensics helped us in forensic investigation.

Mohammed Babiker et al. [33] Web application attacks are an increasingly important area in information security and digital forensics. It has been observed that attackers are developing the capability to bypass security controls and launch a large number of sophisticated attacks. Several attempts have been made to address these attacks using a wide range of technology and one of the greatest challenges is responding to new and unknown attacks in an effective way. This study aims to investigate the techniques and solutions used to detect attacks, such as firewalls, intrusion detection systems, honeypots and forensic techniques. Data mining and machine learning techniques, which attempt to address traditional technology shortcomings and produce more effective solutions, are also investigated. It was aimed to contribute to this growing area of research by exploring more intelligent and convenient techniques for web application attack detection by focusing on the data mining techniques in forensics.

III. CONCLUSION

This document presents a review of several image extraction procedures suggested in previous years. It also conveys a computed summary of the system. Image extraction is simply learning mining in the field of image processing and pattern recognition. Future examination work may include the use of the Bayesian input system and a large number of serious tests with optional samples for unauthorized image work. It is also envisaged that personal tests are carried out with the participation of scientific bodies. Some potential destinations consider what will be done within the image extraction area; capture experiments in parts of optional images such as the surface, shape, etc. You'll also be interested in searching for hidden connections between images. For example, intense experimental search and serious research within existing frameworks in the database may be very useful. The discoveries can be updated later in the extract image field. It has been noted that many studies have been conducted on biometrics, facial art, drawings, iris, palm trees, etc., but the researcher did not reach the high efficiency range due to incorrect information at the entrance. Therefore, this work can be carried out through the use of effective digital forensic techniques.

REFERENCES

[1] Sanjay T. Gandhi, K. T. Talele and Avinash G. Keskar, "Image Mining Using Wavelet Transform", Knowledge-Based Intelligent Information and Engineering Systems, Springer link book chapter, pp. 797-803, 2007.
[2] 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.
[3] Selim Aksoy and R. Gokberk Cinbi "Image Mining Using Directional Spatial Constraints", IEEE Geosciences and Remote Sensing Letters, vol. 7, no. 1, pp. 33-37, Aug 2009.
[4] P. Rajendran and M. Madheswaran, "An Improved Image Mining Technique for Brain Tumor Classification Using Efficient Classifier", International Journal of Computer Science and Information Security, Vol. 6, no. 3, pp. 107-116, Jan 2009.
[5] Ankita Tripathi, Shivam Pandey, Hitesh Jangir, "An improved and efficient image mining technique for classification of textual images using low-level image features", Inventive Computation Technologies (ICICT) 2017.
[6] 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.
[7] 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.
[8] Tao Jiang and Ah-Hwee Tan, Learning, "Image-Text Associations", IEEE Transactions on Knowledge and Data Engineering, Vol. 21, no. 2, February 2009
[9] 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.
[10] B. Sathees Kumar, R. Anbu Selvi, " Feature extraction using image mining techniques to identify brain tumors", Innovations in Information, Embedded and Communication Systems (ICIIECS),2015 International Conference.
[11] Ja-Hwung Su, Wei-Jun Huang, Philip S. Yu, and Vincent S. Tseng, "Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns", IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 3, March 2011.
[12] Noorhaniza Wahid, "A Novel Image Classification Algorithm Using Swarm-Based Technique for Image Database", Springer-Verla-Berlin Heidelberg, pp. 460-470, April 2011.
[13] Satej Wagle, J. Alamelu Mangai and V. Santosh Kumar, "An Improved Medical Image Classification Model using Data Mining Techniques", 2013 IEEE GCC Conference and exhibition, November 17-20, Doha, Qatar.
[14] Surya S. Durbha, Roger L. King, and Nicolas Yunnan, "Wrapper-Based Feature Subset Selection for Rapid Image Information Mining", IEEE Geoscience and Remote Sensing Letters, vol. 7, no. 1, January 2010.
[15] Bambang Harjito and Heri Prasetyo," Image forensics using EDBTC feature", Consumer Electronics - Taiwan (ICCE-TW), 2017 IEEE International Conference.
[16] Rajshree S. Dubey, "Image Mining using Content-Based Image Retrieval System", (IJCSE) International Journal of Computer Science and Engineering Vol. 02, no. 07, pp. 2353-2356, October 2010.
[17] Akshara Ravi, T Raj Kumar, Angelo Renju Mathew, "A method for carving fragmented document and image files" Advances in Human Machine Interaction (HMI), 2016 International Conference.
[18] David A. Fay, Richard T. Ivey, Neil Bamberger, and Allen M. Waxman, "Image Fusion & Mining Tools for a COTS Environment", ISIF pp. 606-613, 2003.
[19] 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.
[20] 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.
[21] 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.
[22] Tanima Dutta, "Medical Data Compression and Transmission in Wireless Ad Hoc Networks" IEEE Sensors Journal, Vol. 15, No. 2, Feb. 2015.
[23] Deepak Kumar Jain, Devansh Gaur, Kavya Gaur, Neha Jain "Image Compression using Discrete Cosine Transform and Adaptive Huffman Coding" International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 1, January - February 2014.
[24] T. M. P. Rajkumar and Mrityunjaya V Latte. "ROI Based Encoding of Medical Images: An Effective Scheme Using Lifting Wavelets and SPIHT for Telemedicine" International Journal of Computer Theory and Engineering, Vol. 3, No. 3, June 2011.
[25] Chenglong Chen,Jiangqun Ni,Zhaoyi Shen, "Blind Forensics of Successive Geometric Transformations in Digital Images Using Spectral Method: Theory and Applications" IEEE Transactions on Image Processing (Volume: 26, Issue: 6, June 2017).
[26] Anil Dada Warbhe, Rajiv V. Dharaskar, Vilas M.Thakare, "Digital image forensics: An affine transform robust copy-paste tampering detection", Intelligent Systems and Control (ISCO), 2016 .
[27] Sepideh Azarian-Pour, Massoud Babaie-Zadeh, Amir Reza Sadri, "An automatic JPEG ghost detection approach for digital image forensics" Electrical Engineering (ICEE), 2016.

[28] Neetu Singh and Rishab Bansal, "Analysis of Benford's law in digital image forensics" Signal Processing and Communication (ICSC), 2015.
[29] Jocelin Rosales Corripio, David Manuel Arenas González, Ana Lucila Sandoval Orozco, "Source smartphone identification using sensor pattern noise and wavelet transform" Imaging for Crime Detection and Prevention, 2013.
[30] Chao-Yung Hsu, Li-Wei Kang, Hong-Yuan Mark Liao, "Cross-camera vehicle tracking via affine invariant object matching for video forensics applications" Multimedia and Expo (ICME), 2013 IEEE International Conference.
[31] H. N. Gangavane, M. C. Nikose, P. C. Chavan, "A novel approach for document clustering to criminal identification by using ABK-means algorithm" Computer, Communication and Control (IC4), 2015 International Conference.
[32]Mandar Jadhav, K.K. Joshi, "Forensic investigation procedure for data acquisition and analysis of Firefox OS based mobile devices" Computing, Analytics and Security Trends (CAST) 2017.
[33]Mohammed, Babiker, EnisKaraarslan, Yasar Hoscan, "Web Application Attack Detection and Forensics:A Survey",978-1-5386-3449-3/18/$31.00 ©2018 IEEE