

Deduplication Analysis of Products in Digital Marketing

P. Amudhavalli¹, N. Rajalakshmi², K.S. Sindhu³

^{1,3}Department of Computer Science and Engineering, Karpagam Academy of Higher Education, India

²Department of Bio Medical Engineering, Karpagam Academy of Higher Education, India

Article Info

Article history:

Received Oct 29, 2017

Revised Jan 2, 2018

Accepted Jan 18, 2018

Keywords:

De-duplication

Forensic Analysis

Pixel Analysis

Least square support vector machine

ABSTRACT

As Digital Marketing is becoming more popular, the number of customer's interpretation on brands is increasing promptly which makes it firmer for companies to evaluate their brand image and to digital market their products on the web. The Forensic Analysis is used to determine and analyze patterns of fraudulent activities on images. Pixel Analysis and Least square support vector machine are used to compare and associate the scores acquired from the images into one result per tweet. We selected these techniques to compare and find the accuracy of the Digital Marketing images with the received product's images to identify the fraudulent activities on images in Digital Marketing. As the result of this project the customer can identify whether the received product is exactly what is given in the online purchase website.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

P.Amudhavalli,

Department of Computer Science and Engineering,

Karpagam Academy of Higher Education,

Pollachi Main Road, Eachanari Post, Coimbatore – 641 021, India.

Email: info@karpagam.com

1. INTRODUCTION

The extensive use of image editing software has induced an ascent interest in systems able to distinguish original images from tampered images and to create the validity of digital photographs. By enabling methods in existing forensic literature, forensic investigators can not only classify the possession environment of digital content, but also sense the processing history that the content has gone through after achievement. Today digital marketing has taken a new route in the industrial field. Now-a-days almost every camera is equipped with digital camera. It consist of various image editing software like Corel draw, Photoshop, Firework, etc. The introduction of these helps designed to enhance and improve the quality of the image. However some people used these tools to create tampered pictures and spread viral throughout the social media. Forensic analysis helps us to stand as an evidence for law enforcement.

2. RELATED WORK

Nowadays, in social media, we see a lot of faked images. These were created by using Forensic techniques. We are using tampering detection method [1], which is able to sense the active attempt to compromise the data integrity. Tampering detection based on image hashing to analyze and gather information about modified image. The modified images and the original images were segmented by image segmentation method. It is the process of partitioning the images. The segmented images were then matched with each other by image alignment technique. Forensic hash technique acts as evidence to extract information about the forensic images [5]. From the resulted forensic image the relationship among all the segments is ranked with the help of a manifold ranking technique. Gaussian Filter method was used by forgers because it eliminates noise and smooth transition [3]. It was also known as Gaussian Blur used to blur the images. It uses

Gaussian Filter Residual (GFR) and Frequency Transform Residual (FTR). To determine whether it is original or not can be detected by using Affine Transform. It is a linear mapping method that preserves points, straight lines, and plains.

A Copy-Paste tampering [4] is the process of copying a region of an image which is scaled before pasting to some other location in that image. A novel method of the JPEG Ghost detection method was used to make judgment whether it is original or tampered [5]. It performs format based image forensic approach. Extracting a finger print of digital camera can be performed by pixel to pixel, non-uniformity, since every picture has an overlaid weak-noise. Knowing the value of digital camera we can identify whether it is source camera. Photo Response Non Uniformity (PRNU) [6] plunk as a reliable technique in fingerprint extraction of digital camera. It is important to find out the limit of capability of forensic by using an information theoretical framework. Forensic ability [7] is used to collect the maximum information that can be obtained from the action by performing some operation. The mutual information obtained having similarity in their concepts such as features and operations [2].

By using Mutual information the forensic ability is measured. The result is used to find out the Error Probability rate. In image processing, the units inherit the images in raster bitmap [8]. In this paper an efficient method is used to identify the previous JPEG compressions. A novel technique is used to find out tamper detection and tamper localization [1]. Image hashing algorithm was used. The tampering was identified in the small regions, corners, an exact location of the tampered image based on this the algorithm is applied using ring partition, Non Negative Matrix Factorization (NMF). The particular tampered region is detected using Image Hash Algorithm.

The variation of JPEG exploitation is used for image authentication. From a JPEG image a camera signature is extracted and it contains the information about quantization tables, Huffman codes, thumbnails etc [8]. 773 different cameras and cell phones were used across 1.3 million images. 62% signature of an image is unique by the single camera and 80% was shared by three or more cameras. The perceptual image hashing maps an image appeared to a human eye to a fixed length string. Its applications are image indexing, authentication and watermarking [10]. In this paper a general framework is used by feature points which should be largely invariant under distortion. An iterative feature detector is used to satisfy this property. In digital image forensic the JPEG [8] plays a vital role. The study of an image forensic is termed as JPEG error analysis. The main errors include quantization, rounding and truncation is obtained in JPEG. Single or double compression of an image is theoretically analyzed.

3. SYSTEM DESIGN

Forensic Data Analysis is the study of Digital Forensic. It helps to create tampered image which is stored in a structured database and to examine the fraudulent activities. Digital Camera captures the image of an object. Pixel analysis technique calculates the value of the two different images stored in database. Orthogonal transformation was carried out between these images. Digital value is calculated and creates a histogram to examine the difference between these images, show in Figure 1.

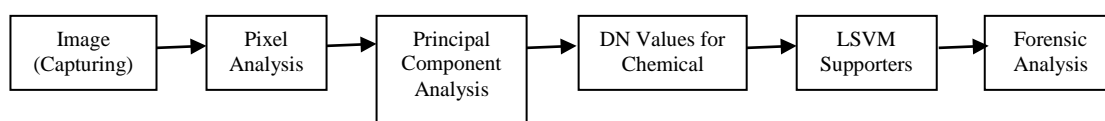


Figure 1. System Design

3.1 Pixel Analysis

The pixel is a basic unit in a computer display or an image. It takes as a logical unit rather than as a physical unit. A pixel is a sample of an image more samples are grouped together to produce a digital image. This tool helps us to drag a spatial position at a particular point. In this technique Color Based Segmentation using K-Means Clustering [13] method is performed. Reads the image from the database and convert the image from RGB color to Space portion using CIELAB. Which help us to differentiate pixel based on the color and partition it. Object in the cluster was performed to group the pixel. Blue Nuclei are the L value in CIELAB which segment the pixel color as light blue and dark blue. A and B represent the color component to segment the images. Various colors are tracked from the image and grouped together.

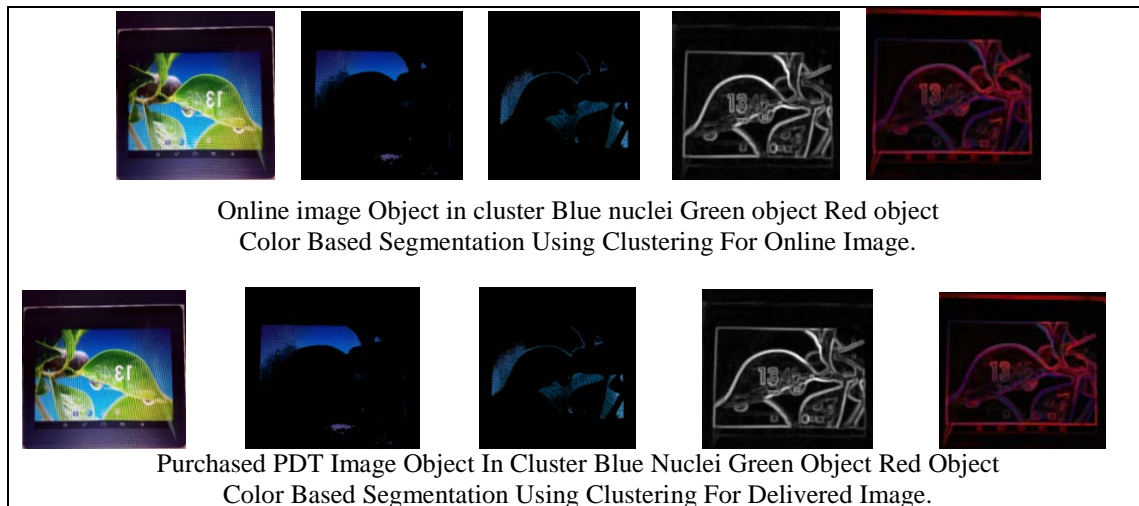


Figure 2. Color Based Segmentation Using K-Clustering

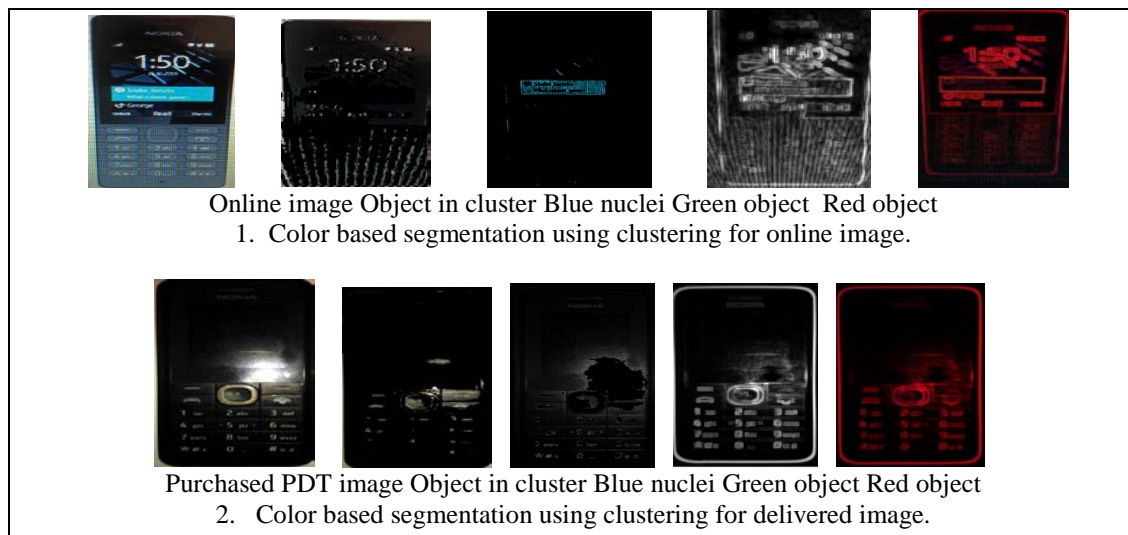


Figure 3. Pixel Analysis based on K-clustering technique to find the difference between the object

3.2 Principal Component Analysis

It is a geometric method for orthogonal transformation. To convert a set of interrelated variables to a set of uncorrelated variables is also known as principal components. PCA is sensitive to the relative scaling of the original variables. It is a statistical procedure used as a tool for data analysis and for making procedural models. The main purpose of PCA is to analyze a data and to identify patterns. In this method captured an image and the tampered image was converted to a grayscale image and then recovered the other image. Recovered images and the original image produce the difference between them in values of digital number and the elapsed time taken to complete the operation. From Figure3, we can see the elapsed time and digital number to distinguish between the object.

3.3 Least-Square Support Vector Machine

Least squares support vector machines (LS-SVM) is a set of related supervised learning techniques that examine data and identify patterns, which is used for classification and regression analysis. It is the fastest technique. It was based on the simple iterative approach. For example, it has been used to classify a dataset with 2 million points and ten features in only 34 minutes on a 400 MHz Pentium-II. Optimization tools in SVM are used for solving machine learning problem. In this method, value has been plotted to two different labels. The graph is drawn to relate these points.

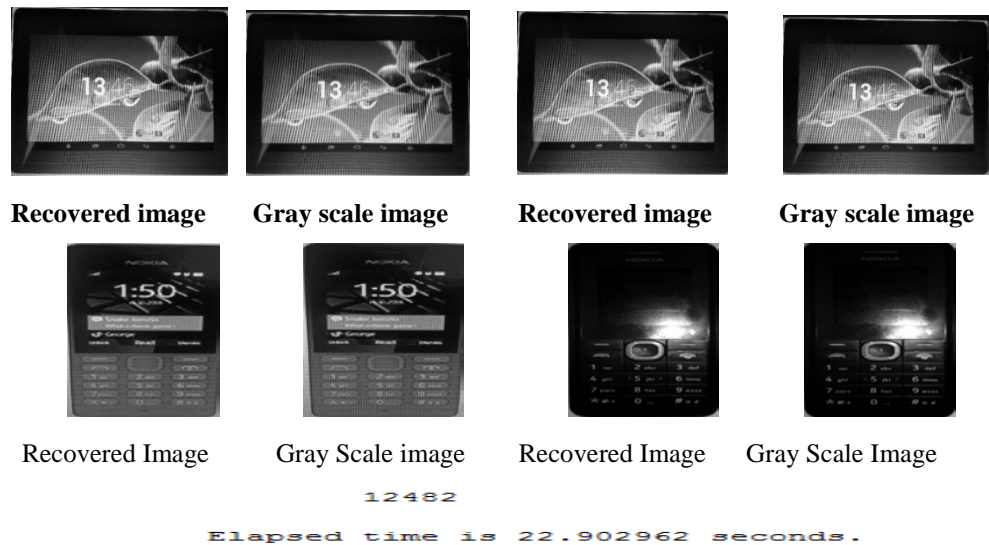


Figure 4. Principal Component Analysis For Tampered Image And Database Image

3.2 Forensic Data Analysis

Forensic Data Analysis (FDA) is a study of Digital forensics. It examines structured data with consider to incidents of financial crime. The aim is to see and investigate patterns of fraudulent activities. Data from function systems or from their underlying databases is referred to as planned data. Analysis of large volume of data is handles by analysing team. The data copies of the images are aligned in separate database to prevent original images. To analyze large structure data with intense of detecting financial crime it takes three expertises from the analysing team.

4. SYSTEM ARCHITECTURE

On plan premises, we have various hardware set up along with software tools. Some of the tools are Microcontroller, Digital Camera, LED, Buzzer.

A microcontroller is a mini computer with Integrated circuits. At first it was programmed in assembly language later, it also supports high level programming languages like C, Python, JavaScript. Nowadays an embedded system software is used to code microcontroller.

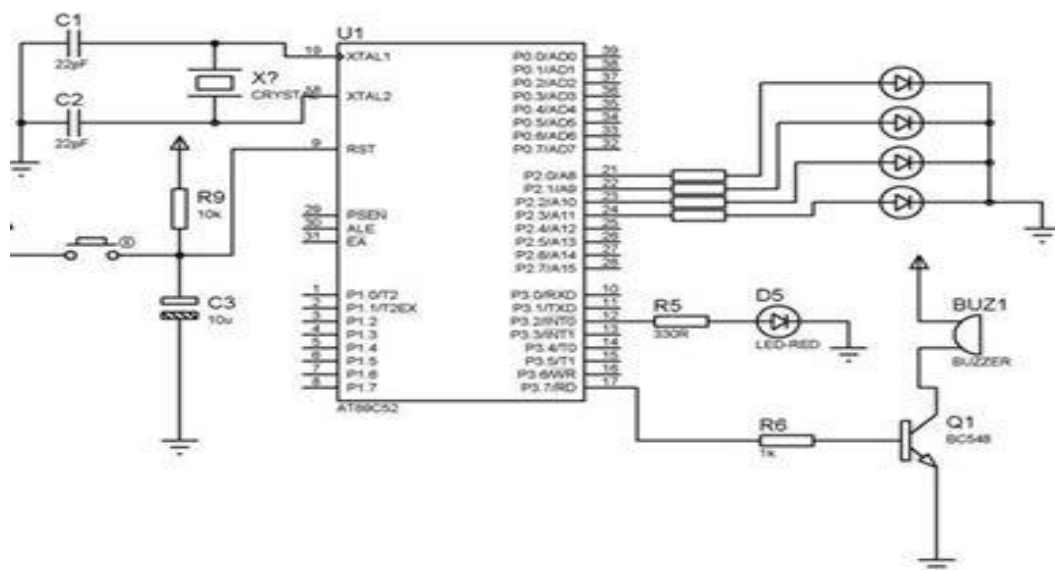


Figure 5. Architecture Diagram of system

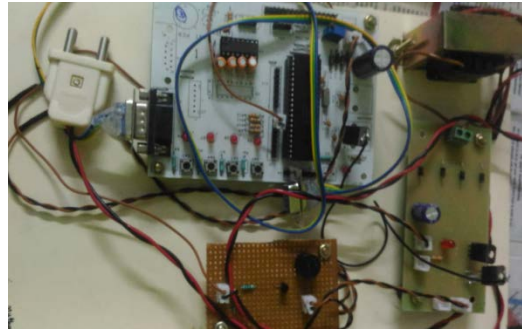


Figure 6. De-Duplication Analysis Kit

4.1 Digital Camera

Digital Camera is used to take pictures and to store it on a computer or memory. Now, almost every phone is equipped with a digital camera. The Resolution of an image can be captured when light falls through the image.

4.2 Light Emitting Diode (LED)

LED is a semiconductor device produce light when Electric current passes through it. Various applications in which LED is performing is indicator light, LCD panel backlighting, fiber optic data transmission, remote control.

4.3 Buzzer

It is an audio signaling device which is a mechanical or piezoelectric. Uses of beeper which is also known as the buzzer is in alarm timings, timer, and conformation of input like mouse clicks or keyboard. The original or tampered image is detected when the led or buzzer glow by forensic technique.

4.4 Serial Adapter

USB adapter is a type of protocol converter which is utilized for converting USB data signals to and from other communications standards. Commonly, USB adaptors are acclimated to convert USB data to standard serial port data and vice versa.

Most commonly the USB data signals are converted to either RS 232, RS 485, RS 422 or TTL serial data. The older serial RS 423 protocol is infrequently utilized anymore, so USB to RS 423 adapters are less prevalent.



Figure 7. USB to Serial Adaptor

USB adapter is a type of protocol converter which is utilized for converting USB data signals to and from other communications standards. Commonly, USB adaptors are acclimated to convert USB data to standard serial port data and vice versa.

Most commonly the USB data signals are converted to either RS 232, RS 485, RS 422 or TTL serial data. The older serial RS 423 protocol is infrequently utilized anymore, so USB to RS 423 adapters are less prevalent.

5. EXPERIMENTAL ANALYSIS

X and Y are matrices holding the training input and training output. The i -th data point is represented by the i -th row $X(i,:)$ and $Y(i,:)$. γ is the regularization parameter: for γ low minimizing of the complexity of the model is emphasized, for γ high, good fitting of the training data points is stressed. σ is the parameter of the kernel; in the common case of an RBF kernel, a large σ indicates a stronger smoothing.

The kernel_type indicates the function that is called to compute the kernel value (by default RBF_kernel). The data copies of the images are aligned in a separate database to prevent original images. To analyze large structure data with intense of detecting financial crime, it takes three expertise from the analyzing time.

Scattered Function Estimation Using LSVM Data Points (Blue .) And Estimation (Red)

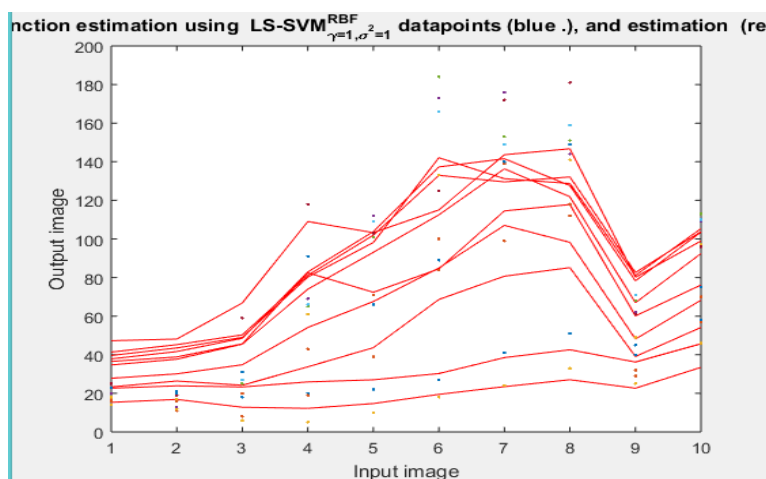


Figure 8. Scattered function estimation using LSVM for the tab

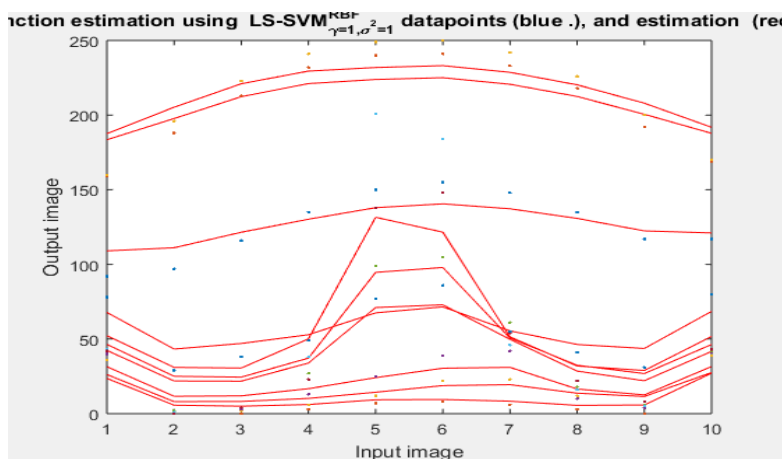
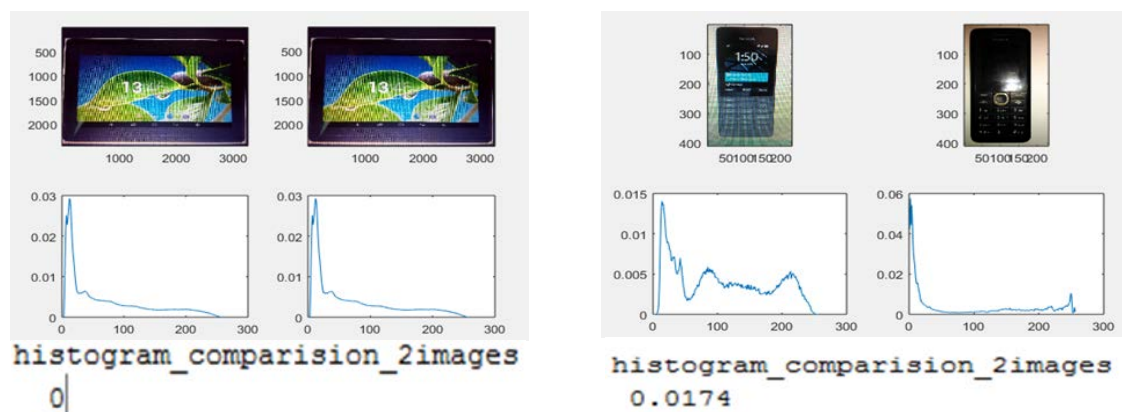


Figure 9. Scattered function estimation using LSVM for the phone

It performs histogram comparison between two images to calculate the difference between both. To calculate its difference it has to read two images and convert images to type double (range from from 0 to 1 instead of from 0 to 255). Reduce three channels [RGB] to one channel [grayscale] and then calculate the Normalized Histogram of Image 1 and Image 2.



1. Histogram comparison of original product.

2. Histogram comparison of duplicate product.

Figure 10. Histogram comparison of 2 images

5. CONCLUSION




Nowadays, in digital marketing has prone with tamper images. Our Proposed method is used to capture the image by a digital camera and to compare it with the structure database in the system to analyze its authenticity of an image. Tampered images are detected with the help of forensic data analysis technique. CIELAB was helpful in calculating the pixel value of an image. Principal Component Analysis is used to classify the Regression Analysis. Support Vector Machine classifies the relation between the images. Our future work is to use these techniques in various types of object. We can also investigate the global level performance of an object.

REFERENCES

- [1] Anil Dada Warbhe, Rajiv V. Dharaska, Vilas M. Dakar, " Digital Image Forensics An Affine Transform Robust Copy-Paste Tampering Detection", IEEE.
- [2] CAI-Ping Yan, Chi-Man Pun, and Xiao-Chen Yuan, " Adaptive Local Feature Based Multi-Scale Image Hashing for Robust Tampering Detection", Department of Computer and Information Science, University of Macau, Macau SAR, China, 2015 IEEE.
- [3] Chi-Man Pun, CAI-Ping Yan, and Xiao-Chen Yuan, "Image Alignment based Multi-Region Matching for Object-level Tampering Detection IEEE Transactions on Information Forensics and Security.
- [4] E. K. Ee, M. K. Johnson, and H. Farid, "Digital image authentication from JPEG headers," *IEEE Trans. Info Forensics, Security*, vol. 6, no. 3, pp. 1066–1075, Sept. 2011.
- [5] H. T. Sencar and H. Menon, " Digital Image Forensics: There is More to a Picture than Meets the Eye.", New York: Springer, 2013.
- [6] Jae Jeong HWANG, Kang Hyeon RHEE, "Gaussian Filtering Detection based on Features of Residuals in Image Forensics" The 2016 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future, IEEE.
- [7] Mustafa Al-Ani, and Fouad Khelifi, " On the Sensor Pattern Noise Estimation in Image Forensics: A Systematic Empirical Evaluation", TIFS.2016.2640938, IEEE Transactions on Information Forensics and Security.
- [8] Sepideh Azarian-Pour, Massoud Babaie-Zadeh, Amir Reza Sadri, "An Automatic JPEG Ghost Detection Approach for Digital Image Forensics", 2016 24th Iranian Conference on Electrical Engineering (ICEE).
- [9] Thanh Hai Thai, R emi Cogranne, Florent Retraint, Thi-Ngoc-Canh Doan, " JPEG Quantization Step Estimation and Its Applications to Digital Image Forensics", TIFS.2016.2604208, IEEE Transactions on Information Forensics and Security.
- [10] Amudhavalli P, "Sparse Based Robust Point Set Matching for Partial Face Recognition" *International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)*, ISSN 2454-9762 (Print), Vol. 2 Special Issue 6, March 2016.
- [11] Xiaoyu Chu, Matthew C. Stamm, and K. J. Ray Liu, "Information Theoretical Limit of Media Forensics: The Forensicability", IEEE Transaction On Information Forensic And Security, vol. 11, No 4, APRIL 2016.
- [12] Z. Fan and R. Queries, "Identification of bitmap compression History: JPEG detection and counter estimation," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 230–235, Feb. 2003.
- [13] Rajalakshmi.N. and Lakshmi Prabha, "Segregation of MRI Brain Image Using Hybrid Evolutionary Clustering Algorithm," *International Journal of Biomedical Engineering and Technology –Inderscience-* SNIP-0.374,SJR-0.034), 2014.

- [14] Pakutharivu P, Srinath M. V “Analysis of Fingerprint Image Enhancement Using Gabor Filtering With Different Orientation Field Values” *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 5, No. 2, pp. 427 ~ 432, February 2017.
- [15] Chawki Youness, El Asnaoui Khalid, Ouanan Mohammed and Aksasse Brahim “New Method of Content Based Image Retrieval based on 2-D ESPRIT Method and the *Gabor Filters*”, *TELKOMNIKA Telecommunication, Computing, Electronics and Control*. 2015; 15(2) pp. 313~320 DOI: 10.11591/telkonnika.v15i2.8377
- [16] Manar A. Mizherl , Mei Choo Ang , Ahmad A. Mazhar “A meaningful Compact Key Frames Extraction in Complex Video Shots”, *Indonesian Journal of Electrical Engineering and Computer Science* Vol. 7, No. 3, September 2017,DOI: 10.11591/ijeecs.v7.i3.pp818-829.

BIOGRAPHIES OF AUTHORS

	<p>Dr.P.Amudhavalli is an Associate Professor in Karpagam Academy of Higher Education, Department of Computer Science and Engineering, Coimbatore. She received a MCA degree in from University of Madras in 2003 and M.E.in Information and Communication from Anna University Chennai in 2008. She finished her doctoral degree in Computer Science and Engineering specializing in Cloud Computing. She has over 12 years of teaching experience. She has published more than 14 papers in international journals and conferences. Her area of interest includes Cloud Computing, Big Data, Image processing and Soft computing.</p>
	<p>Dr.N.Rajalakshmi is an associate professor in Karpagam Academy of Higher Education, biomedical department Coimbatore. She received a B.E degree in Electronics and Instrumentation Engineering from Bharathiar University in 1998 and M.E.in Medical Electronics from Anna University Chennai in 2002. She finished her doctoral degree in information and communication specializing in medical image processing. She has over 10 years of teaching experience. She has published more than 14 papers in international journals and conferences. Her area of interest includes Image processing, Soft computing, Medical image analysis.</p>
	<p>Mrs.K.S. Sindhu is an Assisstant Professor in Karpagam Academy of Higher Education, Department of Computer Science and Engineering, Coimbatore. She received a B.E degree in computer science and engineering, Avinashilingam University and M.E degree in computer science and engineering, Avinashilingam Institute For Home Science And Higher Education For Women. Her area of interest includes Image processing, Cloud Computing and Networking.</p>