International Journal of Engineering & Technology, 7 (4) (2018) 4393-4397

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 15368
Research paper

Intelligent systems to forgery image detection based on the edge characteristics using soft computing techniques

Divyashree S [1], Narendra V. G [2] *, Priya Kamath [2]

[1] *Student (IInd M.Tech.), Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, INDIA-576104*
[2] *Associate Professor Senior Scale, Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, INDIA-576104*
**Corresponding author E-mail: narendra.vg@manipal.edu*

Abstract

Detection of forgery in an image is an important aspect and very much essential in order to maintain the authenticity and the privacy of the information that could be interpreted from the image. In these days images could be manipulated using various tools and techniques, so its originality is lost, because these images may be used in medical diagnosis for monitoring the patients' health care, in smartphones for user authentication, in forensic investigations where images serve as legal evidence. In such cases it is necessary to get the correct details from an image or else it may be misused. Hence there is a need to detect the altered images and prevent the intended people in extracting the false details from the manipulated images. This research focuses on a technique to detect the forged images (tampered images) based on the edge characteristics, that are supported by the functions such as Canny edge detection and Hough transforms, to extract a feature vector for the Gaussian filtering detection. Better performance is obtained by using Random Forest classification for K-Means clustering, SGD classifica-tion for Density based clustering and hence detect the images that are forged.

Keywords: *Forgery Image Detection; Authenticity; Canny Edge Detection; Hough Transforms; Gaussian Filtering.*

1. Introduction

Digital image forensics is an upcoming area of image processing. One of its prominent tasks is image tampering detection. Images are treated as proofs in many real world situations, so image forgery could be defined as intentional manipulation of the digital images for malicious purposes.
Image Forgery is a concept where certain techniques are applied to the images, so that they could be manipulated to extract some useful, meaningful information from the digital images. The need for the detection of the forgery in images is driven by the fact that there is a need to maintain the integrity as well as authenticity of the digital images.
The importance of detecting image forgery is mainly because of its application such as:

- Authentication of images captured from CCD cameras (charge coupled device cameras)
- Fingerprint recognition
- Document authentication
- Authenticity of evidences
- Authentication of information available in an image
- Used in medical diagnosis for monitoring the patients' health care
- Useful in smartphones for user authentication
- In forensic investigations where images serve as legal evidence

In order to serve the above mentioned applications, firstly there is a need to get the correct details from an image or else the false information would be interpreted by the user. Hence there is a need to detect the tampered images and prevent the intended people in extracting the false details from the manipulated images.
The concepts of image processing would be applied to detect the tampered images. There are various forgery image detection techniques that could be used, based on the type of image forgery, i.e. active or passive approach. Active approach deals with digital watermarking and passive approach deals with the tampering in the digital images. Amongst the filtering techniques used to detect forged (i.e. tampered) images, this paper mainly focuses on a forgery detection technique for the Gaussian filtered images.

2. Literature survey

The author of the paper [1] has proposed a new scheme for the Gaussian filtering detection of the forged images, where in the feature is extracted by using the techniques such as Canny edge detection, Hough transforms. The author of the survey paper [2] gives description about the importance of image authenticity in the current world and also provides various techniques for image forgery detection. The author of the paper [3] presented robust Gaussian filtering detection method that uses frequency residual function as well as contrast ingredient of gray-level co-occurrence matrix to extract the feature vector. Junyu Xu et al. [4] demonstrate an approach to detect the altered images of Gaussian low-pass filtering using frequency residual function. Jae Jeong HWANG et al. [5] present a technique where the feature vector is obtained from the Gaussian filter residual (GFR) and Frequency transform residual (FTR), that is trained in the classifier for Gaussian filtering detection of the manipulated images. Anurag Das et al. [6] provide a technique for detecting the image splicing by using the inconsistencies of Gaussian blur, in order to test the authenticity of images. Dijana Tralic et al. [7] present a new database namely CoMoFoD database. It has 260 image sets with 200 images in small image category (512x512) and 60 images in large image category (3000x2000). Here every image group includes the original image, colored mask, the binary mask, forged image. Sagar Adatrao et al. [8] in their paper demonstrate the concepts of Image Pre-processing techniques and analyze the filters that could be used in image processing. B. Chitradevil et al. [9] provide an overview on the steps involved in image processing. It mainly deals with image pre-processing, image segmentation, feature extraction, image classification. Parameswaran Nampoothiri V et al. [10] briefs about the evolution of image forgery and their classification. The paper also describes about the forgery detection methods.

3. Methodology

Data Source: CoMoFoD image database is the dataset that is used for the implementation:

- It has 260 image sets with 200 images in small image category (512x512) and 60 images in large image category (3000x2000).
- Here every image set has original image, colored mask, binary mask, forged image.

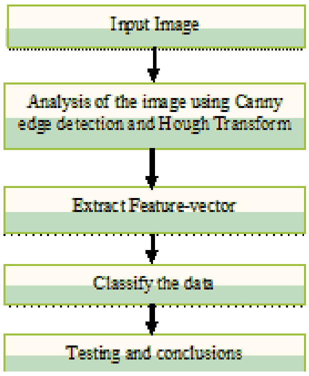

Fig. 1: Modules Involved in Forgery Image Detection.

Fig. 1 describes about the different modules involved in the project. Here the input image would go through the following steps of the image processing.

a) Image Pre-processing Techniques:

The image data may contain some of the errors pertaining to the geometry as well as the intensity value of the pixels of the image. Such images undergo pre-processing techniques such as noise filtering, contrast and edge enhancement, pseudo coloring, sharpening and magnifying.

The noise filtering is used to filter the unnecessary information i.e. noise from images. Here Gaussian filter is used to eliminate noise and detail and is efficient in smoothing image. It is not particularly effective at removing salt and pepper noise. Also median filtering and average filtering techniques are used to remove noise[11-19].

1 D (dimensional) function for Gaussian filter is:

$$G(x) = \frac{1}{\sqrt{2\pi}.\sigma} e^{-\frac{x^2}{2\sigma^2}} \tag{1}$$

Where, σ : Standard Deviation of distribution that has mean μ =0.

For 2D images, the 2D function for Gaussian filter is:

$$G(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

The Gaussian filter when applied on images, it is based on the two dimensional Gaussian distribution as a point-spread function, as shown in Fig. 2.

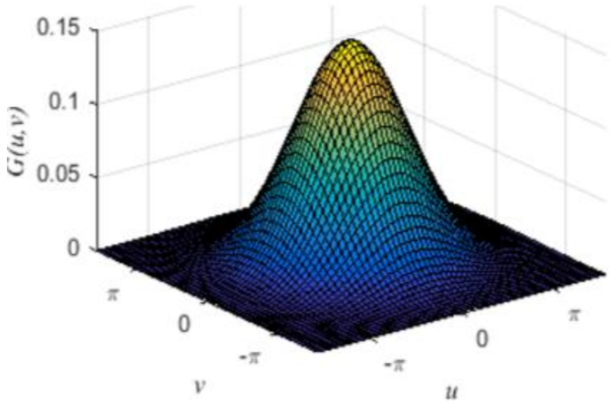

Fig. 2: Two Dimensional Gaussian distribution.

b) Image Segmentation Techniques:

Image segmentation is the process of partitioning a digital image into multiple segments, so that it provides an easier representation of an image which is meaningful and better to analyze.

It involves thresholding techniques, edge detection techniques. Here Canny Edge detection technique is used. It is an edge detection operator which uses a multistage algorithm to detect a wide range of edges in images [11-19].

The process involved in canny edge detection technique is described as follows:

- Initially Gaussian filter is used in order to remove noise from the images and hence smoothen them. Here gaussian kernel size of 5x5,

$$k = \frac{1}{159}\begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix} \tag{3}$$

- To determine the intensity of the image gradient, the Sobel edge detection operator is used to obtain the first-derivate in horizontal (G_x) and vertical (G_y) direction respectively.

The magnitude of the gradient is,

$$G = \sqrt{G_x^{\,2} + G_y^2} \tag{4}$$

Direction (angle) of the gradient is,

$$\theta = \tan^{-1}\left(\frac{G_x}{G_y}\right) \tag{5}$$

Where Gx = $\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$, Gy = $\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$

And θ is rounded to one of the four angles, namely 0°, 45°, 90° and 135°.

- Edge thinning technique such as non-maximum suppression is used in order to retain only the candidate edges i.e. “thin lines” of the image.
- After thinning, some edge-pixels may be caused due to noise, color variation. These are filtered out by applying double threshold, to retain the edge-pixels with the high gradient value.
- There may be some weak edge-pixels which actually form a part of the true edge (there are not noise). So to retain such pixels, connected component labeling technique is applied, by considering the weak-edge pixel and it’s 8-connected neighbor pixels.

Thus, segmentation of images helps to discriminate between objects and background and also separation among different regions.

c) Feature Extraction:

Feature extraction techniques are needed to carry out classification of the targets. The attributes (features) uniquely describe targets such as size, shape, composition, location.

Here Hough transform technique would be used to find the straight lines in an image. It's also used to find imperfect instances of objects within a particular class of shapes based on the voting procedure.

Any point (xi,yi) satisfies the line equation y=mx+c, its parametric representation is $\rho = x\cos\theta + y\sin\theta$. Thus line which is perpendicular has parameters, has distance "ρ" and degree "θ". The (ρ,θ) represents the angle-distance plane [11-19].

d) Feature Optimization:

Feature reduction involves data transformations techniques such as Principal component analysis. Feature vector would be obtained after applying Canny edge detection, Hough transforms to the input images[11-24].

e) Image Classification:

Image classification is a technique of labeling a pixel or a group of pixels based on the grey value. In classification many features are used for a set of pixels i.e., many images of a particular object are needed.

Here Support Vector Machine (SVM) classifier would be used. The feature vector is given as an input to the SVM classifier for training of GF (Gaussian Filtered) classification and it uses a grid search mechanism.

Random Forest classification technique is applied for which feature vector is provided as input. The results obtained from SVM classifier, Random Forest classifier are compared and analyzed as mentioned in the results section.

Next the already trained classifier model performs testing on set of test images. Finally measurements are estimated and necessary conclusions are drawn based on the performance of GFD (Gaussian Filtering Detection) technique [11-26].

4. Results

The Gaussian Filter is applied for the given input image; result is as shown in Fig. 3. Canny Edge Detection technique is applied on this Gaussian filtered image, to detect the edges as shown in Fig. 4. Next this is given as input to the Hough Transform function; resultant image is as shown in Fig. 5. Feature Vector is extracted from them for further processing [11-19].

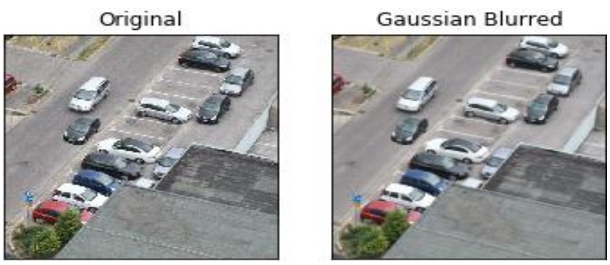

Fig. 3: Gaussian Blurring.

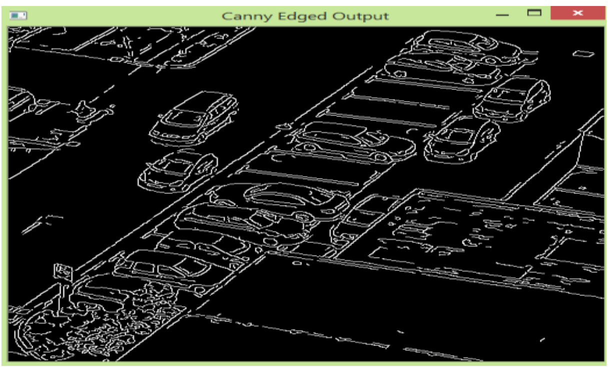

Fig. 4: Outcome of Canny Edge Detection.

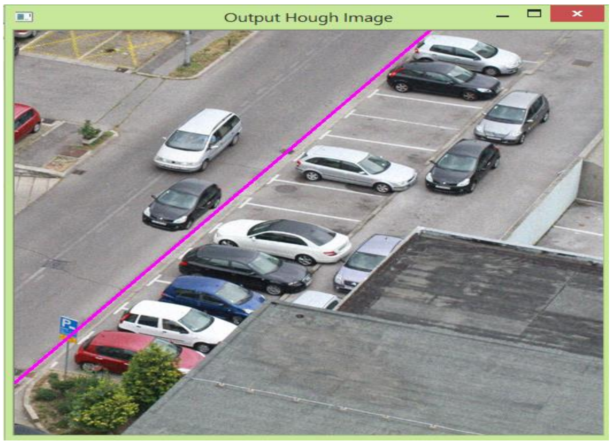

Fig. 5: Result of Hough Transform.

The constructed feature vector is given as input to the classifier model. So before classification, cluster the feature-vector data using k-Means Clustering Technique, Density Based Clustering in order to assign the label. SVM Classifier is used for classification, with cross validation of 10 folds, then it's TP Rate (True Positive Rate) and FP Rate (False Positive Rate) that is, sensitivity and specificity is above 0.9 and is depicted from the ROC curve and summary are shown in Fig. 6 and Fig. 7 respectively.

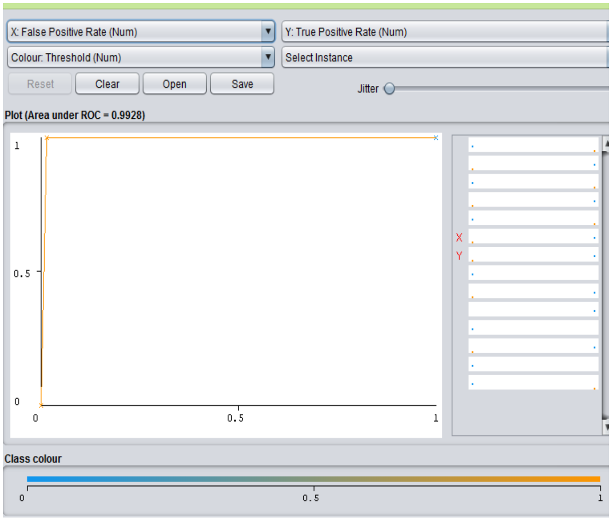

Fig. 6: ROC Curve for SVM Classifier.

```
=== Summary ===

Correctly Classified Instances       20793              99.4262 %
Incorrectly Classified Instances       120               0.5738 %
Kappa statistic                          0.988
Mean absolute error                      0.0057
Root mean squared error                  0.0757
Relative absolute error                  1.1959 %
Root relative squared error             15.4655 %
Total Number of Instances            20913

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.014    0.991      1.000    0.995      0.988    0.993     0.991     cluster1
                 0.986    0.000    1.000      0.986    0.993      0.988    0.993     0.991     cluster2
Weighted Avg.    0.994    0.009    0.994      0.994    0.994      0.988    0.993     0.991

=== Confusion Matrix ===

     a     b   <-- classified as
 12558     0 |   a = cluster1
   120  8235 |   b = cluster2
```

Fig. 7: Summary of Output for SVM Classifier.

When Random Forest Technique is used for classification, with cross validation of 10 folds, it is observed that comparatively better results are obtained here. Here sensitivity and specificity is above

0.99 and is depicted from the ROC curve and summary are as shown in Fig. 8 and Fig. 9 respectively.

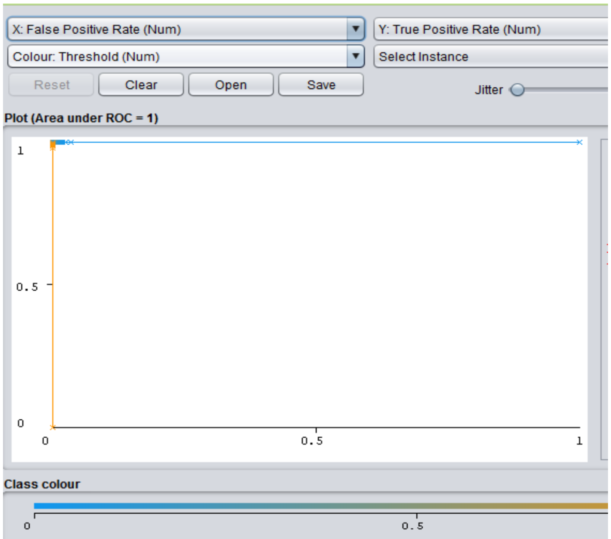

Fig. 8: ROC Curve for Random Forest.

```
=== Summary ===

Correctly Classified Instances       20877               99.8279 %
Incorrectly Classified Instances        36                0.1721 %
Kappa statistic                          0.9964
Mean absolute error                      0.003
Root mean squared error                  0.0358
Relative absolute error                  0.6281 %
Root relative squared error              7.3016 %
Total Number of Instances            20913

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.999    0.003    0.998      0.999    0.999      0.996    1.000     1.000     cluster1
                 0.997    0.001    0.998      0.997    0.998      0.996    1.000     1.000     cluster2
Weighted Avg.    0.998    0.002    0.998      0.998    0.998      0.996    1.000     1.000

=== Confusion Matrix ===

     a     b   <-- classified as
 12544    14 |     a = cluster1
    22  8333 |     b = cluster2
```

Fig. 9: Summary of Output for Random Forest.

The result analysis about the parameters associated with the different classifiers for the K-Means clustering and the Density Based Clustering (that is clustering is performed before the classification of data) is as shown in Table 1 and Table 2 respectively.

Table 1: Classification Models for K-Means Clustering

Classification Models	Training Set					Validation Set				
	Correctly Classified Instances %	Kappa Statistics	Root mean squared error	Sensitivity (TP Rate)	F-Measure	Correctly Classified Instances %	Kappa Statistics	Root mean squared error	Sensitivity (TP Rate)	F-Measure
Random Forest	99.8279	0.9964	0.0358	0.998	0.998	99.9412	0.9988	0.0231	0.999	0.999
SGD	99.7776	0.9953	0.0472	0.998	0.998	99.8909	0.9978	0.033	0.998	0.998
SMO	99.4262	0.988	0.0757	0.994	0.994	99.8993	0.998	0.0317	0.999	0.999
KStar	99.1767	0.9825	0.1203	0.992	0.992	99.4097	0.9881	0.1025	0.994	0.994
Naive Bayes	94.2323	0.8738	0.2041	0.942	0.941	97.7842	0.9552	0.1422	0.978	0.978
Random Subspace	92.3019	0.8313	0.2917	0.923	0.921	97.6051	0.9516	0.2641	0.976	0.976
IBk	72.3634	0.3163	0.5257	0.724	0.670	78.9973	0.5566	0.4582	0.790	0.775

Table 2: Classification Models for Density Based Clustering

Classification Models	Training Set					Validation Set				
	Correctly Classified Instances %	Kappa Statistics	Root mean squared error	Sensitivity (TP Rate)	F-Measure	Correctly Classified Instances %	Kappa Statistics	Root mean squared error	Sensitivity (TP Rate)	F-Measure
Random Forest	99.8297	0.9962	0.0344	0.998	0.998	99.9692	0.9965	0.0161	0.999	0.999
SGD	97.9465	0.9543	0.1433	0.979	0.980	99.9189	0.9984	0.0285	0.999	0.999
SMO	97.8850	0.9529	0.1454	0.979	0.979	99.7566	0.9951	0.0493	0.998	0.998
KStar	99.3376	0.9852	0.1169	0.993	0.993	99.3621	0.9871	0.1025	0.994	0.994
Naive Bayes	94.2276	0.8656	0.1916	0.942	0.941	99.2614	0.9851	0.1233	0.993	0.993
Random Subspace	95.4105	0.8946	0.2613	0.954	0.953	99.3481	0.9868	0.2594	0.993	0.993
IBk	75.6565	0.3414	0.4934	0.757	0.707	79.5233	0.5698	0.4524	0.795	0.781

The various classifiers analyzed here are Random Forest, Support Vector Machine-SVM (SMO), SGD, KStar, Naive Bayes, Random Subspace, IBk Classifier [25, 26]. The results of these classifiers are analyzed based on kappa statistics, Root means squared error, Sensitivity (TP Rate), F-measure, ROC area characteristics.

When K-Means clustering is applied, it is observed that these characteristics have better value for the Random Forest classifier than others. Hence better performance could be achieved using Random Forest Classification Technique.

When Density Based Clustering is applied, it is observed that above mentioned characteristics has better value for the SGD classifier than others. Hence better performance could be achieved using SGD Classification Technique.

Hence, forged image, along with the object that is forged is detected, and is as shown in Fig. 10.

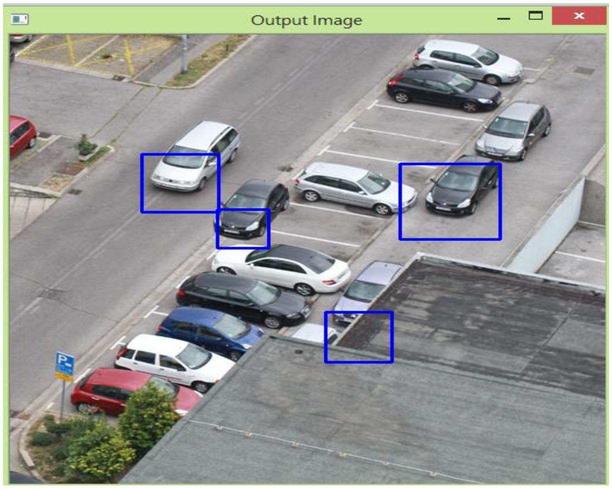

Fig. 10: Forged Object is Detected.

5. Conclusion

In this paper, presents Gaussian Filtering detection method, which is based on canny edge detection and Hough Transform functions. By comparison with the previous schemes, the feature vector length of the proposed scheme is about half of the other schemes. Even so, the measured performance is evaluated as the same or excellently more.

Better performance is obtained by using Random Forest classification for K-Means clustering, SGD classification for Density based clustering and hence detect the images that are forged. Finally, proposed models can also be applied to solve different forensic problems.

Acknowledgement

The authors are much indebted to the Department of Computer Science and Engineering. Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India-576104, for providing excellent lab facilities that make this work possible.

References

[1] Kang Hyeon Rhee, "Forgery Image Detection of Gaussian Filtering by Support Vector Machine using Edge Characteristics", International Conference on Ubiquitous and Future Networks (ICUFN), IEEE, 2017.
[2] Mr. Arun Anoop M, "Image forgery and its detection: A survey", IEEE Sponsored 2J/d International Conference on Innovations in Information, Embedded and Communication systems (ICIIECS), IEEE, 2015.
[3] Kang Hyeon, "Gaussian Filtering Detection using Band Pass Residual and Contrast of Forgery image", IEEE, 2016.
[4] Junyu Xu, Yun Ling, Xuan Zheng, "Forensic Detection of Gaussian Low-pass Filtering in Digital Images", 8th International Congress on Image and Signal Processing (CISP), IEEE, 2015. https://doi.org/10.1109/CISP.2015.7407990.
[5] Jae Jeong HWANG, Kang Hyeon RHEE, "Gaussian Filtering Detection based on Features of Residuals in Image Forensics", IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future, 2016.
[6] Anurag Das, Abhishek Medhi, Ram Kumar Karsh, Rabul Hussain Laskar, "Image Splicing Detection using Gaussian or Defocus Blur", IEEE International Conference on Communication and Signal Processing, 2016. https://doi.org/10.1109/ICCSP.2016.7754350.
[7] Dijana Tralic, Ivan Zupancic, Sonja Grgic, Mislav Grgic, "CoMoFoD - New Database for Copy-Move Forgery Detection", International Symposium ELMAR, 2013.
[8] Sagar Adatrao, Mayank Mittat, "An Analysis of Different Image Preprocessing Techniques for Determining the Centroids of Circular Marks Using Hough Transform", 2nd International Conference on Frontiers of Signal Processing, 2016.
[9] B. Chitradevil, P.Srimath, "An Overview on Image Processing Techniques", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 11, 2014.
[10] Parameswaran Nampoothiri V, Dr. N Sugitha, "Digital Image Forgery - A threaten to Digital Forensics", International Conference on Circuit, Power and Computing Technologies [ICCPCT], IEEE, 2016.
[11] Weibing Liu, Deqiang Han, Yi Yang, "A Novel Weighted SVM Based on Theory of Belief Functions", 20th International Conference on Information Fusion (ISIF), IEEE, 2017.
[12] Chi-Man Pun, Xiao-Chen Yuan, Xiu-Li Bi, "Image Forgery Detection using Adaptive Over-segmentation and Feature Point Matching", IEEE Transactions on Information Forensics and Security, Vol. 10, No. 8, 2015.
[13] Anushree U. Tembe, Supriya S. Thombre, "Survey of Copy-paste Forgery Detection in Digital Image Forensic", International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), IEEE, 2017. https://doi.org/10.1109/ICIMIA.2017.7975613.
[14] Ibrahim Mesecan, Ihsan Omur Bucak, "Searching the Effects of Image Scaling for Underground Object Detection using KMeans and KNN", UKSim-AMSS 8th European Modelling Symposium, IEEE, 2014.
[15] Chun Guan, Kevin Kam Fung Yuen, Qi Chen, "Towards a Hybrid Approach of K-Means and Density-Based Spatial Clustering of Applications with Noise for Image Segmentation", IEEE International Conference on Internet of Things (iThings), IEEE Green Computing and Communications (GreenCom), IEEE Cyber Physical and Social Computing (CPSCom) and IEEE Smart Data, 2017.
[16] Zeyu Li, Oriol Vinyals, Harlyn Baker, Ruzenqa Bajcsy, "Feature Learning using Generalized Extreme Value Distribution based K-Means Clustering", 21st International Conference on Pattern Recognition(ICPR), IEEE, 2012.
[17] Hua Yang, Jipu Gao, Changbao Xu, Zheng Long, Weigang Feng, Shaohua Xiong, Shuaiwei Liu and Shan Tan, "Infrared Image Change Detection of Substation Equipment in Power System using Random Forest", 13th international Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), IEEE, 2017.
[18] Vrushali Y. Kulkarni, Manisha Petare, P.K. Sinha, "Analyzing Random Forest Classifier with Different Split Measures", International Conference on Soft Computing, Advances in Intelligent Systems and Computing 236, Springer, 2014.
[19] M. Octaviano Pratama, Aniati Murni Aryamurthy, "Automatic Land Cover Classification of Geotagged Images using ID3, Naïve Bayes and Random Forest", ICACSIS, IEEE, 2017.
[20] Zhi Liu, Bo Tang, Xiaofu He, Qingchen Qiu, Feng Liu, "Class-Specific Random Forest with Cross-Correlation Constraints for Spectral-Spatial Hyperspectral Image Classification", GeoScience and Remote Sensing Letters, Vol. 14, No. 2, IEEE, 2017.
[21] Amanpreet Kaur, "A Review Paper on Image Segmentation and its various Techniques in Image Processing", International Journal of Science and Research(IJSR), Vol. 3, Issue 12, 2014.
[22] Sharan Agrawal, Shivam Rana, Tanvir Ahmad, "Random Forest for the Real Forests", second International Conference on Computer and Communication Technologies, Advances in Intelligent Systems and Computing 381, Springer, 2016.
[23] Oliveira Julio. R, Soares, Leonardo. B., Costa, E.A.C, Bampi, Sergio, "Energy-Efficient Gaussian Filter for Image Processing using Approximate Adder Circuits", IEEE, 2015.
[24] Nidhi, "Image Processing and Object Detection", International Journal of Applied Research, 2015.
[25] Ahmad Nor Ikwan Masazhar, Mahanijah Md Kamal, "Digital Image Processing Technique for Palm Oil Leaf Disease Detection using Multiclass SVM Classifier", 4th IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), IEEE, 2017.
[26] Neetu Chahal, Anuradha, "A Study on Agricultural Image Processing along with the Classification Model", International Advance Computing Conference (IACC), IEEE, 2015.