An Efficient Forgery Image Detection Method using Hybrid Feature Extraction and Multiclass SVM

Clara Shanthi .G, V. Cyril Raj

Abstract: ***The advancement of image editing software tools in the image processing field has led to an exponential increase in the manipulation of the images. Subjective differentiation of original and manipulated images has become almost impossible. This has kindled the interest among researchers to develop algorithms for detecting the forgery in the image. Image-Splicing, Copy-Move and Image Retouching are the most common image forgery techniques. The existing methods to detect image forgery has drawbacks like false detection, high execution time and low accuracy rate. Considering these issues, this work proposes an efficient method for detection of image forgery. Initially, bilateral filter is used to remove the noise in pre-processing, Chan-Vese Segmentation algorithm is used to detect the clumps from the filtered image utilizing both intensity and edgeinformation, followed by hybrid feature extraction technique.***

Hybrid feature extraction technique comprises of Dual Tree Complex-Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA) and Gray-Level-Co-Occurrence Matrix (GLCM). The DWT has dual-tree complex wavelet transform with important properties, it is nearly shift invariant and directionally selective in two and higher dimensions. Principal Component Analysis (PCA) finds the eigenvectors of a covariance matrix with the highest eigenvalues and uses these values to project the data into a new subspace of equal or less dimensions. Gray-Level-Co-Occurrence Matrix (GLCM) extracts the Feature values such as energy, entropy, homogeneity, standard deviation, variance, contrast, correlation and mean. Classification is done based on the texture values of training dataset and testing dataset using Multi Class-Support Vector Machine (SVM). The performance analysis is done based on the True positive, False positive and True negative values. The experimental results obtained using the proposed technique shows a better performance compared to the existing KNN classifier model.

Keywords: Chan-Vese algorithm, Hybrid Feature Extraction, Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA), Gray-Level Co-Occurrence Matrix (GLCM), Multi-Class Support Vector Machine (Multi-Class SVM), Image Forgery Detection

I. INTRODUCTION

With the development of digital photography, pictures can now be readily modified for anyone. With the growing digital imaging apps, various kinds of software instruments have been implemented for image processing.

Image forgery detection in both academia and the professional world has become a increasing area of studies. Convincingly manipulated pictures are commonly distributed on social media platforms and can quickly spread to groups that think they are true. There are many techniques for detecting forgery in digital pictures. Furthermore, with the significant rise in the use of social media, people would also significantly benefit from being able to identify forgeries in pictures. To detect these forgeries of pictures, some typical techniques used to manipulate pictures must be understood.

Photographs have been used to document for centuries and used in court as evidence [1]. Retouching, picture splicing and copy-moving forgery are few major kinds of picture forgery methods. However, strong software for digital image editing makes it easy to execute picture modifications[2]. Retouching of images deemed the less damaging type of forgery of images. This does not alter an picture considerably, but decreases or enhances the image's characteristic. Splicing of images is more aggressive than retouching of images. It composes a combination of two or more images to produce a false picture. The forgery of moving copies is very comparable to the splicing of the picture. Both methods change the picture in another picture. This forgery copies portion of an initial picture and moves to another image's required place.

In the role of malware, the level of image forgery on the internet has increased, making it possible for any user to upload, download or share images. When an intruder modifies the initial picture to produce a manipulated image for misuse, if one picture is tampering[3].Images are seen on social media, Internet websites like YouTube has been tampered their authenticity can no longer always be taken for granted [4]. Because of the constantly changing techniques used in picture forgery, no algorithm is ideal, and it is difficult to detect any kind of forgery with one algorithm. However, the suggested work will not only enable the user to decide which algorithm is best for them, but will also provide an insight into the most enceinte algorithm to detect fake picture. The proposed method has the greatest detection rate when run on the group of sample images.

Image Composition

The field of image composition is inventing the ways to improve the methods used to alter an image for practical purposes and like to increase the beauty of the image.

Revised Manuscript Received on June 22, 2019.

Clara Shanthi.G, Research Scholar CSE Department, Dr.M.G.R. Educational and Research Institute University, Tamil Nadu, India

V.Cyril Raj, Professor CSE & IT, Dr.M.G.R. Educational and Research Institute University, Tamil Nadu, India

But it also provides way for fraudsters to manipulate the original picture with the image composition tools.

Image forgery

In recent times, many cases of image forgery have been reported with the advancement of technology and Image Processing Software. Important information from an image can now be added, changed or removed easily without leaving any traces of such manipulation. Various methods have been used for altering an image. A brief discussion of the techniques used for image distortion is given below.

Image Retouching

Image retouching is a popular technique used for making the image to look more attractive such as removing wrinkles from the image, removing sweat from the image, removing darkness from the image, etc. This method is used extensively in the media industry and it is seen as a desirable method as it enhances the image by removing unwanted features in the image.

Copy-move forgery

Copy-move forgery technique is also referred to as cloning technique. In this technique, parts of an image are copied and pasted to a different location in the same picture. There is no significant change in the visible features of the tampered image. This method is a challenging task for researchers to detect the forgery in the image as there is not much change in the texture properties like colour, noise and texture of the forged image.

Image Splicing

Image-Splicing technique needs some operations to be done on the spliced regions of the image. The spliced image is created by using image parts from one or more images. To make the changes in the image invisible, some geometric transformations like stretching, flipping, rotating, skewing, scaling, etc. are done in the spliced regions.

Forgery Detection Techniques

Government have introduced e-Government services to reduce the use of papers and other documentation works. This initiates citizens or users to upload their certificates and other important documents in digital format. Therefore, it becomes a necessity for forensic department to validate the credibility of digital images. Nowadays, it is becoming a real challenge to detect the authenticity of an image due to the advancement in image tampering techniques. This challenge provoked interest among researchers to work on the techniques to detect the tampered image. Forensic techniques to detect image forgery are mainly classified into two categories – active and passive techniques. New techniques have been designed to figure out the forgery in images. Several algorithms have been proposed by researchers to detect the forgeries in images like block-based forgery detection algorithms and key-point based on forgery detection algorithms. However, these techniques are proved to be computationally inefficient for detecting the duplication in digital images.

II. LITERATURE REVIEW

Cao, Y. et. al. (2011) designed a copy/move forgery detection technique based on the Discrete Cosine Transform (DCT). To represent specific blocks, they used DCT coefficients for each block which are selected. The main drawback of their method is it detects homogenous blocks wrongly and it is much prone to noise and blurring. [5]

Fridrich et. al. (2003) detected copy-move forgery by matching the blocks with quantized DCT coefficients. The drawbacks of their technique are false identification of some copied areas and the small copied images are of less reliability. [6]

A DWT based image forgery detection technique is designed by Myna et. al. (2007) which first identifies the matching blocks and then utilizes phase correlation for detecting the copied regions. The limitation of their technique is that if the copied region is slightly rotated or scaled, it gives very poor results. [7]

Pixel matching and DWT techniques are used by Zhang et. al. (2008) to minimize the dimensions. To identify the copied and pasted regions, phase correlation is utilized. However, performance of their technique relies on the scene of the copied and moved image. [8]

Chi-Man Pun et al. (2015) have used segmentation technique to detect the forgery in images using Simple Linear Iterative Clustering (SLIC) algorithm Thus, the results depict that various copy-move forgery techniques such as blurring, JPEG compression, brightness adjustment, slight image rotation can be obtained by overlapping the blocks of HOGN.

To identify the combination of post-processing working by single technique, a copy-move forgery detection method is designed by Kaur et al (2015) using DCT and SIFT approaches. Besides, to rotate and scale the image, a key point based SIFT method is used. Thus, the mixture of SIFT and DCT approaches has the capability to determine the forgery under post-processing of rotation, Gaussian noise, scaling, JPEG compression. Thus, a proficient forgery detection is possible, but detection method has many false matches.

Based on the Gray Level Co-occurrence Matrices (TF-GLCM), Splicing Image Forgery detection was explained by the researchers Xuan Jing Shen et. al. (2016) by using textural features. In TF-GLCM, the GLCM is measured using the Differential Blocks Discrete Cosine Transform (DBDCT) arrays that is used to get the textual information and the spatial relationship sufficiently among the image pixels.

III. PROPOSED METHODOLOGY

The proposed system comprises of a novel image forgery detection technique that involves efficient filtering mechanism followed by hybrid feature extraction technique and the classification. The framework of the proposed technique is given in the figure 1.

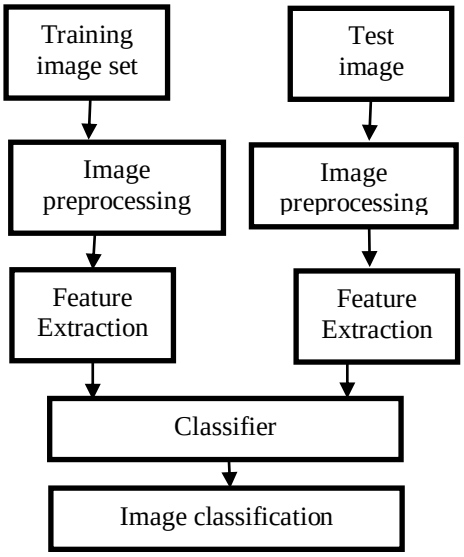

Fig. 1 Framework of proposed methodology

Structure of image forgery detection

The steps followed in the forgery detection of image is given in the fig.2 below.

ImagePre-Processing
Segmentation
Feature Extraction
Classification

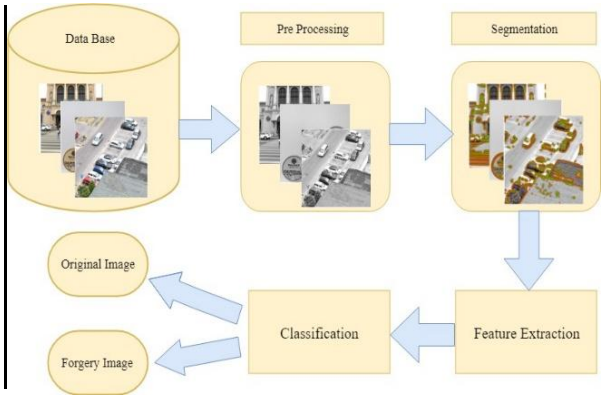

Fig. 2 General structure of image forgery detection

Image Pre-Processing

In preprocessing the RGB image (Color image) will be converted into a Grayscale image. Generally, Combination of Red, Green and Blue plane form the colour image in 3 Dimensions. To reduce the complexity RGB image is converted to 2D grayscale format. We use weighted method to convert an RGB image to a grayscale image.

The grayscale image is obtained by the following equation.

New Grayscale Image = ((0.3*R) + ((0.59*G) + (0.11*B)) (1)

The grayscale image must be resized into a nominal image range of 256 X 256-dimension rate. The main reason for resizing the image is to convert all the images into same size. The final step of the image pre-processing stage, a bilateral filter is utilized for eliminating the noise present in the images. We use bilateral filter because it replaces the intensity of every pixel by a weighted average of intensity values from nearby pixels. Bilateral filter has an added advantage of preserving the sharp edges along with the energy of the image

A low pass domain filter image is derived as I(x):

Pixel location (x,y) and output I(x,y) of bilateral filter measured by the following equation:

$$I'(X,Y) = \in y_{\in N(X)} e^{\frac{-(i(y)-i(x))^2}{2\sigma_r^2}} e^{\frac{-(i(y)-i(x))^2}{2\sigma_D^2}} i(x,y) \quad (9)$$

Where &σr - spatial & intensity of fall-off of weights control parameter, N(x) - pixel of I(x), σd - geometric spread parameter, used to select the required low pass filter. The bilateral filter helps to understand its entire variants.

Segmentation

The image is partitioned into multiple segments or set of pixels. The main intention of segmentation is to simplify and to modify the depiction of an image to analyze. To perform image segmentation, a multi-level set algorithm isused. The multi-level set algorithm works in three stages. The first step of Multi-level set algorithm is to discover the clusters from the filtered image utilizing multi-level set technique, using both intensity and edge data. In the second step, intensity-based level set is applied to recognize the middle regions.

The intensity-based level set is applied several times until it reaches final edge in the image. In the present process Chan-Vese Algorithm is implemented for multilevel set segmentation. Chan-Vese Active Contour Model is a strong and versatile technique that can segment many image kinds, including some that would be hard to segment.

Feature Extraction

Textural Feature extraction enhance the speed and efficiency of supervised learning method. In the proposed system for detecting the image forgery, a new hybrid feature extraction technique is employed. The Hybrid Feature Extraction technique comprises of Dual Tree Complex-Discrete Wavelet Transform (DTC-DWT), Principal Component Analysis (PCA) and Gray-Level Co-Occurrence Matrix (GLCM). The Hybrid Feature Extraction Technique is given in equation form as below.

HFE = DTC-DWT + PCA + GLCM (10)

Three steps are involved in the stage of feature extraction.

In the first step, wavelet coefficients are extracted by using wave transorm technique which works only on the greyscale images.Essential coefficients are selected by employing PCA technique. With the combination of DTC-DWT and PCA techniques, Efficienttexture-feature extraction of image is done by GLCM with spatial difference of the pixel.

Dual Tree Complex Dwt

Discrete Wavelet Transform (DWT) is a popular transform for its susceptibility to reduce most of the image information into a compact size in proportion to the image size.

The image information is reduced to the lowest energy sub-band so that it can be exposed to forensic analysis with less complexity.

A dual-tree complex DWT of a signal x is implemented using two critically sampled DWTs in parallel on the Data as shown in the Fig.3

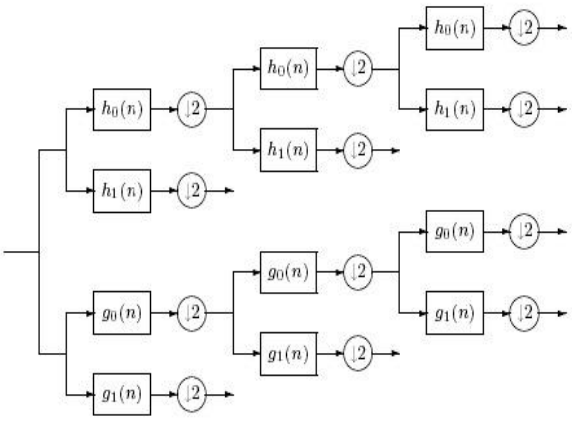

Fig. 3 Dual-tree Complex

The transform is 2-time expansive as it provides 2N DWT coefficients for an N-point signal. If the upper and lower DWT filters are the same, then there is no benefit. If the filters are designed, however, is a particular way, then the upper DWT sub-band signals can be interpreted as the real part of a complex wavelet transformation, and the lower DWT sub-band signals can be interpreted as the imaginary part. Equally, the wavelet associated with the upper DWT may be an approximate Hilbert transformation of the wavelet associated with the lower DWT for specially designed filter sets. The dual-tree complex DWT is almost shift-invariant when built in this manner. In addition, the DWT dual-tree complex can be used to introduce 2D wavelet transforms where each wavelet is oriented, which is particularly helpful for processing images. The DWT dual-tree complex exceeds the critically sampled DWT for apps such as picture de-noise and improvement.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical operation to turn a set of possible correlated variables observations into a set of linearly uncorrelated variables values using orthogonal conversion. PCA discovers the immediate lower-dimensional data illustration. PCA's goal is to safeguard the recreated data distinction. The primary reason for using the PCA in the suggested method is to decrease the wavelet coefficients ' dimensionality.

Gray-Level Co-Occurrence Matrix (GLCM)

The gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix, is a statistical technique for examining texture that considers the spatial connection of pixels. The combination of DTC-DWT and PCA methods increases the effectiveness of GLCM in extracting texture features. The GLCM functions characterize an image's texture by calculating how often pairs of pixels with specific values occur in an image and in a specified spatial relationship, creating a GLCM, and then extracting statistical measurements from the matrix.

$$G = \begin{bmatrix} h(1,1) & \cdots & h(1,Bg) \\ \vdots & \ddots & \vdots \\ h(Bg,1) & \cdots & h(Bg,Bg) \end{bmatrix} \quad (11)$$

Where h (i, j) shows the relative frequencies;

I and j- pixel couple values of image. h(i, j) can be derived as follow:

$$h(i,j) = \#\{(p_1,q_1),(p_2,q_2) \in A * B \setminus f(p_1,q_1) = i, f(p_2,q_2) = J\} \quad (12)$$

Where, # - number of elements in a set.

The diverse arrangements of separation (indicated by d) and edges (meant by θ) between the two pixels can impact the method for computing the quantity of the pixel couples in GLCM. In TF-GLCM, the separation is set at 1 and the edges are 0°, 45°, 90°, and 135° [11]. What's more, these two parameters (d and θ) would bring about four types of grouping. Subsequent to being standardized, for example, the four GLCM removed from the horizontal contrast coefficient array can be given by

$$G_{d=1,\theta=0^*}(F_H(i,j)), \ldots . G_{d=1,\theta=135^*}(F_H(i,j)) \quad (13)$$

Gray-Level-Co-Occurrence Matrix (GLCM) extracts the Feature values such as energy, entropy, homogeneity, standard deviation, variance, contrast, correlation and mean. These extracted feature values are used in the training and testing phase of the Multi-SVM classifier

Classification

Classification is the final stage of the image forgery detection process. This is supervised learning method. In supervised learning, adequate number of values are given, and the target variable is well known, whereas in unsupervised learning either the target variable is not known, or it has been observed only for small number of data.

Support Vector Machine (SVM) is one of the supervised learning models that has higher classification efficiency compared to other classifier models. But, the utilization of SVM is restricted due to higher training time required for large sets of data. Hence to acquire reduced dimensional data, feature selection techniques are integrated with SVM.

Multi-Support Vector Machine (M-SVM) classifier is employed to classify the original and forgerydetection. MC-SVM classifier also enhances the accuracy of detecting forgery. The classification of image frames is done based on the test values and training values.

MC-SVM classifier is constructed by f 1, f 2,..$f_{M,\ each}$ set is trained by each class. Multi-class maximum output is derived from $g^i(x)$:

$$g^j(x) = \sum_{i=1}^{m} y_i \alpha_i^j k(x, x_i) + b^j \quad (14)$$

The detection of forgery image was done by using M-SVM classifier on systematic basis by planning a simple process containing of two phases, namely training phase and testing phase.

IV. RESULTS AND DISCUSSION

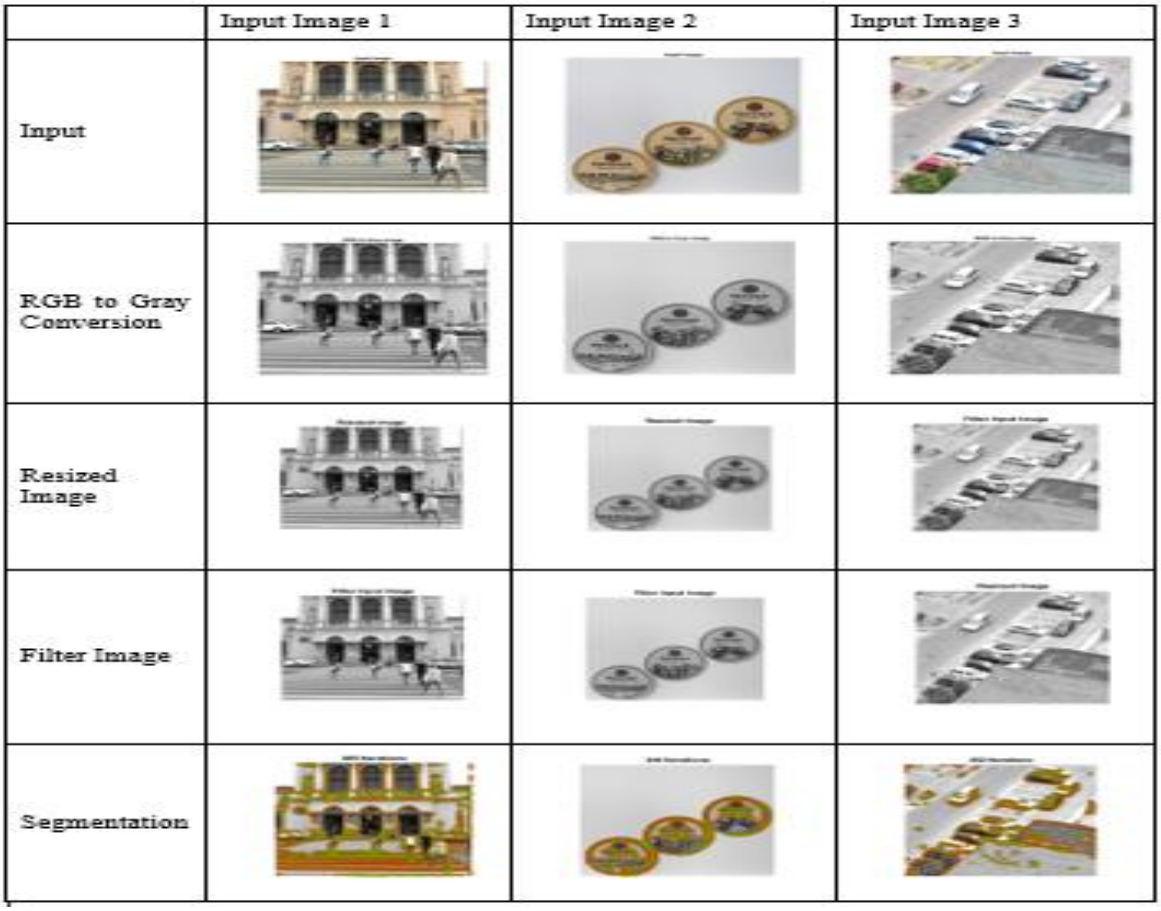

Fig. 4 Resultant images from every stage of image forgery detection

Table. 1 Features extracted for 3 images

Feature	Image1	Image2	Image3
Std Deviation	0.0707	0.0707	0.0707
Entropy	4.0336	5.2155	3.9841
Mean	153. 5348	187. 9546	178. 6338
Correlation	0.0157	0.0382	0.0108
Contrast	0.0851	0.0959	0.0864
Energy	0.8479	0.8516	0.8437
Homogeneity	0.9604	0.9610	0.9591
Variance	0.0050	0.0050	0.0050

The above table shows the feature values extracted for the three input images. It is seen from the above table that standard deviation and variance is same for all the three images 0.0707 and 0.0050 respectively. The values of other extracted features are described above.

With the increase in forgery testing image, the specificity and precision have fluctuated. The table2 shows the performance of the existing system and table 3 shows the performance analysis of the proposed system. From the below tables it clearly, shows when the forgery images increase, the parameters are varied. The tables depict the improvement of the proposed methodology in terms of accuracy.

Table. 2 Results obtained by KNN method

Ratio	accuracy	sensitivity	specificity	precision
1:3	58.33%	62.78%	45%	77.40%
1:2	57.78%	64.17%	45.00%	70.00%
1:1	54.17%	95%	52.29%	95%

Table. 3 Results obtained by Multi-SVM method

Ratio	accuracy	sensitivity	specificity	precision
1:3	73.75%	94.44%	11.67%	76.23%
1:2	67.22%	95%	11.67%	68.26%
1:1	56.67%	66.67%	55.36%	66.67%

Accuracy is the ratio of the correct assessments to the total assessments number. The accuracy obtained for the proposed classifier for different input ratios is 0.7375, 0.6722, and 0.5667. Figure 9 shows the accuracy comparison of the proposed Multi-SVM classifier and KNN classifier.

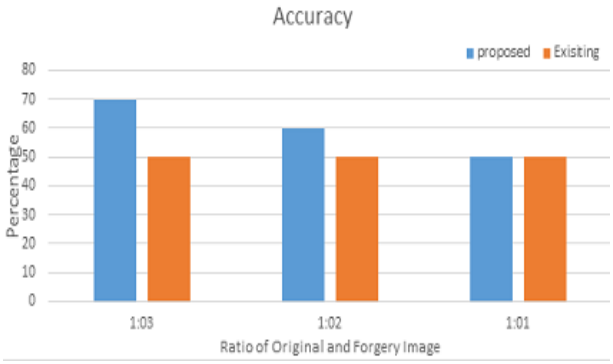

Fig. 5 Accuracy comparison

Sensitivity is the ratio of the number of true positive assessments to the overall positive assessments. The sensitivity obtained for the proposed classifier for different input ratios is 0.9444, 0.9500, and 0.6667. Figure 10 shows the sensitivity comparison of the proposed Multi-SVM classifier and KNN classifier.

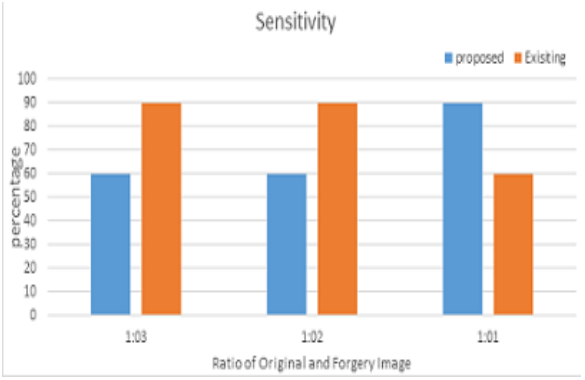

Fig. 6 Sensitivity comparison

Specificity is the ratio of the number of true negative assessments to the overall negative assessments. The specificity obtained for the proposed classifier for different input ratios is 0.1167, 0.1167, and 0.5536. Figure 11 shows the specificity comparison of the proposed Multi-SVM classifier and KNN classifier.

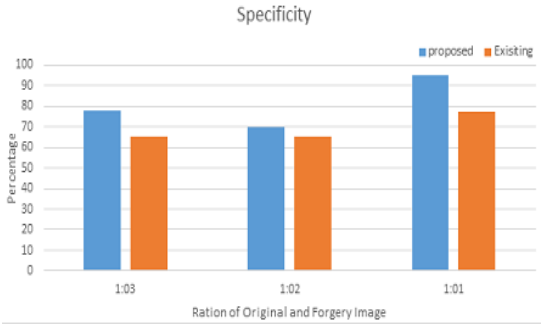

Fig. 7 Specificity comparison

Precision defines how many positively classified results were relevant. It is defined as the ratio of the number of true positive assessments to the total number of positive assessments. The precision obtained for the proposed classifier for different input ratios is 0.7623, 0.6826, and 0.6667. Figure 12 shows the precision comparison of the proposed Multi-SVM classifier and KNN classifier.

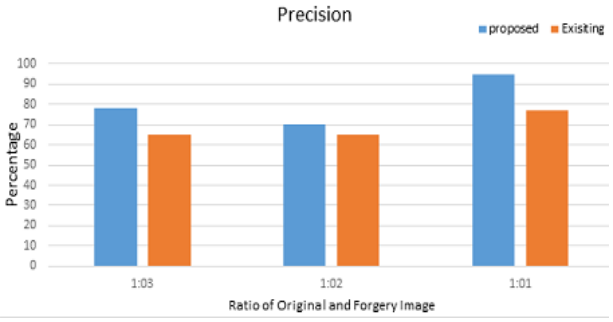

Fig. 8 Precision comparison

KNN classifier is thus a lazy learner without any building models. It keeps all the training samples in memory. For a test sample, it searches all the K nearest neighbors. K nearest neighbors is a small subset of the entire training sample space. Hence for improved performance of classification, Multi-SVM classifier is employed in work. Furthermore, the above-given comparison graphs prove that the performance of the proposed Multi-SVM classifier is superior compared to the existing KNN model of classification.

V. CONCLUSION

With the advancement of image processing field, image forgery detection is the one of the vital research areas in the department of forensics. The proposed method of detecting the forgery in images involve Pre-Processing the input images, where it is converted to a grayscale image, resized and denoised using Bilateral filter. Then the Pre-Processed image is segmented using the Chan-Vese algorithm, followed by the Feature extraction method, which involves Dual Tree DWT-PCA-GLCM techniques. Using the extracted features, the forgery images are recognized by Multi-class SVM classifier. The proposed Multi-class SVM classifier helps to achieve better results compared to the existing KNN method of classification. Evaluation of the performance in detecting forgery of images was done by computing parameters such as accuracy, sensitivity, specificity, and precision. Also,the proposed technique is better than the existing technique. However, the precision of the Multi-Class SVM is comparatively less. In the future, the performance of classification can be improved using neural networks.

REFERENCES

1. Mankar, S. K., &Gurjar, A. A. (2015). Image forgery types and their detection: A review. International Journal of Advanced Research in Computer Science and Software Engineering. 5(4).
2. de Carvalho, T.J.Riess, C. ; Angelopoulou, E.; Pedrini, H., “Exposing Digital Image Forgeries by Illumination Colour Classification” Information Forensics and Security, IEEE Transactions on,June 2013.
3. A. Rocha, W. Scheirer, T. Boult, S. Goldenstein, Vision of the Unseen: Current Trends and Challenges in Digital Image and Video Forensics", ACM Computing Surveys (CSUR), Volume 43Issue 4, October 2011, Article No. 26,
4. Saurabh Upadhyay, Sanjay Kumar Singh, “Video Authentication: Issues and Challenges" in IJCSI International Journal of Computer Science

Issues,Vol. 9, Issue 1, No 3, January 2012 ISSN (Online):1694-0814.
5. Cao, Y., Gao, T., Fan, L., & Yang, Q. (2012). A robust detection algorithm for copy-move forgery in digital images. *Forensic science international*, *214*(1-3), 33-43.
6. Fridrich, A. J., Soukal, B. D., &Lukáš, A. J. (2003). Detection of copy-move forgery in digital images. In Proceedings of Digital Forensic Research Workshop.
7. Myna, A. N., Venkateshmurthy, M. G., & Patil, C. G. (2007, December). Detection of region duplication forgery in digital images using wavelets and log-polar mapping. In International Conference on Computational Intelligence and Multimedia Applications(ICCIMA 2007) (Vol. 3, pp. 371-377). IEEE.
8. Zhang, J., Feng, Z., &Su, Y. (2008, November). A new approach for detecting copy-move forgery in digital images. In 2008 11th IEEE Singapore International Conference on Communication Systems (pp. 362-366). IEEE.
9. Joshi, D., and Pansare, S. (2015). Combination of Multiple Image features along with KNN classifier for classification. IEEE.
10. Reshma, P.D., and Arunvinodh, C. (2015). Image forgery detection using SVM Classifier. IEEE.
11. Teerakanok, S., and Uehara, T. (2018). Copy Move Forgery Detection Using GLCM-based Rotation Invariant Feature. IEEE. (COMPSAC).
12. Zawish, M., Siyal, A. A., Ahmed, K., Khalil, A., and Memon, S. (2019, January 14). Brain Tumour Segmentation in MRI Images using Chan-Vese Techniques in MATLAB. IEEE.
13. Hayat, K., and Qazi, T. (2017, August). Forgery detection in digital images via discrete wavelet and discrete cosine transforms. Computer and Electrical Engineering, Vol.62. Pp. 448 – 458.
14. Warif, N. B. A., Wahab, A. W. A., Idris, M. Y. I., Ramli, R. and Choo, K. R. (2016, November). Copy-move forgery detection: Survey, Challenges, and Future Directions. Journal of Network and Computer Applications. Vol. 75. Pp. 259 – 278.
15. Manu, V. T., and Mehtre, B. M. (2019, February). Tamper detection of social media images using quality artifacts and texture features. Forensic Science International. Vol. 295. pp. 100 – 112.
16. Qian, H., Mao, Y., Xiang, W., and Wang, Z. (2010). Recognition of human activities using SVM Multi-Class Classifier. Pattern Recognition Letters. Vol. 31. Iss. 2. Pp. 100 – 111.
17. Goel, A., and Vishwakarma, V.P. (2017). Fractional DCT and DWT hybridization based efficient feature extraction for gender classification. Pattern Recognition Letters. Vol.95. pp. 8 – 13.