Copy move forgery detection in digital images using Discrete Cosine Transform and Singular Value Decomposition

[1]Nirmalpreet Kaur, [2]Gurbinder Singh Brar

[1]Mtech Student, [2]Professor
[1]CSE Department
[1]Adesh Institute of Engineering and Technology, Faridkot, Punjab.

Abstract: There is a strong demand for robust authentication methods that can discern whether an image is forged or not. In this work, a hybrid approach based on discrete cosine transform and singular value decomposition (SVD) to efficiently detect and localize the copy-move forged region is proposed. In order to reduce the computational complexity of the classification procedure, we propose to arrange image blocks in a descending ordered sorted matrix based on the mean values of intensity of blocks. Further edge detection is applied to find the dominant edge pixels in the image by using horizontally and vertically derived sobel edge filter. For matching process, edge pixel blocks are taken from the sorted matrix and matching is carried out for those edge blocks only which have similar mean values. This reduces computation time of the algorithm as very few edge blocks need to be compared. Pair of blocks within each segment has been compared using singular values of DCT coefficients of the blocks, to find regions that exhibit maximum resemblance. Further morphological operation has been applied for the detected forged edge pixels which results in dilation of region in between these pixels. Re-matching is applied in similar manner to extract the actual forged regions in the image. The experimental results has been evaluated using standard copy move forgery CoMoFoD database and quality metrics i.e. sensitivity, specificity and accuracy show how effectively the method identifies the duplicated region as approx. 99% forged pixels has been truly detected as forged on whole database. Furthermore it also detects multiple copy-move forgery within the image.

***Index Terms* - Copy move forgery, DCT, SVD, CoMoFoD dataset.**

I. INTRODUCTION

Today, digital images are used extensively in various fields in our life through important areas such as news reports, forensics sciences, surveillance services, online marketing and medical diagnosis. Moreover, they can be used as proof in tribunals, and in press to adjust the meaning of pictures in order to affect the readers' points of views. Thus, this area of digital image forensics to specify the originality of digital image has become an important area of research to regain trust in digital image. The forensic analysis for digital images helps in providing information to support law enforcement, security, and intelligence agencies. Various techniques can be introduced to examine and legitimize the digital image's content. The image forgery detection is analyzed to active and passive methods. The active method relies on digital signatures or watermarking. That method depends on the information taken previously from the original image. It is clear to notice that those methods are not powerful. Because they require certain equipment like particular cameras to add watermark or a signature to a captured image; moreover, it can be manipulated. On the other hand, passive methods are optimized to examine images without resorting to previous information, where we have to make vague decision concerning how images have been manipulated. The majority of passive methods depend on supervised learning by using the extraction of certain characteristics to distinguish the original picture from the fake one [7].

The copy-move image forgery involves duplication of some parts of the image within the same image. Such tampering is generally performed with the intention of concealing some useful information or to replicate the things in order to mislead the people. The latest image editing tools allow user to perform copy-move forgeries with such sophistication that it is almost impossible to say anything about the authenticity of an image just by looking at the image. Therefore, it is required to develop a copy-move forgery detection technique that can identify and locate the forgery in digital images [1]. There are many kinds of image forgeries and among them, the "copy move/copy paste/cloning forgery" is a challenging one to detect because this forgery is done by copying a portion of the image and pasting it in another portion of the same image. Moreover, the pasted portion also has similar patterns of the copied portion. This kind of forgery can be done either for the purpose of concealing some information or to raise the number of objects. The same has been pictorially presented in Fig. 1, where, a truck in the first row of column (a) has been concealed by copy pasting the region contained the leaves and is shown in the first row of column (b). Similarly, the second row of column (a) contains only one ball, which has been copied and pasted in the same image in another portion to increase the number of objects and it is shown in the second row of (b). Such kinds of tampering are identified in this work by finding the matches among the extracted novel local symmetry features.

As the symmetry features are potentially stable and robust when it is explored at multiple scales and multiple locations, local symmetries also can be relatively descriptive. The numerous work has been previously implemented are based on the SIFT (Scale Invariant Feature Transform) and SURF (Speeded Up Robust Features) features. The main disadvantage of SIFT features is a high computational cost, especially in the process of feature extraction and matching. Even though the SURF features are an efficient one, but it lacks in finding the symmetrical pairs of feature points [3].

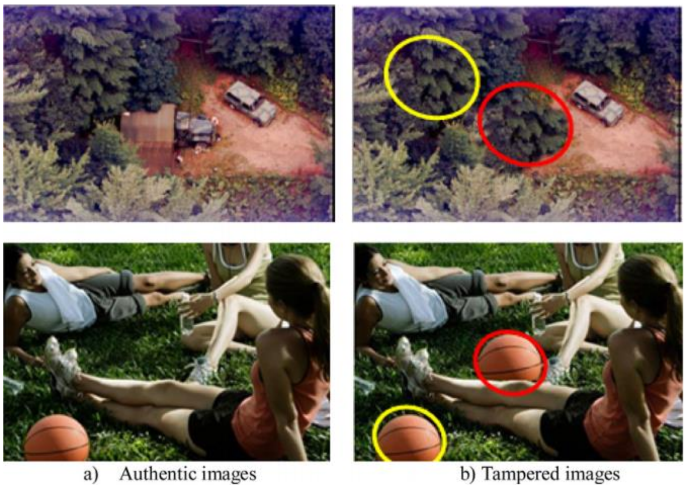


Figure 1: Sample original and copy move forged pictures [3]

II. LITERATURE REVIEW

In 2020, Meena and Tyagi [1] described a new technique to detect and locate the copy-move forgery in digital image using Tetrolet transform. Tetrolet features represent an image with all possible geometric directions. Experimental results show that the proposed technique can detect copy-move forgery very effectively under common post-processing and geometric transformations. Also, the proposed technique is very accurate in locating the shape or size of copy-moved regions. The experiments also show that the proposed technique is faster than state-of-the-art copy-move forgery detection techniques. In future, this technique can be extended to detect the copy-move forgery in noisy and distorted forged images.

Zhong & Pun (2019) [2] proposed a novel two-pass hashing detection method for the accurate localization and identification of forgery regions with high efficiency. First, it proposed the polar cosine transform (PCT) as an example to illustrate how to extract 15-dimensional block features from an HSV image with different orders and repetitions. The multiple-dimensional features of each pixel are projected into the corresponding hashing bin to obtain the corresponding hashing features. Then, a novel two-pass hashing feature representation method is proposed to concatenate multiple hashing features as the 27-bits two-pass hashing sequence. Two-pass hashing tables containing the forward and backward hashing tables are constructed to store the localization of each pixel with the same 27-bits two-pass hashing sequence in the same bucket. . Compared with state-of-the-art methods, the proposed (T) method is one of the highest efficiency methods.

Vaishnavi & Subashini (2019) [3] proposed a novel image copy move forgery detection scheme for digital images and implemented successfully. In this scheme, a new symmetry-based image features are extracted to detect the forgery and the proposed scheme is modified to detect multiple copy move forgeries also. The MICCF220, MICC-F600 and CMH datasets were used. 83.64%. 5.45% of TPR and FPR respectively were achieved the detection results in MICC-F220 dataset and 5.80% FPR and 75% TPR in MICC-F600 dataset. Also, it achieved 90.2% and 89.43% of forged region localization accuracies with compressed and uncompressed JPEG images of CMH dataset. Though the proposed scheme obtained a good forgery detection results, it needs to be enhanced.

Raju & Nair (2018) [4] proposed work focused on detection of copy-move forgery by integrating the conventional block-based and keypoint based techniques. Image keypoints can be used to detect the matching regions of an image. The proposed method used the Binary Discriminant Features to extract the keypoints and identify the matching points. The descriptor being binary and of low dimension, reduces the complexity of feature matching. SLIC super pixels replace the matching points so as to give an indication of the suspected forged regions and then a Forgery Region Localization algorithm extracts the forged regions, using color histogram matching. A final morphological close operation obtains the detected forged region. Experimental analysis proved the effectiveness of the proposed method in copy-move forgery detection. The method offers higher detection and accuracy rates. It has shown remarkable improvement in precision and F1 Score values when compared to the other state-of-the-art algorithms.

Hegazi et al. (2019) [5] presented an improved SIFT features based method for copy-move forgery detection. The main contributions of this work are introducing a density-based clustering algorithm and Guaranteed Outlier Removal algorithm that can effectively reduce false matches. Various datasets have been tested containing different typologies and resolutions of fake and original images. Experimental results exhibit that the proposed technique performs well in the existence of various attacks such as scaling, rotation, a composition of these attacks, JPEG compression and Gaussian noise compared to other similar state-of-the-art techniques. Moreover, it can handle multiple copy-move forgeries with the least false matches.

Warif et al. (2016) [6] provided a comprehensive overview of existing CMFD techniques for the entire process. Specifically, this research discussed the importance of the CMFD techniques, and outlined the common process involved in the CMFD workflow. The key processes are categorized into two categories; namely block-based and keypoint-based. We described the major classes of techniques in both categories, and listed the associated activities related to the CMFD including datasets and validations. Furthermore, it classified the copied regions to determine their relevancy in existing CMFD techniques. It also discussed how advances in big data solutions could be influence and/or solve CMFD challenges.

Alberry et al. (2018) [7] optimized the key point based techniques for detection of Copy Move forgery. While raising the number of key points, the computational requirements will raise in these techniques, so minimal execution time will be needed. In this research, the researcher optimized FCM technique for clustering the SIFT key points to decrease time complexity. The experimental results indicate that the propose algorithm decreases the detection time of appreciably same accuracy standards and minor enhancement in some cases. This research detects also in the status of rotation, scaling and multiple Copy Move attacks. In this research, a new data set is created for CMFD that includes more manipulated pictures that were performed deliberately by professionals. The obtained data set is an open source and free to be optimized as benchmarking for more comparisons. According to this research, it is highly recommended that optimizing multiple clustering algorithms or even using the FCM by matrix optimization rather than the sequential optimization done.

III. PROPOSED METHODOLOGY

3.1 RGB image convert to grayscale

Following RGB image conversion to grayscale image, the image is partitioned into overlapping block to detect forged area by using block matching to find the duplicated or identical block.

Step 1 Assuming a n × n grayscale image I, is partitioned into overlapping blocks of m × n pixels, m = 4, 8. The neighboring blocks will only have one different column or row. Each block is indicated as B_{ij}, where i and j signifies the beginning point of the block’s row and column, respectively, [Eq. (2)].

$$B_{ij}(x,y) = f(x+j, y+1) \qquad (2)$$
$$where\, x, y \in \{0,....,B-1\},\ i \in \{1,.....,M-B+1\}\ and\ \ j \in \{1,.....N-B+1\}$$

Hence, obtain N_{blocks} of overlapped sub-blocks from suspicious image using Eq. (3).

$$\mathrm{N_{blocks}} = (M-B+1)\times(N-B+1) \qquad (3)$$

Step 2 Let $\mathrm{N_{blocks}} = (M-B+1)\times(N-B+1)$, DCT is applied for each block $B_i(i=1,2,3,......,\mathrm{N_{blocks}}$. Then, exploit a DCT coefficients framework with an indistinguishable size from the block, which the comparing block could be represented [9]. Applied common quantization mask with same size as the DCT coefficients matrix and rounding to integers result as feature vector for each block.

3.2 Rearrange the coefficient

The feature vector is rearranged into row vector using zigzag scanning. Zig-zag scanning converts 2D matrix into a 1D array (row vectors). Figure 3 represents the direction of the way of zig-zag scanning arrangement.

3.3 Lexicographically sorting

The A is then sorted using lexicographically sorting and left corner’s facilitates every block that is indicated by a circle block is recorded. The sorted set could be characterized as A _ since each element of A is a vector. In lexicographic sorting, a matrix of feature vectors is developed and each feature vector appears as a row in the matrix. This matrix is further sorted in row-wise fashion and similar features in sequential rows are appeared. Figure 4 shows the feature vector in row before and after sorting.

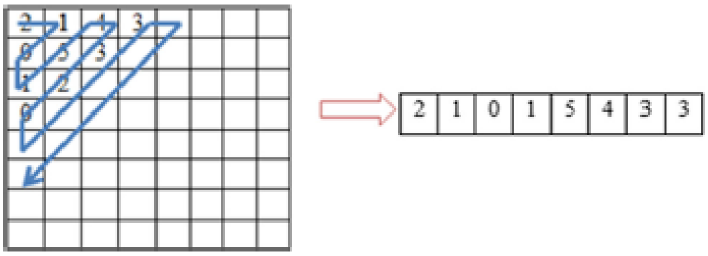


Figure 2: Zig-zag scanning converting 2D matrix into 1D array

3.4 Block matching

The block using quantized DCT coefficients is represented in robust match method. In calculating DCT coefficients, quantization process is involved decided by a user-specified parameter Q. Higher the values of Q factor means finer quantization so Q factor plays vital role in quantization steps for DCT transform coefficients. The blocks should match intently to recognize as comparable. However, more matching blocks are produced for the lower values of the Q factor and it will lead to false matches. Consequently, it may affect the accuracy of the final result. Based on A _ after the lexicographically sorting, calculated the Euclidean distance m_match (A_i, A_{i+j}) between adjacent pairs of A _ . Initialize a black map P with the size M × N and consider the looked blocks as a couple of possibility for the forgery, if the separation is littler than a preset limit D similar.

$$m\_match(A_i, A_{i+j}) = \sqrt{\sum_{k=1}^{4}(v_i^k - v_{i+j}^k)^2 < D_{similar}} \qquad (4)$$

Moreover, due to the neighboring squares might have the comparative component vector, the real distance between two comparable pieces calculated using Eq. (5). m distance

$$m\_match(A_i, A_{i+j}) = \sqrt{(x_i - x_{i+j})^2 + (y_1 - y_{i+j})^2 N_d} \qquad (5)$$

as (x,y) is the circle center of the corresponding block, m_match and m_distance is used to determine the duplicated blocks. In short, set two thresholds to make the detection: likeness threshold $D_{similar}$ and distance threshold N_d where the amounts of neighboring feature vectors are controlled, only if the test satisfies the following condition [Eq. (6)]. Where $j \in [i - N_{number}]$, for the actual block, denotes a shading map and another guide for the copied block.

$$m\_match(V_i, \mathrm{V}_j) < D_{similar}\ and\ m\_dis\,\mathrm{tance}(V_i, \mathrm{V}_j) > N_d \qquad (6)$$

3.6 Forgery decision

Since most of the natural images would have many similar blocks, the method of block matching is insufficient to make the forgery decision. In the case, that there are more than a specific number of blocks that are linked to each other within a same distance, the forgery decision could be determined. Meanwhile, the distance between the two blocks those have the similar feature vectors, $A_i\ and\ A_j$ .Let(i_1, i_2) and (j_1, j_2) represents matching blocks location. In next step, we calculate shift vector between two blocks to be compared. Refer to Eq. (7).

$$s=(s_1,s_2)=(i_{1-j_1},i_2-j_2) \qquad (7)$$

Due to the shift vectors -s and s correspond to the same shift, if necessary, normalize the shift vectors s by multiplying by -1 so that s1 ≥ 0. Increase the standardized move vector counter C by one for each coordinating pair of blocks using Eq. (8).

$$C(s_1,s_2)=C(s_1,s_2)+1 \qquad (8)$$

At the beginning, initialize the values of C to zero. At least one of the values of Cðs1; s2Þ should be more than a threshold value, if there are many blocks which give the similar feature values within the same separation. In the event that these blocks are associated with each other, then the forgery decision can be made.

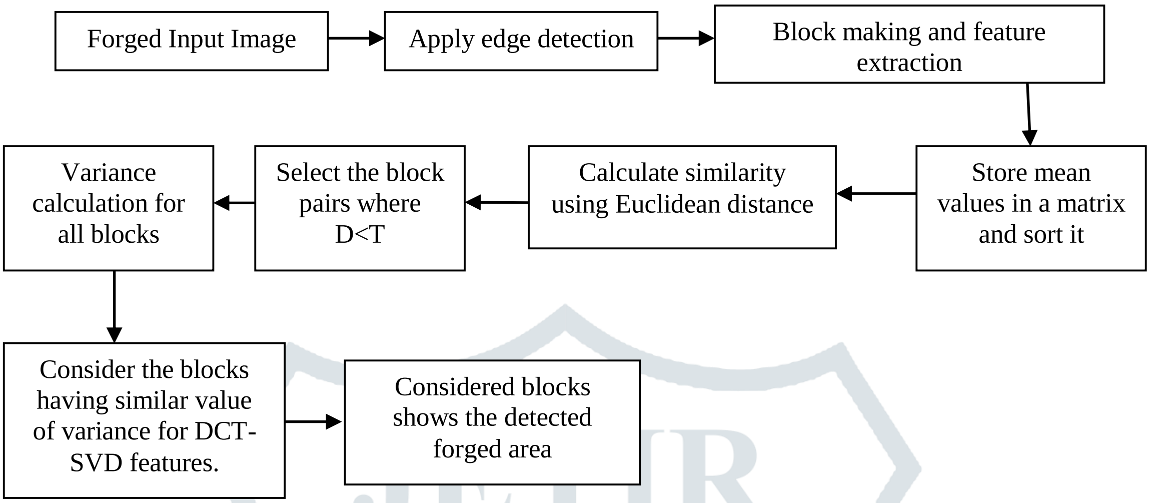

Forged Input Image
Apply edge detection
Block making and feature extraction
Store mean values in a matrix and sort it
Calculate similarity using Euclidean distance
Select the block pairs where D<T
Variance calculation for all blocks
Consider the blocks having similar value of variance for DCT-SVD features.
Considered blocks shows the detected forged area

Figure 3: Flow chart of the proposed method

IV. Results and Discussions

For the current research work, the CoMoFoD dataset is selected and then MATLAB is used to perform the experiments and calculate the performance measures of proposed algorithm.

CoMoFoD dataset

CoMoFoD is a standard dataset for benchmarking the detection of image tampering artifacts [10]. This dataset comprises of 200 images: 100 original images and 100 tampered images. The standard image size has been set as 512 × 512. In this research, 10 images will be chosen as experimental images. Each of the images will implement the block-based copy move image forgery detection approaches using DCT coefficients with 4 × 4 and 8 × 8 pixel block sizes. Figure 6 shows the sample of datasets used as input images.

Figure 4: Input images used in this experiment

The presented edge-DCT based methodology has been executed in MATLAB. Our test data consists of a set of 512 × 512 RGB images, taken from the CoMoFoD Database [10]. For the sake of experimentation, we have selected test images with copy–move forgery induced into them. For performance evaluation of the proposed method, sensitivity, specificity and accuracy has been calculated for each image. First of all, Forgery detection has been extracted from whole dataset and feature extraction has been carried out using DCT texture algorithm and sobel edge detection. After that forged pixels has been calculated. The classification accuracy is the extent to which the classifier is able to correctly classify the exemplars and is summarized in the form of confusion matrix to the test data. This is defined as the ratio of the number of correctly classified patterns (TP and TN) to the total number of patterns (species) classified. In simple words,

$$Accuracy = TP+TN/(TP+TN+FP+FN)$$

Sensitivity- The sensitivity of a classifier is the fraction of the image samples correctly classified as that specific species class. It is defined by equation below:

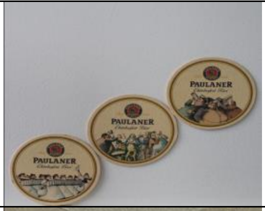
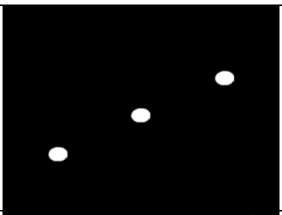
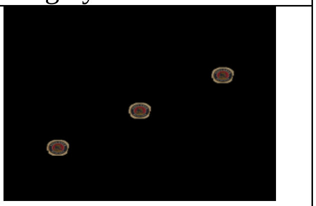
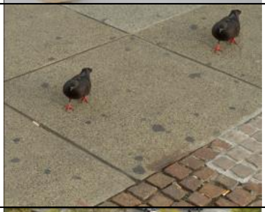
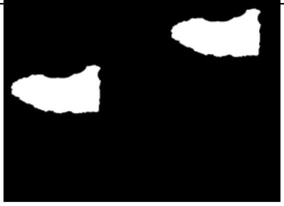
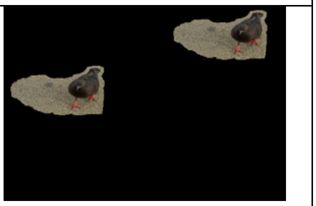
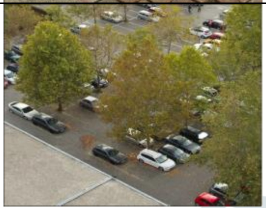
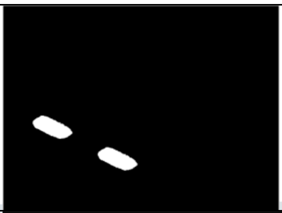
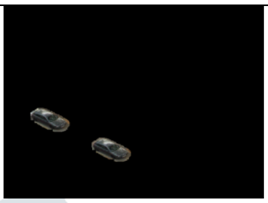
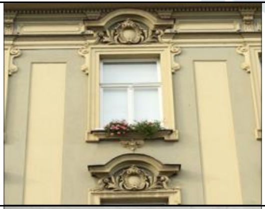
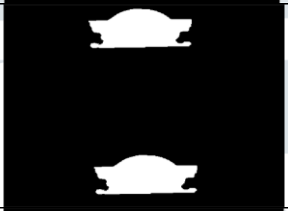
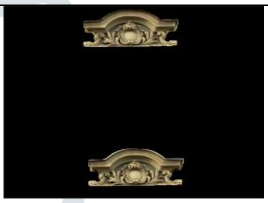
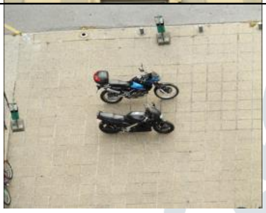
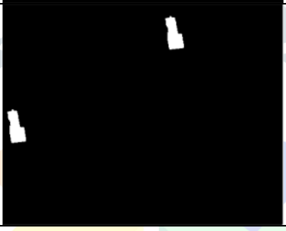
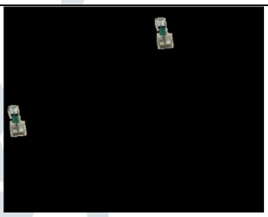
$$Sensitivity = TP/(TP+FN)$$

Specificity- The specificity is the fraction of normal pixels correctly classified as normal class. It is also called selectivity.

$$Specificity = TN/(TN+FP)$$

The results for the actual pixel location using ground truth images and that of resulted outputs has been described with above parameters.

Table 1: Forgery detection results for the first five copy-move forged images taken from CoMoFoD Database

Image used	Ground Truth	Forgery area detected
		
		
		
		
		

Below are the bar graphs for sensitivity, specificity and accuracy parameters

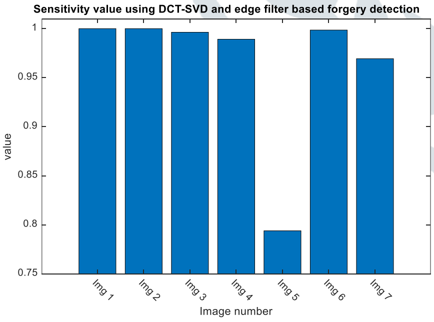


Figure 5: Sensitivity value for the copy move forged pixels detected by proposed method

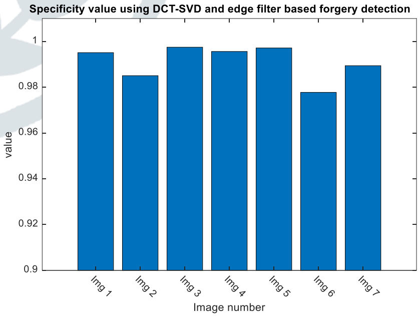


Figure 6: Specificity value for the copy move forged pixels detected by proposed method

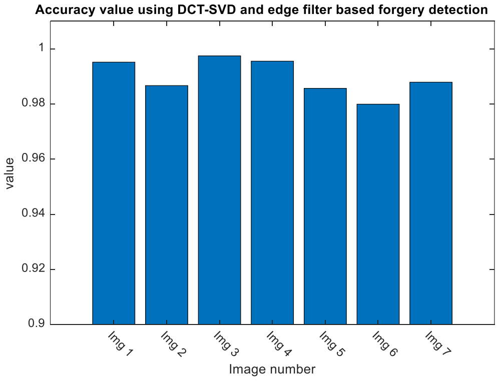


Figure 7: Accuracy value for the copy move forged pixels detected by proposed method

It has been found that proposed method of forgery detection is effective when a patch or block of an image is copy moved to another places. As Euclidian distance gives minimum error or difference between two feature sets, an exact replica of a patch of forging can be easily detected by the proposed method. Small artefacts or noisy forged detected blocks have been eliminated using morphological operations, higher accuracy has been achieved which conforms accuracy of about 99-100 per cent.

V. CONCLUSION

Copy-move forgery is one common manipulation among various digital image forgeries, where one or several regions of an image are pasted elsewhere in the same image in order to hide or duplicate objects of interest. Many examples of forgery image in history. Lately, the dark rooms were unwieldy used to perform image forgery. but todays, the forgery in digital image no need to dark room because there are many tools available to make forgery in image processing software. The forgery seen very day in press and social media. The main disadvantages of the previously developed algorithms is that they are very complex and resulted in less accuracy rate while detecting the forgery. So, in this work, we proposed a new approach for detecting the forgery from the images using DCT-SVD features. The main focus was to reduce the computational complexity of the classification procedure. For this, we introduced a new method to arrange the blocks of image in the descending order matrix that is based on the average values of the intensity blocks. The results using the proposed method have been evaluated using the CoMoFoD dataset. The result is evaluated for the performance metrics sensitivity, specificity and accuracy. The results show that the proposed method identifies the forgery area effectively and approx. 99% accuracy was obtained using the proposed method. This shows the efficiency of the proposed method in detecting the forged area from the images.

VI. REFERENCES

[1] Meena, K. B., & Tyagi, V. (2020). A copy-move image forgery detection technique based on tetrolet transform. Journal of Information Security and Applications, 52, 102481. doi:10.1016/j.jisa.2020.102481

[2] Zhong, J.-L., & Pun, C.-M. (2019). Two-Pass Hashing Feature Representation and Searching Method for Copy-Move Forgery Detection. Information Sciences. doi:10.1016/j.ins.2019.09.085

[3] Vaishnavi, D., & Subashini, T. S. (2019). Application of local invariant symmetry features to detect and localize image copy move forgeries. Journal of Information Security and Applications, 44, 23–31. doi:10.1016/j.jisa.2018.11.001

[4] Raju, P. M., & Nair, M. S. (2018). Copy-move forgery detection using binary discriminant features. Journal of King Saud University - Computer and Information Sciences. doi:10.1016/j.jksuci.2018.11.004

[5] Hegazi, A., Taha, A., & Selim, M. (2019). An Improved Copy-Move Forgery Detection Based on Density-Based Clustering and Guaranteed Outlier Removal. Journal of King Saud University- Computer and Information Sciences. doi:10.1016/j.jksuci.2019.07.007

[6] Warif, N. B. A., Wahab, A. W. A., Idris, M. Y. I., Ramli, R., Salleh, R., Shamshirband, S., & Choo, K.-K. R. (2016). Copy-move forgery detection: Survey, challenges and future directions. Journal of Network and Computer Applications, 75, 259–278. doi:10.1016/j.jnca.2016.09.008

[7] Alberry, H., A. Hegazy, A., & I. Salama, G. (2018). A fast SIFT based method for copy move forgery detection. Future Computing and Informatics Journal. doi:10.1016/j.fcij.2018.03.001

[8] Yang, F., Li, J., Lu, W., & Weng, J. (2017). Copy-move forgery detection based on hybrid features. Engineering Applications of Artificial Intelligence, 59, 73–83. doi:10.1016/j.engappai.2016.12.022

[9] Cao Y, Gao T, Fan L, Yang Q (2012) A robust detection algorithm for copy-move forgery in digital images. Forensic Sci Int 214(1-3):33–34

[10] Zampoglou M, Papadopoulos S, Kompatsiaris Y (2015). Detecting image splicing in the wild (WEB). In: Multimedia and expo workshops (ICMEW), 2015 IEEE international conference on 2015, pp 1–6