ISSN: 2277 – 9043
International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE)
Volume 6, Issue 3, March 2017

An Improved SIFT-PCA-Based Copy-Move Image Forgery Detection Method

Mona F. Mohamed Mursi, May M. Salama, Mohamed H. Habeb

***Abstract*—Copy–move image tampering is one of the most frequently tampering types that contaminate images authenticity due to its ease. Tampering detection becomes more difficult when the tampered parts are subjected to post operations like scaling, rotation, compression or noise. In this paper, a blind copy-move tampering detection and localization method is proposed. Its novelty lies in the combination of SIFT, PCA and DBSCAN techniques. The proposed method shows its potential to disclose and localize tampered regions of different sizes and shapes. Furthermore, our method requires no prior information about the image or the manipulation operations carried on it. A comparative analysis between the proposed method and other tampering detection methods is evaluated based on various performance measures. Our experimental results show that the proposed method outperforms the previous reported techniques and is quite reliable in copy-move tampering detection and localization.**

***Index Terms*—Copy–move, DBSCAN, PCA, SIFT, Tampering.**

I. Introduction

Highlight Due to its importance, copy-move forgery detection has been a focus in forensics research. A copy-move forgery followed by rotation and scaling of the forged part is a real challenging detection task[1], [2]. Most of the detection methods are categorized into two classes: block based and key point based methods. Both extract features, describe the local patches, and evaluate the similarity among various regions. The major difference between the two classes is that the feature vector is extracted for each block in block based methods, while in key point based schemes, a descriptor is obtained for each salient point which is mostly located in high entropy regions[1], [2].

Copy-Move detection has been studied intensively in the last few years. Techniques based on Scale Invariant Feature Transform (SIFT) are widely used to detect copy-move tampering. Various approaches under the SIFT based framework are more acceptable ways to detect copy-move tampering due to their robust performance like [1], [3] , [4], [5], [6], [7]. Other approaches have been used for instance *Fan, et al.*,[8] applied a robust clustering strategy. Another approach developed by *Yohannan & Manuel*[9], use Gabor filters. While *Dixit & Naskar*[10] proposed a technique for identification of copy-move tampering in digital images. While *Isaac & Wilscy*[11] detects duplicated regions, which had undergone simple translation, rotation and scaling.

This paper presents a blind key point based copy-move tampering detection technique. It combines SIFT, PCA and DBSCAN as shown in Fig.1*Fig. 1.* SIFT is used as a feature extractor, while PCA is used for feature dimension reduction. Then key points are matched using Euclidean distance. Finally, the outlier matches in the previous step is eliminated by using DBSCAN (Density-based spatial clustering of applications with noise). Further, a performance comparison of the proposed method, with and without applying DWT (Discrete Wavelet Transform), with several recent state of the art copy–move tampering detection algorithms is illustrated. The proposed method has also been tested against affine transformations.

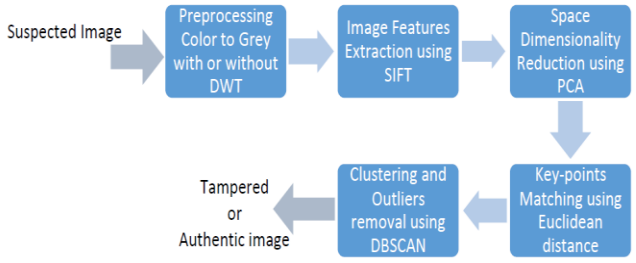

Fig. 1 the proposed method

The rest of the paper is structured as follows: The proposed forgery detection methodology is presented in Section 2. Experimental evaluation and analysis are done in section 3. Section 4, concludes the paper.

II. Proposed Methodology

The proposed method is based on SIFT combined with PCA and DBSCAN. It is applied on JPEG images from MICC-F220 dataset and implemented on MATLAB R2016a platform. MICC-F220 dataset contains 220 images: 110 tampered images and 110 authentic images. The proposed methodology stages are: Preprocessing, Image Features Extraction, Space dimensionality reduction, Key-points Matching, clustering and removal of outliers. The following explains each stage in details.

Manuscript received March, 2017.
Mona F. Mohamed Mursi, *Computer Engineering Dept, Benha / Shoubra Faculty of Engineering, (e-mail: monmursi@yahoo.com). Cairo, Egypt*
May M. Salama, *Computer Engineering Dept, Benha / Shoubra Faculty of Engineering, (e-mail: may.mohamed@feng.bu.edu.eg). Cairo, Egypt*
Mohamed H. Habeb,*Computer Engineering Dept, Benha / Shoubra Faculty of Engineering, (e-mail:* mohamed.rehan@feng.bu.edu.eg*). Cairo, Egypt.*

A. Preprocessing:

In this step, the RGB image to be examined is converted to grey scale, as it is more complex to process color image than grey scaled one. Moreover, luminance is more important in distinguishing visual features rather than RGB. There are two scenarios that could be followed; either to apply DWT on the image before the extraction of the feature descriptors or to skip applying DWT. DWT scales the image size to half, however, it suppresses the high details in the image. The input image to DWT will be decomposed into four sub images: LL, LH, HL and HH. LL corresponds to the coarse level coefficients or the approximation image from which the features will be extracted. Therefore, the number of features extracted will decrease reflecting on remarkable time reduction for feature extraction step. Fig. 2 illustrates the effect of DWT on the matching process time of 102 images using different techniques. It is well noted that matching time is strongly reduced when applying DWT.

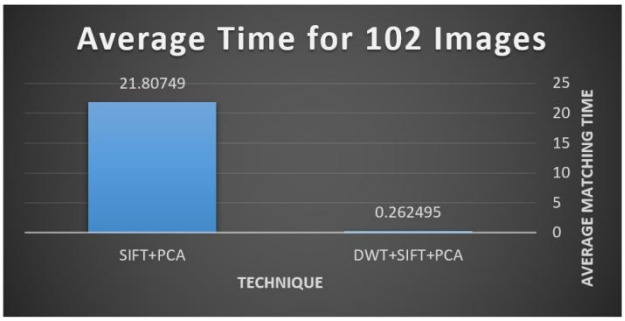

Fig. 2 Effect of DWT on Matching Process Time

B. Image Features Extraction:

An image could be expressed by a set of key-points that represent its characteristics. Each key-point is represented by a feature vector. This vector is a group of statistics values that are derived from the key-point and its surrounding neighbors. A good key-point's features reflect a good image representation specially if it is invariant to local geometrical transformations, illumination variations, additive noise, and other degradations. SIFT is an efficient and powerful key-points' features extractor algorithm that is invariant to orientation, scale, affine transformations and illumination deformations. It provides 128-dimensional features vector for each key-point. An image location is considered a key-point when it is local extrema with respect to its neighbors in the scale space. The key-point's neighbors local gradients histograms are calculated and used to form the 128-D feature vector. Furthermore, the outstanding scale of the key-point and the orientations of the neighbors' gradients assigned to the key-point provide robustness to scaling and rotation. Moreover, the formed vector is normalized to unit length to make the feature vector independent to local illumination variations.

C. Space dimensionality reduction:

It is essential to find a suitable representation of multivariate data that simplifies computations. PCA is an approach for representing data in lower-dimensional space, which discards unimportant information. For a vector of data, PCA orthogonalizes its components, orders the resulting orthogonal components (principal components) with highest priority and hence components with a little variation could be eliminated. In our work, the PCA is applied to the key-point features vector obtained from SIFT to produce another 128-D feature vector sorted in descending order according to each feature's importance. The features vector's length is reduced by cutting it off at a threshold, resulting in a vector with less than 128 elements and containing only the most important features to be matched in the following processing step.

D. Key-points matching:

A matching process finds out which key-points are similar, which could indicate the detection of duplicated areas. As the copied and pasted regions are supposed to have the same statistics and characteristics and hence the extracted key-points from the two regions are similar. Applying the approach in [12] where, for m extracted key-points, there are:

K = {k1, k2,... km}, a set of key-points extracted from the image,

V = {v1, v2, ..., vm}, the corresponding reduced feature vectors set.

E = {e1, e2, ..., em-1}, calculated similarity vector

The superior candidate match for each key-point is discovered by determining its closest neighbor in the remaining (m - 1) key-points, which is the key-point with the minimal Euclidean distance. Therefore, for any key-point ( ) in K-set, we use its features vector ( ) to calculate the similarity vector E= {e1, e2... em-1}, which is sorted Euclidean distance with respect to the other (m-1) feature vectors in the V set. Therefore, a key-point ( ) is said to be matched if and only if the distance proportion between the nearest neighbor to that of the second- nearest one is less than a specific threshold T (here T = 0.6) as shown in (1).

$$e1 / e2 < T \qquad T \in (0,1) \qquad (1)$$

However, the matched key point could be a false one since it can be a part of the same region. To overcome this issue, a 9 × 9 window centered at the key-point is created beyond which the investigation of the neighbor key-points is done.

E. Clustering and Removal of Outliers:

DBSCAN[13],[14] is one of the most common clustering algorithms and most cited in scientific literature [14]. It can identify clusters of any shape in large spatial sets of points by examining the local density distribution of those points. Hence, it can categorize those points into separate clusters that define different classes and determine also what information should be classified as noise or outliers, using only two input parameters[15], [16]. Therefore, minimal knowledge of the domain is required. Based on the above, we applied DBSCAN to cluster the detected matched key-points and to eliminate any noise or outliers as shown in Fig. 3.

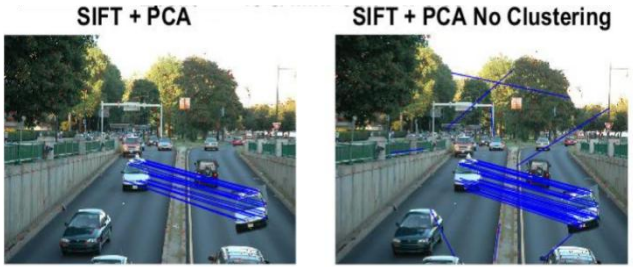

Fig. 3 Image after applying PCA with and without applying clustering to remove outliers

III. Experimental Results

Images from MICC-F220 dataset are used to evaluate the performance of the proposed method and to compare it with three tampering detection methods, namely: DWT+SIFT, DWT+SIFT+PCA and SIFT+PCA where DBSACN is applied in each. The comparative analysis was carried out based on Sensitivity, Specificity, Accuracy, False positive rate (FPR) and False negative rate (FNR) represented as follows:

$$Sensitivity\ (TPR) = \frac{TP}{TP+FN} \quad (2)$$

$$Specificity\ (TNR) = \frac{TN}{TN+FP} \quad (3)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

$$FPR = 1 - TNR \quad (5)$$

$$FNR = 1 - TPR \quad (6)$$

Where TP (True Positive) is forged images identified as forged, FN (False Negative) is forged images identified as authentic, FP (False Positive) is authentic images identified as forged and TN (True Negative) is authentic images identified as authentic. In each of the methods, four images are examined, two tampered and two authentic. In Fig. 4, Fig. 6 and Fig. 8, the first column always shows the four images, the second column shows either the tampered or the authentic image while the last column demonstrates the detection results.

A. *DWT+SIFT*

The tampered images in Fig. 5 were exposed to scaling and rotation. The detection results prove that DWT+SIFT is invariant to affine transformations.

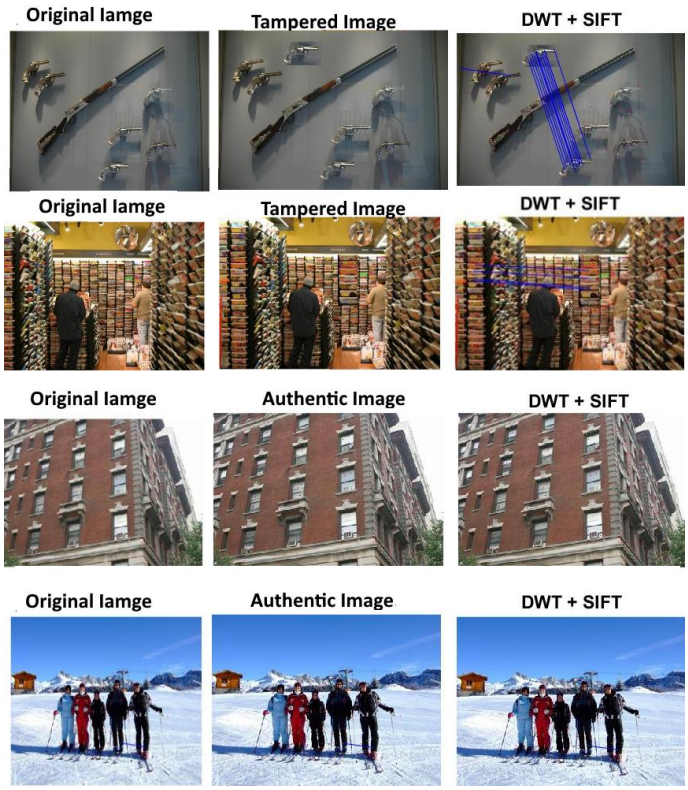

Fig. 4: Results of applying DWT+SIFT.

Although DWT+SIFT method reduces the time needed to detect tampered image, yet it was observed that it does not detect duplicated regions perfectly since DWT reduces the image size to half, resulting in the loss of some details. That's why this method is not 100% reliable. Table 1 gives the results of TP, TN, FP and FN for this method.

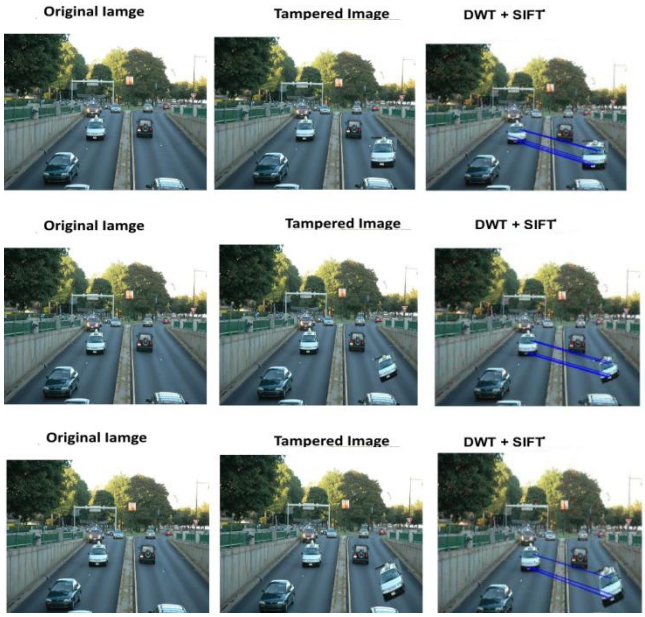

Fig. 5 Results of applying DWT+SIFT on images with transformations

Table 1: TP, TN, FP and FN for DWT+SIFT

Technique	No. of Tamp Images	No of Auth. Images	TP	TN	FP	FN
DWT + SIFT	110	110	67	104	6	43

Table 2: TP, TN, FP and FN for DWT+SIFT +PCA

Technique	No. of Tamp Images	No of Auth. Images	TP	TN	FP	FN
DWT + SIFT + PCA	**110**	**110**	**70**	**102**	**8**	40

B. DWT+SIFT+ PCA

Detection results are presented in Fig. 6, where tampering is detected.

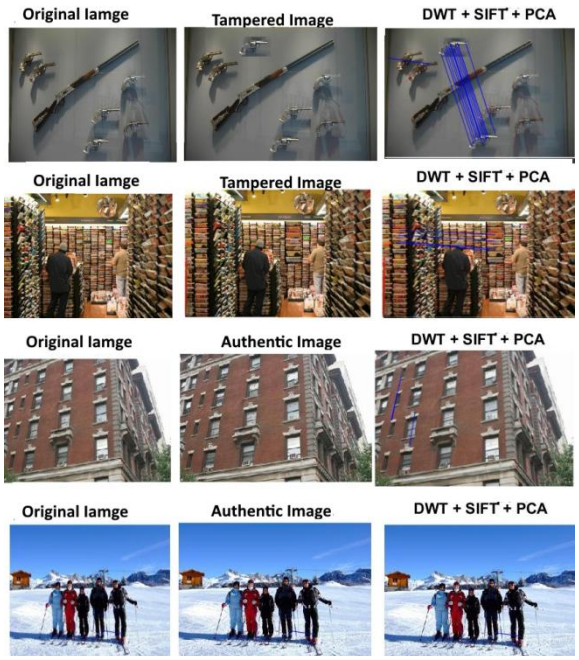

Fig.6: Results of DWT+SIFT+PCA.

The results shown in Fig.7 prove that DWT+SIFT+PCA is also invariant to affine transformations such as rotation, scaling, both rotation and scaling. Table 2 gives TP, TN, FP and FN for this method which show a slight improvement relative to previous method.

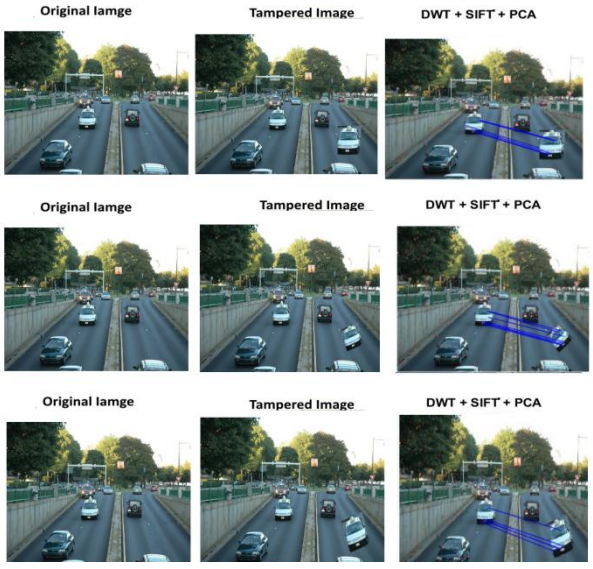

Fig. 7 Results of applying DWT+SIFT+PCA on transformed images.

C. SIFT+ PCA (without DWT)

Fig. 8 and Fig. 9 show that the SIFT+PCA method can accurately detect and locate tampered regions and has resistance ability for post processing such as rotation, scaling, both rotation and scaling effectively.

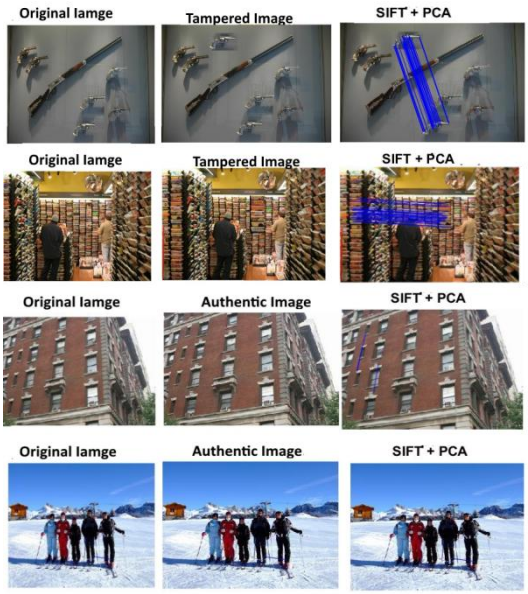

Fig.8 Results of SIFT+PCA

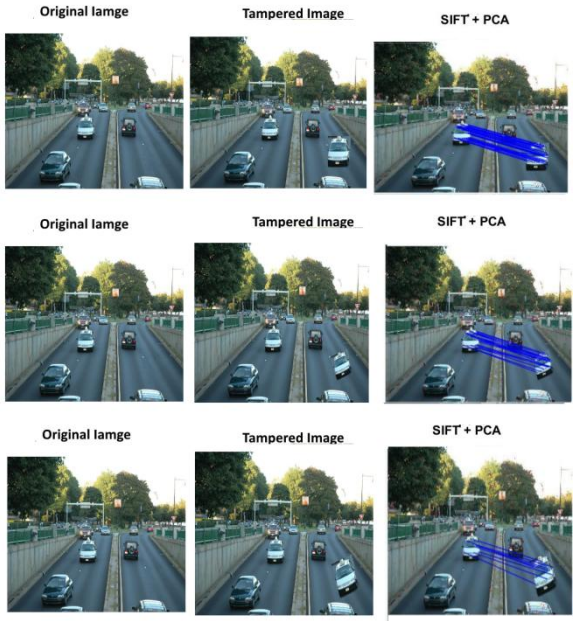

Fig.9: Results of SIFT+PCA on transformed images

Table 3 gives the results of TP, TN, FP and FN for forgery detection using SIFT+PCA.

Table 3: TP, TN, FP and FN for SIFT +PCA

Technique	No. of Tamp Images	No of Auth. Images	TP	TN	FP	FN
SIFT + PCA	110	110	106	107	3	4

D. Quantitative Comparison of DWT+SIFT, DWT+SIFT+PCA and SIFT+PCA

The performance of an image-forgery detection system can be measured in terms of Sensitivity, Specificity, Accuracy, FPR and FNR. Table 4 presents the values of these metrics for the three methods. These performance metrics were calculated using the definitions given in (2)-(6).

Table 4: TPR, TNR, Accuracy, FPR and FNR for SIFT combined with different techniques

Technique	Sensitivity (TPR)%	Specificity (TNR) %	Accuracy %	FPR	FNR
DWT + SIFT	70	95	78	0.05	0.3
DWT + SIFT + PCA	64	93	80	0.07	0.36
SIFT + PCA	96	97	97	0.03	0.04

Fig. 10 demonstrates graphically that SIFT+PCA method achieves higher sensitivity, accuracy and has smaller FNR as compared with DWT+SIFT and DWT+SIFT+PCA methods.

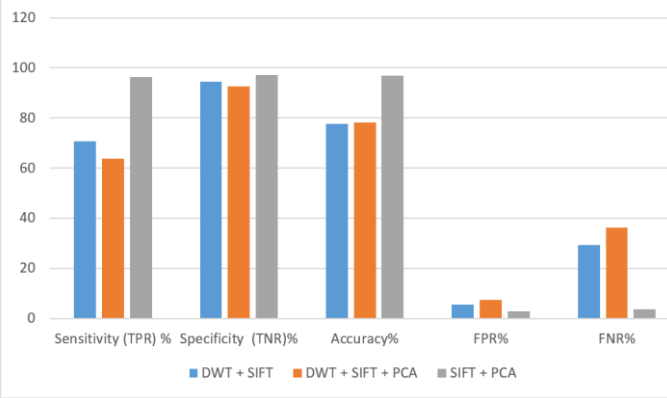

Fig.10: Comparative analysis of obtained results from different techniques

While the efficiency of the methods was satisfactory enough, the computational time was still an issue to be improved. The time taken to detect forgery is not directly dependent upon the image size; rather it depends on the number of key-points generated and their corresponding descriptors. We found that, the low dimensionality of the DWT+SIFT and DWT+SIFT+PCA, compared to SIFT+ PCA mechanisms, decrease the computational time but also reduce the efficiency of the copy-move forgery detection.

A comparative analysis between the proposed method and other copy-move tampering detection methods is shown in Table 5. For a reliable comparison, we compared our method with other methods that use either the same combination of techniques (i.e. SIFT+PCA) or the same dataset (i.e. MICC-F220) based on accuracy, TPR and FPR measures. As shown in Table 5, the detection accuracy of our method is 97% which is the highest accuracy among the other methods in addition to the high TPR=96% and a low FPR=3%, The obtained results indicate that the proposed method is superior in the detection and localization of copy-move tampering even if the tampered images are exposed to different post-processing operations.

Table 5 Comparative results between several state-of-the-art methods and our method

Methodology	**The used technique(s)**	**Used dataset**	**Accuracy %**	**TPR %**	**FPR %**
Our Proposed method	SIFT+PCA+ DBSCAN	MICC-F220	97	96	3
***Kaur, et al.*[6]**	PCA+SIFT	MICC-F220	95	97.20	7.27
***Uliyan, et al.*[17]**	Hessian features + CSLBP	MICC-F220	-	92	8
***Kaur & Sharma*[18]**	DyWT+SURF	MICC-F220	-	76	23
***Naincy & Bathla*[19]**	DCT + SIFT	MICC-F220	86.36	79.09	6.36
***Mahajan*[20]**	SIFT+PCA	only 10 images	92	-	-
***Hashmi*[21]**	DWT + SIFT	100 images (50 authentic and 50 forged)	94	92	4
***Fan, et al.*[8]**	SIFT+ T-Linkage	MICC-F600	89.64	80.34	1.06
***Amerini,et al.*[22]**	SIFT+ J-Linkage	MICC-F600	-	81.6	7.27
***Fan, et al.*[8]**	SIFT+ T-Linkage	MICC-F600	89.64	80.34	1.06
***G. Muhammad*[23]**	DyWT	10 images + CASIA v1.0 [Only 160 images (80 authentic and 80 forged)]	-	93.08	3.52
***Yohannan & Manuel*[9]**	Gabor filter	CoMoF-oD	80	90.90	50

IV. CONCLUSION

One of the biggest issues that image tampering detection techniques face is being able to detect the duplicated image regions even if it undergoes image processing operations. In this paper, copy-move image forgery detection method using SIFT+PCA is introduced

combined with DBSCAN clustering. Image feature descriptors extracted by SIFT exist in high dimensional space, which can be effectively reduced using PCA. A comparative analysis of this method has been carried out (with/without) DWT as a pre-processing step. Sensitivity, Specificity, Accuracy, FPR and FNR were evaluated as performance metrics for the proposed method and the other reported methods. Our experimental results show that the detection accuracy of SIFT combined with PCA method is superior to DWT combined with SIFT alone or together combined with PCA (DWT+SIFT and DWT+SIFT+PCA). In addition, a comparison is carried out to validate our proposed method. It reveals that the proposed method has very good performance (97%) when using SIFT followed by PCA, then performing clustering using DBSCAN.

REFERENCES

[1] S. F. Z. B. Q. &. B. L. Wenchang, "Improving image copy-move forgery detection with particle swarm optimization techniques," *China Communications*, vol. 13, no. 1, pp. 139-149, 2016.

[2] M. Zandi, A. Mahmoudi-Aznaveh and A. Talebpour, "Iterative Copy -Move Forgery Detection Based on a New Interest Point Detector," *Transactions on Information Forensics and Security*, 2016.

[3] R. Nithiya and S. Veluchamy, "Key point descriptor based copy and move image forgery detection system," in *Science Technology Engineering and Management (ICONSTEM), Second International Conference on. IEEE*, 2016.

[4] B. Liu and C. Pun, "A SIFT and Local Features Based Integrated Method for Copy-Move Attack Detection in Digital Image," China, 2013.

[5] R. C. Pandey,, S. K. Singh, K. K. Shukla and R. Agrawal, "Fast and Robust Passive Copy-Move Forgery Detection Using SURF and SIFT Image Features," in *9th International Conference on Industrial and Information Systems (ICIIS). IEEE*, 2014.

[6] H. Kaur, J. Saxena and S. Singh, "Simulative Comparison of Copy-Move Forgery Detection Methods for Digital Images," *International Journal of Electronics, Electrical and Computational System*, vol. 4, no. Special, pp. 62-66, 2015

[7] L. Liu, R. Ni, Y. Zhao and S. Li, "Improved SIFT-based Copy-move Detection Using BFSN Clustering and CFA Features," in *10th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2014.

[8] Y. Fan, Y. S. Zhu and . Z. Liu, "An Improved SIFT-Based Copy-Move Forgery Detection Method Using T-Linkage and Multi-Scale Analysis," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 7, no. 2, pp. 399-408, 2016.

[9] R. P. Yohannan and M. Manuel, "Detection of copy-move forgery based on Gabor filter," in *2nd IEEE International Conference on Engineering and Technology (ICETECH)*, India, 2016.

[10] R. Dixit and R. Naskar, "DyWT based Copy-Move Forgery Detection with Improved Detection Accuracy," in *3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, India, 2016.

[11] M. M. Isaac and M. Wilscy, "A Key point based Copy-Move Forgery Detection using HOG features," in *International Conference on Circuit, Power and Computing Technologies [ICCPCT]*, 2016.

[12] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, p. 91–110, 2004.

[13] "DBSCAN," 2016. [Online]. Available: https://en.wikipedia.org/wiki/DBSCAN. [Accessed 14 11 2016].

[14] D. Z. &. N. W. Peng Liu, "VDBSCAN: Varied Density Based Spatial Clustering of Applications with Noise," *Service Systems and Service Management, 2007 International Conference on* , pp. 528-531, 2007.

[15] H. A. N. N. Bäcklund H, "DBSCAN - A Density-Based Spatial Clustering of Application with Noise.," Linkoping University–ITN, Rep. TNM033., 2011.

[16] M. Ester, H. P. Kriegel, J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Kdd*, vol. 96, no. 34, pp. 226-231, 1996.

[17] D. M. Uliyan, H. A. Jalab and A. W. A. Wahab, "Copy move image forgery detection using Hessian and center symmetric local binary pattern," in *IEEE Confernece on Open Systems (ICOS)*, 2015.

[18] R. Kaur and T. Sharma, "Image Forgery Detection using Speed up Robust Feature Transform, Wavelet Transform, Steerable Pyramid Transform and Local Binary Pattern," *International Journal of Modern Computer Science and Applications (IJMCSA)* , vol. 4, no. 5, 2016.

[19] Naincy and A. K. Bathla, "An Enhancement of Copy Move Forgery Detection in Digital Images Using Hybrid Technique," in *2nd International Conference on Science, Technology and Management, University of Delhi(DU), Conference Center, New Delhi(India).*, 2015.

[20] N. K. a. N. Mahajan, "Image Forgery Detection using SIFT and PCA Classifiers for Panchromatic Images," *Indian Journal of Science and Technology*, vol. 9, no. 35, pp. 1-6, 2016.

[21] M. H. A. a. K. A. Hashmi, "Copy Move Forgery Detection using DWT and SIFT Features," in *Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference on*, 2013.

[22] I. Amerini, L. Ballan, R. Caldelli, A. D. Bimbo, L. D. Tongo and G. Serra, "Copy-move forgery detection and localization by means of robust clustering with J-Linkage," *Signal Processing:Image Communication*, vol. 28, no. 6, p. 659–669, 2013.

[23] M. H. a. G. B. G. Muhammad, "Passive copy move image forgery detection using undecimated dyadic wavelet transform," *Digital*, vol. 9, no. 1, p. 49–57, 2012.

Mona F. Mohamed Mursi, Professor, Computer Engineering Dept., Shoubra Faculty of Engineering, Benha University, Cairo, Egypt
Email: monmursi@yahoo.com

May A. Salama, Assistant Prof, Computer Engineering Dept., Shoubra Faculty of Engineering, Benha University, Cairo, Egypt
Email: may.mohamed@feng.bu.edu.eg

Mohamed H. Habeb Teaching Assistant, Computer Engineering Dept., Shoubra Faculty of Engineering, Benha University, Cairo, Egypt
Email: mohamed.rehan@feng.bu.edu.eg