IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-3, Mar 2019 E-ISSN: 2347-2693

A SIFT with RANSAC Based Spatial Tampering Detection in Digital Video

Jayashree D. Gavade[1]*, Sangeeta R.Chougule[2]

[1]Department of Electronics, Textile and Engineering Institute, Ichalkaranji-416115-India
[2]Department of Electronics and Communication, KIT's College of Engineering, Kolhapur-416234-India

*Corresponding Author: jayashree2k2@gmail.com, Tel.: 9764233811

DOI: https://doi.org/10.26438/ijcse/v7i3.11561163 | Available online at: www.ijcseonline.org

Accepted: 20/Mar/2019, Published: 31/Mar/2019

Abstract—This paper presents passive blind forensic scheme to detect spatial tampering in MPEG-4 (Moving Picture Experts Group-4) digital video. In spatial tampering, small region of frame is copied and pasted at some other location in same frame. A proposed algorithm uses SIFT (Scale Invariant Feature Transform) and RANSAC (Random Sample Consensus) to detect the tampering. In this local features from each frame are extracted using SIFT and those features are matched to identify forged area. At the end RANSAC homography is used to remove the false matching to increase the detection accuracy. The proposed method performance is measured with respect to detection accuracy and computational time and verified on compressed and uncompressed videos. To create test data various geometric alterations used in forgery such as scaling, rotation are considered. The simulation results proves that the proposed method finds the forged area efficiently for all the above mentioned cases with average detection accuracy of 99.5%. The algorithm is tested for various compression rates to check its robustness. The detection accuracy of the algorithm increases as the compression rate increases. The performance of the proposed algorithm is compared with two other methods reported in literature which shows that the proposed scheme has higher detection accuracy compared to other methods. The average computational time observed is 0.56 seconds.

Keywords—Spatial tampering, Forgery detection, SIFT, RANSAC, Forensic scheme

I. INTRODUCTION

In recent days digital videos and images are used to convey the important information through newspapers, news channels and social media sites such as YouTube, WHATSUP and Facebook. The digital media has got key impact on day to day life as it is most effective media of quick information delivery, but this has dark side too. This shared information may not be true facts and there is a possibility of manipulation in the video using forgery techniques. The manipulation in image and video information is easily possible due to easily available editing and processing tools. These tools are such powerful and easy to use that even a novice can handle them to modify the contents of digital video without leaving any visible traces of manipulation. This process of modifying/altering the contents of video to change its meaning is called as video forgery or video tampering [1], [2], [3].

The most common method used to tamper the video is copy move tampering. In this method, two types of attacks are used. 1) The contents in the frame are changed, for example, object from frame is copied and pasted at some other location in the same frame. This is called as spatial tampering or region duplication tampering. As the forged object belongs to same frame, its statistical properties are uniform which makes it difficult to identify this type of tampering. 2) In second case, the sequence of frames is changed or altered to hide specific activity in the video, called as frame duplication attack. In this, particular sequence of frames is copied and pasted at other sequence in video. This type of tampering is called as temporal tampering [4].

The intense manipulations in the video can lead to the serious concern, as it may create vulnerable situation in the society. For example, in court trials, someone can delete the specific objects in evidence video to hide its presence in order to mislead the court of law, hence the authenticity of the video must be checked before to present it as an evidence. The authenticity of the video is checked by verification of the video contents. A video forensic is the branch which deals with the verification of integrity of the digital video. There are two methods used to detect the video tampering. 1) Active method: In this digital signature or watermarking are used to validate information for authentication during recording of the video. 2) Passive

method: In this internal properties of the video are used to detect the tampering. The second method doesn't require any specialized hardware or source video to detect the tampering, so it is called as passive blind forgery detections method and most preferred over other [4], [5].

This paper focuses on spatial tampering detection method developed for MPEG-4 video. The spatial tampering is same as that of image copy move forgery in which small region of image is copied and pasted at other location in the same image. In thisprocess, before pasting the region, various image processing operations such as scaling, rotation, compression are used to retouch the region which makes it difficult to identifyit by naked eyes. At present, most of the researchers have focussed on image forensic and most of them suggested methods for image forgery detection but for video forgery very few are available, so video forensic is highly opportunistic area in the current research.

The rest of the paper is organized as follow: The section II, takes brief review of the related work reported in literature to detect spatial tampering in digital video. The section III, focuses on the method used in this work to detect the spatial tampering. The section IV the detailed result analysis is done and the results of proposed methods are compared with existing methods. The last section V concludes the proposed work with future scope.

II. RELATED WORK

In image copy move forgery detection methods, feature point extraction algorithms such as SIFT, SURF, HOG, FFT, and FMT are used, whereas for feature matching clustering algorithms such as KD means, g2nn, k-means clustering etc. are used [6],[7],[8],[9],[10]. While to handle the image forgery detection problem, many researchers have considered geometric transformations on forged region such as rotation, compression. The performance of the algorithm is tested against these geometric transformations to decide the robustness of the algorithm [11], [12], [13], [14], [15], [16],[17].

However, few algorithms have been suggested for detecting spatial tampering in digital video. The Wang and Farid (2007) are the first those who addressed video tampering detection problem. They have proposed a method in which each frame is divided into overlapping blocks of size 16 by 16. The correlation coefficient of each block is calculated and the blocks whose correlation is above specified threshold are considered as candidate of duplicated region. However, the detection accuracy is very low for small forged region [18]. The Subramanyam and Emmanuel (2012) have used HOG features to detect forgery, in which first the frames are allocated into suitable block size and HOG descriptors of each block are generated. In next step HOG descriptors of each block are matched against other block to find the duplicate region. The performance of this algorithm is very good and robust against various attacks at the cost of high computational time [19]. In Pun et al. (2015), combination of block based and key point based methods are used, in which first the image is segmented into non-overlapping irregular blocks. The features are extracted from each block and matched to locate the forged region. The video is down sampled to reduce the computational time however, down sampling loses the features which reduces accuracy [20]. In Pandey et al. (2014), SIFT algorithm is used to extract the key points from each frame, then K-NN matching algorithm is used to find best 10 matches and finally dynamic thresholding is applied to find the forged region [21]. In [22], the authors suggest a method in which noise correlation properties between spatially collocated blocks are used to detect video tampering. In [23], the detection of forged region is done based on the inconsistencies of noise characteristics, which occurs due to the forged patches from different videos. Generally, the noise properties depend on the intrinsic properties of camera, and hence the noise characteristics are not useful when the forged patch comes from the same video. For low compression rates the noise properties may not be estimated correctly. In the upcoming sections, the proposed method, result analysis and conclusion is discussed rigorously.

III. METHODOLOGY

The flowchart in Figure 1, shows the steps used to detect the spatial forgery in video. To find the tampered area in frame of video, first the video is transformed into frames. Then the SIFT key points of individual frame are extracted and matched to find the tampered region. At the end, RANSAC homography matching is used to remove false positive. The detailed algorithm is discussed in the sub-sections.

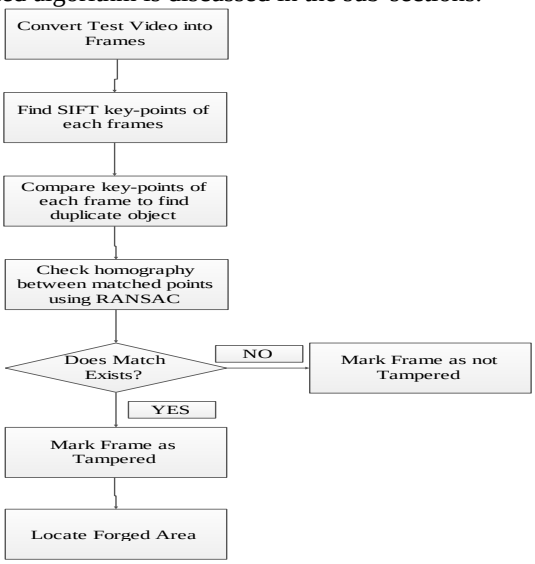

Figure1: Flowchart of Proposed Method

A. Scale Invariant Feature Transform (SIFT):

The SIFT is a computer vision algorithm used for detecting and extracting local feature descriptors. These descriptors does not change with variation in illumination, noise, rotation, scaling, and small variation in viewpoint. This is the reason SIFT is used to extract features in this work since the forger may change the illumination or add noise to the copied region before pasting it at other location. This algorithm is divided in to three steps, 1) the gradient oriented histogram is used to calculate first set of key points which normally are scale and illumination. 2) The scale-space structure of image is used to generate invariant features. 3) The features which are stable over affine transformation and having adequate contrast are selected and others are discarded.

Following are the steps to obtain SIFT features from the image:

For each octave
    Create Gaussian blur intervals
    Create difference of Gaussian intervals
    Find edges for each interval
End For

Search each octave for stable extrema
Create key points of dominant orientation of extrema

For each key point
    Rotate sample grid to key point orientation
    Sample region and create descriptor
End For
    Save Descriptors

Once the SIFT features are extracted, the next step is of feature matching. To decide the matched pair of features, the angle between features are compared. If this angle is less than the predefined threshold value, then these features are considered to be matched. The angle is calculated using dot product of coordinates of SIFT features given by Equation 1.

The dot product is given as

$$a.b = |ab^T| \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots [1]$$

Where, a and b are the coordinates of SIFT features and b^T stands for the b transpose. We have to check whether the nearest neighbour has angle less than distance ratio. This angle is calculated by equation 2

$$a.b = |a||b|cos\theta \dots\dots\dots\dots\dots\dots\dots\dots\dots [2]$$

The inverse cosine transform is applied to the dot product and match the nearest neighbour.

This procedure generates the matched pairs of key points, there are chances of getting some false matched pairs as some identical points may present in the frame due to similar objects in the frame. This leads to the false positive rate means authentic region will be detected as forged region, which directly affect the detection accuracy of the algorithm. To remove this false positive matching, RANSAC is used.

B. Random Sample Consensus (RANSAC):

To apply RANSAC, minimum four matches are required between the clusters. Homography, H is estimated by random selection of any four from matched points. All the remaining matched points are transformed according to H and compared in terms of distance with respect to their corresponding matches. The distance metric given in Equation 3 is used for RANSAC.

$$d = \sum_{i=1}^{NUM} min\big(D(p_{ib}\varphi(p_{ia}: H), t)\big) \dots\dots [3]$$

Where, p_{ia} and p_{ib} are the points in cluster a and b respectively. (p_{ia}: H), represents the projection of point p_{ia} of cluster based on transformation matrix H, t is the threshold value and *NUM* represents the number of points. The points with distance greater than t are termed as inliers while others as outliers and are discarded.

Following is the pseudo code of the proposed algorithm:

Get SIFT features of two objects namely obj1 and obj2

[des1, loc1] = sift (obj1)

Here des1 are SIFT descriptors of obj1

Similarly,

[des2, loc2] = sift (obj2)

For Matching: Find dot product between two descriptors

x = des1.des2^T

Find angular distance array between two descriptor

θ () = cos(x)

Sort angular distance which returns sorted array.

Val [] = sort (θ)

Now, for each value in sorted array compare with its next value along with threshold called as distance ratio.

if Val[1] < Val[2] X Distance Ratio

then matching is there

else not matching

In this study the distance ratio is decided by trial and error basis and we get the best results for

$$dr = 0.47.$$

C. Details of Test Video Data Set:

To test the performance of proposed method, eleven test videos are used. Following parameters are considered while selecting the videos.

1. Both stationary and moving camera recorded videos.
2. Compressed and uncompressed videos, surveillance videos with moving objects.
3. When the video is tampered intentionally to change its meaning, the forger will not simply copy and paste the object as it is, he/she may apply signal processing operations on the objects or regions before pasting it at other location to make the detection process complex. To address this issue, we have considered various geometric transformation such as rotation, scaling, illumination change, noise addition while creating forged regions.
4. Compression rate.

Three to four frames of each video are tampered to create region duplication forgery. All the eleven videos are taken from internet. Four videos are HD and remaining all are MPEG-4 compressed videos.

D. Performance Parameters:

Three parameters are used to test the performance of the proposed method. Following are the parameters.

$$PrecisionRate(PR) = \frac{TP}{TP + FP}$$

$$RecallRate(RR) = \frac{TP}{TP + FN}$$

$$DetectionAccurcy(DA) = \frac{TP + TN}{TP + TN + FP + FN}$$

Where,

TP (True Positive) = Authentic is detected as Authentic
TN (True Negative) = Forged is detected as Forged
FP (False Positive) = Authentic is detected as Forged
FN (False Negative) = Forged is detected as Forged

(TP+TN) represents the total number of detections, and (TP+TN+FP+FN) represents the total number of frames in the experiments. DA is the percentage of correct detection. High value of DA corresponds to better detection accuracy.

IV. RESULTS AND DISCUSSION

To verify the performance of the algorithm, an experimentation is carried out in two phases. In first phase, the detection accuracy is taken as performance measure and in second phase, computational time is taken as measure. The following sub sections A and B explore the results in detail.

A. In terms of Detection Accuracy:

Table 1: Performance parameters for video data

Sr. No	Test Video	Recall Rate	Precision Rate	Detection Accuracy
1	T1	100	100	**100**
2	T2	99	100	**99**
3	T3	98	100	**98**
4	T4	98	100	**98**
5	T5	99	100	**99**
6	T6	100	100	**100**
7	T7	100	100	**100**
8	T8	100	100	**100**
9	T9	100	100	**100**
10	T10	100	100	**100**
11	T11	100	100	**100**

Table 1, indicates the performance measure of proposed method in terms of Recall Rate, Precision Rate and Detection Accuracy for all test videos. The detection accuracy is slightly less(98%-99%) for the videos T2, T3, T4, T5, as these videos contains higher motion activity (sport videos), illumination change, and similar objects such as cars (street video), still the average detection accuracy of proposed method is 99.5%.

B. In terms of Simulation Time:

Table 2: Simulation Time of Video Data

Sr. No	Test Video	Number Of Frames	Computational Time (in seconds)
1.	T1	32	0.020
2.	T2	200	0.022
3.	T3	100	0.434
4.	T4	50	1.769
5.	T5	78	0.204
6.	T6	34	1.306
7.	T7	100	0.122
8.	T8	100	0.081
9.	T9	100	0.170
10.	T10	100	0.525
11.	T11	100	1.524

Table 2, shows the total simulation time (second) taken by each video. T3, T4, T6 and T11 are the HD videos. Hence, the simulation time required for these videos is more as compared to other videos. However overall, the total time

taken by each video is considerably small. The average time consumed to detect the forged region is 0.56 seconds.

C. Simulation Results:

To summarize the simulation results and to explore all the attacks graphically, a single video T12 has been considered as an example. This video consists of the 40 frames, out of which 15 frames are tampered using various attacks. In this while considering a particular attack, variations in that attack are done. For example, in rotation attack, the object is rotated by an angle like 300, 450 etc. The frame number and details of the attack are as listed in Table 3.

Table 3: Details of the attacks for Video Forgedv5

Sr. No	Frame No	Details of Attack
1	4	Rotation: Object rotated by 30^0
2	5	Rotation: Object rotated by 45^0
3	6	Rotation: Object rotated by 60^0
4	7	Rotation: Object rotated by 180^0
5	14	Scaling: Object scaled down by 20%
6	15	Scaling: Object scaled up by 20%
7	16	Scaling: Object scaled down by 40%
8	17	Scaling: Object scaled up by 40%
9	24	Gaussian Noise: Noise added by 10%
10	25	Gaussian Noise: Noise added by 20%
11	26	Gaussian Noise: Noise added by 30%
12	34	Illumination Variation: Brightness increased by 10%
13	35	Illumination Variation: contrast increased by 10%
14	36	Illumination Variation: Brightness decreased by 10%
15	37	Illumination Variation: contrast decreased by 10%

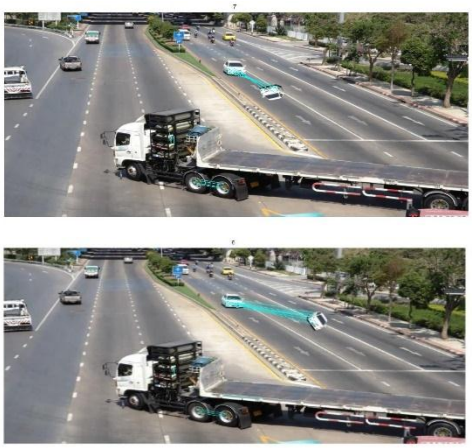

Figure 1. Rotation Attack
2a: Object rotated by 30% 2b: Object rotated by 45%
2c: Object rotated by 60% 2d: Object rotated by 180%

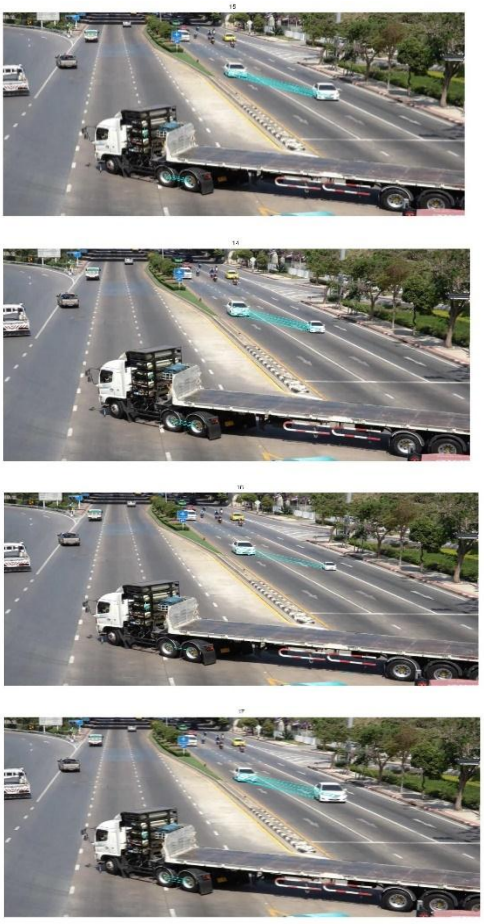

Figure 2. Scaling Attack
3a: Object scaled down by 20% 3b: Object scaled up by 20% 3c: Object scaled down by 40% 3d: Object scaled up by 40%

Figure 3. Gaussian Noise addition Attack
4a: Noise added by 10% 4b: Noise added by 20%
4c: Noise added by 30%

Figure 4. Illumination Variation Attack
5a: Brightness increased by 10% 5b: Contrast increased by 10% 5c: Brightness decreased by 10% 5d: Contrast decreased by 10%

Figure [2-5] summarises the simulation results for the various attacks for test video T12 and it is clear that the algorithm detects the smallest size of the object, 180 degree rotated object and blurred object successfully. This proves the robustness of the proposed method for various geometric transformations. The detection accuracy for the above video is 98.5%.

Finally, the performance of the proposed method is verified against various compression rates. For demonstrating the result, test video T12 is considered as an example. The video is compressed for Q ranging from 10 to 50. The performance in terms of detection accuracy is listed in Table4 and it is clear that the detection accuracy increases with respect to the compression rate.

Table 4: Performance parameter for various compression rates for test video Test12

Sr. No	Compression Rate	Detection Accuracy
1	10	80
2	15	85
3	20	88
4	25	90
5	30	95
6	40	98
7	50	98.5

D. Comparison of proposed method with existing methods:

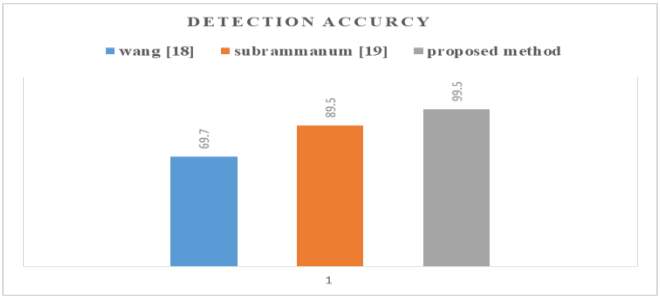

Figure 5. Comparison of proposed method with existing method

The Figure 6 shows, comparison of the results for proposed method in terms of detection accuracy with respect to existing methods reported in [18] and [20]. From the graph it is clear that the detection accuracy of the proposed method is high (99.5%) as compared to other methods.

V. CONCLUSION AND FUTURE SCOPE

The security and authenticity of the information are the key issues in the digital world. The information tampering may leads to the vulnerable situation in the society and through this work we have contributed by suggesting new tempering detection technique. This paper explores one of the most common and difficult tampering attack in MPEG-4 video i.e. spatial tampering. The proposed method uses SIFT to extract key points from each frame of test video and match to take the decision on tampering. In this work the RANSANC algorithm is used to remove the false matches and localize the accurate tampered area in each frame. The proposed method performance is tested with 12 tampered videos. The test dataset has been developed by tampering the videos using different possible attacks. Also proposed method is tested for compressed video with varying compression rate. The performance of the method is evaluated in terms of detection accuracy and simulation time. The average accuracy of the proposed method is 99.5% while the average time taken to detect the tampering is 0.5 second. The results shows that proposed method is capable to detect all the attacks, which shows the robustness of the method. The proposed method is accurately and efficiently detects and locates the tampered regions across the frames in given test video. The extension of the work is to apply proposed method to detect the Spatio-temporal attack in video.

REFERENCES

[1] K. Sitara, B. M. Mehtre, “Digital video tampering detection: An overview of passive techniques”, Digital Investigation,vol 18,pp 8-22,2016.

[2] S. Milani, M. Fontani, P. Bestagini, M. Barni, A. Piva, M. Tagliasacchi, and S. Tubaro, “An overview on video forensics,” APSIPA Transactions on Signal and Information Processing, vol.1, pp.1-18,2012.

[3] S. Upadhyay and S. K. Singh, “Video authentication: Issues and challenges,” International Journal of Computer Science, vol. 9, no. 1-3, pp. 409–418,2012.

[4] H. Yin, W. Hui, H. Li, C. Lin, and W. Zhu, "A Novel Large-Scale Digital Forensics Service Platform for Internet Videos," IEEE Transactions on Multimedia, vol. 14, pp. 178-186, 2012.

[5] T. Stütz, F. Autrusseau, and A. Uhl, "Non-blind structure-preserving substitution watermarking of H. 264/CAVLC inter-frames,” IEEE Transactions on Multimedia, vol. 16, pp. 1337-1349, 2014.

[6] Ardizzone, A. Bruno, G. Mazzola, “Copy-move forgery detection by matching triangles of key points”, IEEE Transactions on Information Forensics and Security vol.10, no.10, pp. 2084-2094, 2015

[7] J. Li, X. Li, B. Yang, X. Sun, “Segmentation-based image copy-move forgery detection scheme”, IEEE Transactions on Information Forensics and Security, vol. 10, no 3,pp.507-518,2015.

[8] V. Christlein, C. Riess, J. Jordan, C. Riess, E. Angelopoulou, “An evaluation of popular copy-move forgery detection approaches”, IEEE Transactions on information forensics and security, vol. 7,no 6,pp.1841-1854,2012.

[9] R. C. Pandey, S. K. Singh, K. Shukla, 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, pp.1-6, 2014.

[10] S. Prasad, B. Ramkumar, “Passive copy-move forgery detection using SIFT, HOG and SURF features”, in IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE, pp.706-710, 2016

[11] Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, “A sift-based forensic method for copy-move attack detection and transformation recovery,” IEEE Transactions on Information Forensics and Security, vol. 6, no. 3, pp. 1099–1110, Sep. 2011.

[12] W. Li and N. Yu, “Rotation robust detection of copy-move forgery,” in Proc. IEEE International Conference on Image Processing ICIP’10, pp. 2113–2116,2010.

[13] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features,” Computer Vision–ECCV, pp. 404–417, 2006.

[14] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Proc. IEEE computer Society Conference on Computer Vison and Pattern Recognition, CVPR’05, 2005.

[15] T. Van Lanh, K. Chong, S. Emmanuel, and M. Kankanhalli, “A survey on digital camera image forensic methods,” in Proc. IEEE International Conference on Multimedia and Expo ICME’07, pp. 16–19,2007.

[16] X. Pan and S. Lyu, “Region duplication detection using image feature matching,” IEEE Transactions on Information Forensics and Security, vol. 5, no. 4, pp. 857–867, Dec. 2010.

[17] Bestagini P, Milani Tagliasacchi M, Tubaro S (2013) Local tampering detection in video sequences, IEEE MMSP, pp. 488–493

[18] W. Wang and H. Farid, "Exposing digital forgeries in video by detecting duplication," in Proceedings of the 9th workshop on Multimedia & security, pp. 35-42, 2007.

[19] V. Subramanyam and S. Emmanuel, "Video forgery detection using HOG features and compression properties," in IEEE International Workshop on Multimedia Signal Processing, pp. 89-94, 2012

[20] M. Pun, X. C. Yuan, and X. L. Bi, "Image forgery detection using adaptive over segmentation and feature point matching,” IEEE Transactions on Information Forensics and Security, vol. 10, pp. 1705-1716, 2015.

[21] Ramesh Chand Pandey, Sanjay Kumar Singh and K.K.Shukla, “Passive Copy- Move Forgery Detection in Videos,” 5th International Conference on Computer and Communication Technology, pp.301-306,ICCCT-2014.

[22] C.-C. Hsu, T.-Y. Hung, C.-W. Lin, and C.-T. Hsu, "Video forgery detection using correlation of noise residue," IEEE 10th Workshop on Multimedia Signal Processing, pp. 170-174, 2008.

[23] M. Kobayashi, T. Okabe, and Y. Sato, "Detecting video forgeries based on noise characteristics," in Advances in Image and Video Technology, Springer, pp. 306-317, 2009.

Authors Profile

Miss. J.D.Gavade has completed her UG and PG in Electronics Engineering from Walchand College of Engineering, Sangali. Currently pursuing her Ph.D. from Shivaji University, Kolhapur. Currently working as Assistant Professor at DKTE's Textile and Engineering Institute, Ichalakranji. She is life member of ISTE. Her area of interest are image and video forensic, neural networks etc. She has total 16 years of experience in teaching.

Dr Mrs S.R.Chougule has completed her Ph. D. from Shivaji University. Currently working as Professor and HOD in Electronics and Telecommunication Department at KIT' College of Engineering, Kolhapur. She has total 23 years of teaching experience. Her are of interest includes image processing, communication and digital signal processing. She has total 67 publications in her account.

