Video Forgery Detection Based on Variance in Luminance and Signal to Noise Ratio using LESH Features and Bispectral Analysis

Aniket Pathak¹, Dinesh Patil²

¹Department of CSE, SSGBCOET Bhusawal, Maharashtra, India
²Associate Professor, Department of CSE, SSGBCOET Bhusawal, Maharashtra, India
¹aniketpathak89@gmail.com; ²dineshonly@gmail.com

Abstract— one among the most popular topics today is authentication of the digital content one has. Due to easy availability of advanced tools to modify content it has become difficult to judge whether a given content produced as piece of evidence is valid or forged one. The main aim of this paper to introduce a method based on statistical properties useful for detecting the forgeries in digital content specially videos. In this the key properties such as luminance and signal to noise ratio has been explored and use of Local Energy based shape histogram and Bispectral analysis. This paper mainly proposes a method which proves useful in detecting the different forms of tampering done with videos such copy-paste, inpainting etc. Experimental results shows that proposed method works in an efficient manner for videos tampered using different techniques.

Keywords — Video Forensics, Multimedia, Forgery Detection, Copy-paste tampering, LESH

I. INTRODUCTION

In recent few years the issue of authenticating a given multimedia content has become more and more complex as a result of possible diverse origins and the potential alterations that are been operated on it. As availability of much less expensive and easy to operate digital devices (such as digital cameras, smart-phones, etc.), together with the advancement of high-quality data processing tools and algorithms resulted in process of acquisition and processing much more simple. The task of validating a given multimedia content has become harder task because of the huge amount of possible alterations operated on it. Experts from field after various attempts has come up with techniques to trace such tampering much of them are related to image processing but are useful in extension to video forgery detection we see details about it section 2.

In paper we introduce a method to detect video forgery which can be concluded from the statistics obtained at the end. The detail idea is discussed later. The paper is organized in following manner Section 2 deals with related work done by the researchers. Section 3 deals with the proposed methodology discussing the framework, terminology along with implementation details. Section 4 overviews the results obtained and in last we conclude with the scope for the new researchers can expand in this field of video forensics.

II. RELATED WORKS

All Smartly used digital video editing techniques has been a concern increasing the difficulty in distinguishing the authentic video from the tampered one. For example, in [2] authors have referred to the
forgery created in film “Speed” by duplicating the frames hiding the actual activity. Detection technique in [1] used to detect the spatial and temporal copy paste tampering. As it’s challenging to detect this type of tampering in videos as the forged patch may invariably vary in terms of size, compression rate and type (I, B or P) or other changes such as scaling and filtering. The algorithm as in [1] is based on Histogram of Oriented Gradients (HOG) feature matching and video compression properties. The advantage of using HOG features is that they are robust against various signal processing manipulations. Image or frame can be represented by using a set of local histograms [2]. The copy-paste tampering performed in a convincing manner without much of difficulty and is practically difficult to detect. Therefore, it’s likely that copy-paste tampering often applied for forging video. Authors in [1] used the intrinsic properties of the captured media to detect the copy-paste tampering in a video. Copy-paste video forgeries are classified into two categories - spatial and temporal tampering. In first region is copied from a location in a frame and pasted to a different location on the same frame or other frames possibly after few modifications. While in second type, a complete frame may be duplicated. Also in addition to that complete region may also be duplicated across the frames at the same spatial location. In both motives may be to hide unfavourable actions or objects in a scene by pasting other objects or background, or to implant false evidences. Several techniques related to image have been proposed, but still there are very few in video forgery detection works. Among few image copy-paste forgery detection techniques, some algorithms are [4], [5], [6]. The techniques proposed in [4], [5], are based on Scale Invariant Feature Transform (SIFT) features matching. In [5] the authors used Fourier-Mellin Transform (FMT) for achieving robustness against geometric transformations. Few more techniques can be found in [7]. Although techniques in [4],[5] are used for spatial copy-paste forgery detection and cannot be easily extended for detecting temporal copy-paste forgery. The reason being algorithms in [3] [4] [5] assumed that the tampered region is at a different spatial location than the source region. However, copy-paste regions are spatially colocated. The forgery has been addressed in [1], [2]. In [1] authors use temporal and spatial correlation in order to detect duplications. A temporal correlation matrix is computed between all frames in a given sub-sequence of frames and, spatial correlation matrix is computed for each frame in a given sub-sequence. In addition, this technique assumes that the forged regions belong to the same video sequence which may be a limitation when considering forged patches from other videos. However, the authors of [1] proposed another forgery detection scheme [8] which detected forged regions belonging to different videos. However, this algorithm detects forgery only in interlaced or deinterlaced videos. In [3], the authors proposed detection of forged regions based on inconsistencies in noise characteristics, occurred due to the forged patches from different videos. However, as the noise characteristics depend on the intrinsic properties of the camera, the noise characteristics are not useful when the forged patch comes from the same video. In addition, the noise characteristics may not be estimated correctly under low compression rates. Anti forensic techniques have also been reported in [9] against few existing forensic techniques. In [10], the authors propose an algorithm for video in-painting tampering. The idea is to use the noise correlation properties between spatially collocated blocks to detect the forgery. Some other forgery detection techniques proposed as in [11], [12]. In these basic idea is that a forged video will be recompressed and the artifacts owing to double quantization of coefficients are used to detect forgery. However, the techniques [10] - [12] suffer from the limitation that they are robust for a limited range of compression rates only.

In image forgery detection [4] - [6] features such as SIFT, transforms such as FMT are used. While in video forgery detection normalized correlation [2], noise characteristics [3], [10], or quantization parameters [11], [12] are used. Although features such as SIFT gave good performance [4], [5] owing to their robustness, other features such as SURF [13] or HOG [14] were also used for detection purpose.

In few years, due to the advances of network technologies, low-cost multimedia devices, sophisticated editing software and wide adoptions of digital multimedia coding standards, digital multimedia applications have become popular in our life. However, digital nature of the media files, can now be easily manipulated, synthesized and tampered in numerous ways without leaving visible clues. Due to which integrity of video content can no longer be taken for granted and a number of forensic-related issues arise. Scheme are widely used for forgery detection are active schemes and passive schemes. With active schemes, the tampered region can be extracted using a pre-embedded watermark. However, this scheme must have source files to embed the watermark first; otherwise, the detection process will fail [12]. On contrary, the passive schemes extract some intrinsic fingerprint traces of image/video to detect the tampered regions. When a real-world scene is captured the information about the scene is processed by pipeline of various camera components as color filter array (CFA), demosaicing, white-balancing, automatic gain control (AGC), Gamma correction, post-processing, and JPEG compression, before the final digital image is generated. In many forensic applications, the intrinsic fingerprint traces, such as the process of CFA [15], [16], Camera Response Function (CRF) [17], sensor pattern noise [18][19][20], and compression artefacts [21] are used to detect tampering such as resampling [22], copy and paste, slicing [16][23], and double compression [24]. Besides, the imperfect information in the camera is also used for forensic applications [2]. Recently, sensor pattern noise has been successfully used as intrinsic sensor biometrics for nonintrusive forensic analysis [18][19][20]. Method proposed in [18] first extracts the pattern noise images in training images captured with some specific cameras and the reference pattern noise.
image can then be obtained via averaging operation. The correlation measurement between the reference pattern noise image and pattern noise image is used here. The sensor pattern noise has also been used for scanner model identification and tampering detection of scanned images [20].

III. PROPOSED METHOD

A. Problem Formulation

In the area of video forensic we propose a new approach using the basic properties such as luminance and signal to noise ratio. It is important that the detection is to be done on the content modified by one or more method. The earlier contribution by the researchers has focused on compression properties, HOG features, noise characteristics and more but over here we have focused on using the statistical properties and use the local nonlinearities. The problem lies given the multimedia content to validate it on basis of some reference. The main idea is to use the variance in luminance and signal to noise ratio and then find out the entropy and display the energy based histogram and tabulating the results gives the output that the video is forged if large variation in the values appears. In this section we further state some terms and workflow of method.

B. Luminance

In video, luminance termed as luma represents the brightness in an image (the "black-and-white" or achromatic portion of the image). It is typically paired with chrominance and represents the achromatic image, while the chroma components represent the color information. While luma is more often used, (photometric) luminance is sometimes used in video engineering when referring to the brightness of a monitor. The formula used to calculate luminance uses coefficients based on the CIE color matching functions and the relevant standard chromaticities of red, green, and blue (e.g., the original NTSC primaries, SMPTE C, or Rec. 709). For the Rec. 709 primaries, the linear combination, based on pure colorimetric considerations and the definition of luminance is: 

\[ Y = 0.2126 R + 0.7152 G + 0.0722 B \]

C. LESH

LESH (Local Energy based Shape Histogram) is an image descriptor in computer vision used to get a description of the underlying shape and is built on local energy model of feature perception. It encodes the shape by accumulating local energy of the underlying signal along several filter orientations, several local histograms from different parts of the image/patch are generated and concatenated together into a 128-dimensional compact spatial histogram, also it is designed to be scale invariant. The LESH features are useful in applications like shape-based image retrieval, object detection, and pose estimation as in [24]. The Local Energy Model was first introduced in [22] proving that features can be extracted at those points from an image where local frequency components represent maximum uniformity. The extended framework of local energy model is given in Equation below and is normalized by the summation of noise cancellation factor \( T \), Sin of phase deviation and factor \( W \), which is the weighting of the frequency spread. For more details of this extended framework refer contribution in [23].

\[
E = \sum_n W(x) A_n(x) \left[ \cos(\phi_n(x) - \bar{\phi}(x)) - \sin(\phi_n(x) - \bar{\phi}(x)) \right] - T
\]

Local Energy gives reliable information to extract the interest points from an image in an invariant manner to illumination and noise. This raw energy indicates the corners, contours or edges of underlying shape in an image. LESH [24] features are obtained firstly by dividing the candidate image into 16 sub-regions and then the local energy information is calculated for each sub-region along 8 different orientations with the help of Gabor Wavelets kernels [25]. The local histogram \( h \) is calculated in equation as follows:

\[
w_r = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-x_{ro})^2+(y-y_{ro})^2}{2\sigma^2}}
\]

\[
h_{r,b} = \sum w_r \times E \times \delta_{lb}
\]

Where \( w \) is the Gaussian weighting function of region \( r \) calculated as in equation, \( E \) represents the local energy computed as equation (8), \( Lb \) represents Kronecker’s delta, \( L \) orientation label map and \( b \) current bin. From the above description it can be seen that the LESH descriptor of a shape is 8x16=128 dimensional feature vector.
D. Proposed Framework

The proposed work implemented can be stated as follows is also illustrated in figure

a. Firstly input the video which is to be authenticated and along with it the reference video
b. Both videos are converted into GOP according to frame rate respectively
c. Using the formula find luminance after separating the RGB components.
d. For each frame of both videos find the local entropy and the average maximum generalized Emax and Emin value
e. Compare frames from both videos and state number of modified frames.
f. For the same generate the histogram based on the entropy energy of pixels and calculate the Ediff and the signal to noise ratio variation in unmatched frames
g. Observe the results obtained if Emax and Emin remain same for both video we guarantee that video is original and not tampered if not so we say that forgery is detected.
h. Simultaneously we take both the video signals and apply bispectral analysis if the unnatural correlations or non-linearities are found then forgery is detected else we say authenticate video.

![Proposed System Framework](image)

Fig. 1

IV. EXPERIMENTAL RESULTS

The dataset is composed by video sequences: original and forged ones. Each sequence has a resolution of 320x240 pixels, and a frame-rate of 30 fps. Original sequences have been recorded using low-end devices, thus they have all been compressed at the origin (using either MJPEG or H264 codecs). Forged sequences have been saved as uncompressed file (RV24, 24 bit RGB). The sequences come from the Surrey University Library for Forensic Analysis (SULFA) database [34].
Fig. 2 SULFA Dataset
(a-c) Original Video sequences
(d-f) Forged Video Sequences for Cricket Ball rolling across the table.
(h-j) Original Video sequences
(k-m) Forged Video Sequences for outdoor lake view with ducks.

For testing purposes from the dataset we had taken the few sequences includes (a) Cricket Ball rolling over a table top. (b) Outdoor lake view with two ducks (c) Moving van with maroon colour and car passing across street (d) Inpainted video based on contributions in [36] man walking across a lamp post (e) girl walking along a building path.

The table below give the summarize details of the video sequences used with their respective frames generated, unmatched frames in original and forged one with frame rate and bit rate along with the energy values computed.

<table>
<thead>
<tr>
<th>Video Sequences Used</th>
<th>No. of Frames</th>
<th>No. of Unmatched Frames</th>
<th>Frame Rate in fps</th>
<th>Bit Rate in kbps</th>
<th>Energy Value of Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cricket ball rolling across the table</td>
<td>319</td>
<td>63</td>
<td>30</td>
<td>200</td>
<td>Emin 7.3, Emax 7.4617</td>
</tr>
<tr>
<td>Cricket ball rolling across the table</td>
<td>319</td>
<td>63</td>
<td>30</td>
<td>200</td>
<td>Emin 7.3, Emax 7.4715</td>
</tr>
<tr>
<td>Lake outdoor scene</td>
<td>210</td>
<td>210</td>
<td>30</td>
<td>200</td>
<td>Emin 7.3661, Emax 7.5098</td>
</tr>
<tr>
<td>Lake outdoor scene</td>
<td>210</td>
<td>210</td>
<td>30</td>
<td>200</td>
<td>Emin 7.2853, Emax 7.4327</td>
</tr>
<tr>
<td>Van and a car passing</td>
<td>412</td>
<td>128</td>
<td>30</td>
<td>200</td>
<td>Emin 7.0764, Emax 7.2943</td>
</tr>
<tr>
<td>Van and a car passing</td>
<td>412</td>
<td>128</td>
<td>30</td>
<td>200</td>
<td>Emin 6.38, Emax 6.5197</td>
</tr>
<tr>
<td>Girl walking inpainting</td>
<td>48</td>
<td>48</td>
<td>10</td>
<td>1840</td>
<td>Emin 6.4867, Emax 6.674</td>
</tr>
<tr>
<td>Girl walking inpainting</td>
<td>48</td>
<td>48</td>
<td>10</td>
<td>1844</td>
<td>Emin 6.519, Emax 6.6849</td>
</tr>
<tr>
<td>Moving person inpainting</td>
<td>49</td>
<td>49</td>
<td>10</td>
<td>1851</td>
<td>Emin 6.4867, Emax 6.674</td>
</tr>
<tr>
<td>Moving person inpainting</td>
<td>49</td>
<td>49</td>
<td>10</td>
<td>1851</td>
<td>Emin 6.38, Emax 6.5197</td>
</tr>
</tbody>
</table>

The variance in the energy levels of the original and forged video sequences along with maximum value for each is shown in the figure 3 below.

Fig. 3 (a) Graph showing the Emax and Emin values for the original and forged video sequences (b) Graph showing the parameters of the videos in dataset such as bit rate and frame rate.
The figure 4 below shows the difference in energy levels frame wise between the video sequences forged using different methods red shows the forged while blue shows the original video sequences (a) the cricket ball rolling over across the table tampered using frame duplication (b) The copy paste tampering in outdoor scene of lake view (c) inpainting used to forged the man walking across the lamp post on street.

Fig. 4

The figure 5 below on the left shows sample frame of original and forged video sequence and to right is computed histogram based on the entropy values. Figure shows for outdoor lake scene having duck below it is the variance in the signal to noise ratio for both sequences and generated histogram.
The figure 6 below on the left shows sample frame of original and forged video sequence and to right is computed histogram based on the entropy values. Figure shows for Cricket ball rolling across the table below it is the variance in the signal to noise ratio for both sequences and generated histogram.
Fig. 6 (a) Sample frame of original video cricket ball rolling across table. (b) Histogram showing the values of energy for each frames (c) Sample frame of forged video cricket ball rolling across table. (d) Histogram showing the values of energy for each frames (e) PSNR value difference for frames that differ in both video sequences.

The results shows that when the video sequences are forged using method such as copy paste, frame/region duplication or inpainting done we get an indication of the forgery based on the method proposed as the variance in Luminance seen in the LESH is itself indicative that the video is forged also the varying signal to noise ratio also proves it. We over here have taken the two video sequences and illustrated the results in above figures. The table below shows the possible tampering method used the difference in energy level and signals to noise ratio.

<table>
<thead>
<tr>
<th>Video Sequences</th>
<th>Possible Tampering method</th>
<th>Modified Frames out of total frames in video sequences</th>
<th>Luminance Variance LESH values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cricket ball rolling across the table</td>
<td>Frame duplication</td>
<td>63 out of 319</td>
<td>$E_{\text{min}}$, $E_{\text{max}}$</td>
</tr>
<tr>
<td>Outdoor scene of lake view</td>
<td>Region duplication</td>
<td>210 out of 210</td>
<td>0.0808, 0.0771</td>
</tr>
<tr>
<td>Man walking across lamp post</td>
<td>Inpainting</td>
<td>49 out of 49</td>
<td>0.1067, 0.1543</td>
</tr>
</tbody>
</table>

The last column in table is the indicative measure for the three different types of tampering proposed method shows that the forgery has occurred and can be simplified. The one major concern is that we always need a reference video to deal with to compare between the original and forged one.

V. CONCLUSIONS

In this paper we introduce a new technique for forgery detection based on the variance in luminance and signal to noise ratio using LESH features. The main focus was on GOP considered and the unmatched frames found in which variation was found. Experimental results showed that this method is useful and efficient in detection of forgery done using more than one method. Also as use of statistical properties make its slight better as compared to the other schemes available at present. It would be interesting to find out as to many more tampering ways can be detected using this method. At present we have shown three methods on the basis of dataset available. Video forensics still remains very large research domain we have only just proposed a method and hope it to be an useful contribution so that the forgery in video is detected.

ACKNOWLEDGMENT

We would like to dedicate this work our parents and would like to thank all the staff members from Department of CSE, SSGBCOEET Bhusawal Maharashtra. Special thanks to Dr. R.P. Singh our Principal for his value guidance and support. Lastly all those who directly or indirectly supported us.
REFERENCES


