COMPRESSION NOISE BASED VIDEO FORGERY DETECTION

Hareesh Ravi†, A.V. Subramanyam†, Gaurav Gupta‡, B. Avinash Kumar†

† Electronics and Communication Engineering
†† Indraprastha Institute of Information Technology
‡ Department of Electronics and Information Technology
New Delhi, India

ABSTRACT
Intelligent video editing techniques can be used to tamper videos such as surveillance camera videos, defeating their potential to be used as evidence in a court of law. In this paper, we propose a technique to detect forgery in MPEG videos by analyzing the frame’s compression noise characteristics. The compression noise is extracted from spatial domain by using a modified Huber Markov Random Field (HMRF) as a prior for image. The transition probability matrices of the extracted noise are used as features to classify a given video as single compressed or double compressed. The experiment is conducted on different YUV sequences with different scale factors. The efficiency of our classification is observed to be higher relative to the state of the art detection algorithms.

Index Terms— Video Forgery Detection, Double Quantization Noise, Markov Process

1. INTRODUCTION
Video cameras and surveillance systems are being increasingly used in today’s world and many of these systems utilize MPEG (MPEG-2 and MPEG-4) encoding for compressing the captured video. In order to forge the videos captured using these systems, an adversary has to decompress it first, forge the video and re-compress the forged video while saving. As this is often the case, double compression can identify a video forgery. The detection of forgery is also of paramount importance for Law Enforcement Agencies during forensic investigation as they need to verify the integrity of a video in question, which could be a potential evidence. Although video forgery requires high levels of sophistication, some convincing forgeries have been pointed out in literature [1]. Figure 1 shows a kind of forgery where from a video, certain frames are deleted in such a way that there is only one person shown walking along the corridor when there were actually two. Several forgery detection techniques have been proposed till date [2 - 10]. In the technique proposed in [2] the basic idea is that, in a recompressed video the statistics of quantized or inverse quantized coefficients exhibit a deviation from that of original video. In [3, 4], noise characteristics are used to detect forgery. In [5], the authors detect double compression by capturing empty bins exhibited in the distribution of quantized coefficients in a recompressed video. Wang et al [6] also proposed detection of MPEG-4 video double compression by Markov modeling of difference of DCT coefficients. The techniques [7, 8] are also based on similar principles while [9, 10] proposed forgery localization techniques. However, these [2, 5, 6] techniques have limitations over the relationship between the scaling factors used for first and second compression. In order to overcome the limitations of aforementioned references, we use compression noise for detection of double compression thereby detecting forgery. The compression noise present in spatial domain in a video has been shown to be correlated [11]. When a single compressed video is re-compressed, the correlation of spatial domain noise is disturbed. This phenomenon can be effectively captured using Markov process and can be used for forgery detection by detecting double compression.

In this paper, we propose a video forgery detection scheme by detecting double compression. A block diagram of the scheme used is given in Figure 2. In order to extract compression noise from a given video frame, we use a modified HMRF prior model [11]. The prior model is modified in order to incorporate the effect of compression. Since, Markov statistics has been proven to be a distinguishable feature for single and double compression in JPEG images [12] and
The noise extraction process is explained as follows. The parameters of quantization error [11], as derived in Section 2 can be used probabilistically to remove compression artifacts. In this removal technique, the quantization error becomes a likelihood term that ensures that the final frame estimate agrees well with the observed data. A maximum a posteriori (MAP) criterion is used for estimating the denoised image as,

\[
\hat{Z} = \arg \max_Z p(Z|Z_q) \quad (5)
\]

where \( \hat{Z} \) is the final frame estimate after removing the compression noise. Equation (6) considers a priori term and a maximum likelihood term. The likelihood can be determined from eq (2) as \( Z_q = Z + e_z \), \( Z_q \) which is a Gaussian random variable with mean \( Z \) and auto covariance \( K_{ez} \). Uniform frequency domain model [11] is used for the likelihood term. The prior model will be based on Huber Markov Random Field (HMRF) wherein the Huber function is as follows,

\[
p(Z) = \frac{1}{G}exp\left(-\lambda \sum_{c \in C} \rho_T(d^c_iZ)\right) \quad (7)
\]

Where \( G \) is a normalizing constant, \( \lambda \) is a regularization parameter, \( c \) is a local group of pixels called cliques and \( C \) is the set of all such cliques which depends on neighbourhood structure of the Markov random field. The Huber function \( \rho_T(\cdot) \) is defined as,

\[
\rho_T(u) = \begin{cases} 
  1u^2, & |u| \leq T, \\
  l(T^2 + 2T(|u| - T)), & |u| > T 
\end{cases} \quad (8)
\]

where,

\[
l = \begin{cases} 
  1 & \forall Z(m,n) : m, n \notin S, \\
  1.5 & \text{otherwise}
\end{cases} \quad (9)
\]

where, we introduce \( l \) as weight in order to incorporate the effect of single compression on a frame and \( m, n \) is the indices of the frame \( Z \) of dimension \( M \times N \), \( S \) is the set of pixels which belong to the border pixels in each 8x8 block. \( d^c_i \) extracts the differences between a pixel and its neighbors so that,

\[
p(Z) = \frac{1}{G}exp\left(-\sum_{n=0}^{M-1} \sum_{m \in N_n} \rho_T(Z[n] - Z[m])\right) \quad (10)
\]

Where \( N_n \) is the index set of neighbors for the nth pixel, and \( M \) is the number of pixels in the frame. Now eq (6) can be written as

\[
\hat{Z} = \arg \max_Z p(Z|Z_q) \quad (5)
\]

where \( \hat{Z} \) is the final frame estimate after removing the compression noise. Equation (6) considers a priori term and a maximum likelihood term. The likelihood can be determined from eq (2) as \( Z_q = Z + e_z \), \( Z_q \) which is a Gaussian random variable with mean \( Z \) and auto covariance \( K_{ez} \). Uniform frequency domain model [11] is used for the likelihood term. The prior model will be based on Huber Markov Random Field (HMRF) wherein the Huber function is as follows,
The video files are obtained from various open sources [14] in 4 : 2 : 0 Common Intermediate Format (CIF) of resolution 352×288. Sixteen different sequences of 300 frames each are taken and are encoded using ‘ffmpeg’ MPEG encoder. Details of MPEG-2 video detection are given here and that of MPEG-4 is discussed in Section 4.3. The encoding sequence is considered as “IIPPP” and these 5 frames constituted a single Group Of Picture (GOP). All the clips were first encoded in Variable Bit Rate mode with Quality scale factor (QF) ranging from 2 to 15. In order to simulate forgery, 28 frames are deleted from the middle (frames 221 to 248) of single compressed videos. These videos are then compressed again with different scaling factors. Each YUV sequence is divided into 10 clips of 30 frames or 6 GOPs each. Totally 160 clips are considered for each scale factor pair (single compression scale factor QF1 and double compression scale factor QF2) resulting in 160*162 (total number of pairs) = 25920 clips. In Table 1, values for scale factors such as 5, 7, 8, 11 and 12 are not given due to space constraints and to include a broader range of values. However, these values also give results similar to those given in the table.

4.1. Classification

For each compression pair as given in Table 1, the total number of samples available is 320 (160 each for single and double). 50% of the total samples was trained using SVM with linear kernel and other parameters being set to default [15]. The testing samples constituted the other 50% of the total samples. It was ensured that a sequence if present in the training sample will not be a part of the testing samples. The experiment was repeated for 10 times by changing the training and testing samples each time maintaining the 50-50 ratio. Each frame is considered for classification and based on a voting mechanism, when the number of frames classified as authentic/single compressed in a given clip is above a threshold th = 0.5 ([16]), then the clip is classified as single compressed. Similarly, the clip is classified as forged when the number of frames classified as forged/double compressed is above th.

4.2. Performance Comparison

Classification accuracy for each compression pair is given in Table 1. Here, the accuracy is given as (TPR + TNR)/2, where TPR is the ratio of classified forged clips to that of total number of forged clips. TNR is the ratio of classified authentic clips to that of total number of authentic clips. It is observed that the accuracy is more than 95% except for very few pairs like 9-10,13-14,14-15 and 14-13 but are still considerably higher. Further, it is also observed that the accuracy is 100% for most of the pairs in the lower left of the Table 1 as well as for a few in the upper right.
Table 1. Accuracy Rate for Various Compression Pairs

<table>
<thead>
<tr>
<th>QF₁ \ QF₂</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>9</th>
<th>10</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>x</td>
<td>94</td>
<td>95</td>
<td>97</td>
<td>98.9</td>
<td>99.4</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>97</td>
<td>x</td>
<td>95.8</td>
<td>96.3</td>
<td>98.5</td>
<td>99.7</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>96</td>
<td>95.3</td>
<td>x</td>
<td>93</td>
<td>96.8</td>
<td>97.3</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>99.4</td>
<td>96.5</td>
<td>94</td>
<td>x</td>
<td>97.6</td>
<td>97.4</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>99.2</td>
<td>98.6</td>
<td>95.4</td>
<td>x</td>
<td>83.4</td>
<td>90.2</td>
<td>97.9</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>100</td>
<td>97.5</td>
<td>97.5</td>
<td>94.6</td>
<td>x</td>
<td>88</td>
<td>93.6</td>
<td>95</td>
</tr>
<tr>
<td>13</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>95.2</td>
<td>93.8</td>
<td>x</td>
<td>82.8</td>
<td>82.4</td>
</tr>
<tr>
<td>14</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>94.2</td>
<td>x</td>
<td>75.6</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>84.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Detection accuracy comparison

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Odd Multiple</td>
<td>98.98%</td>
<td>51.53%</td>
<td>50.32%</td>
</tr>
<tr>
<td>Even Multiple</td>
<td>98.62%</td>
<td>96.28%</td>
<td>59.46%</td>
</tr>
</tbody>
</table>

Fig. 3. ROC Curve showing True positive and False positive rate for three different scaling factors and the average.

5. CONCLUSION

An efficient method to detect forgeries in video by detecting double compression is proposed. The effectiveness of this method is threefold. First, the detection accuracy rate is above 95% for all scale factors and in most of the cases, the efficiency is as high as 100%. Second, modeling of compression noise as Markov process clearly characterizes the form of compression which is single or double. It also detects double compression in both MPEG-2 and MPEG-4 videos. This is also validated through experimental results. Further, the proposed algorithm performs better than most of the present techniques. In future works, we want to perform the localization of tampering in a video. This localization can be in terms of GOP, frames or as small as a macroblock.

6. ACKNOWLEDGEMENT

The authors thank the ‘Cybersecurity Education and Research Centre (CERC)’, IIITD for partially funding this work.
7. REFERENCES


