PIXEL ESTIMATION BASED VIDEO FORGERY DETECTION

A.V. Subramanyam

Electronics and Communications Engineering
Indraprastha Institute of Information Technology
Delhi, India

Sabu Emmanuel

School of Computer Engineering
Nanyang Technological University
Singapore

ABSTRACT

In this paper, we propose a novel technique to detect double quantization, which results due to double compression of a tampered video. The proposed algorithm uses principles of estimation theory to detect double quantization. Each pixel of a given frame is estimated from the spatially colocated pixels of all the other frames in a Group of Picture (GOP). The error between the true and estimated value is subjected to a threshold to identify the double compressed frame or frames in a GOP. The advantage of this algorithm is that it can detect tampering of I, P or B frames in a GOP with high accuracy. In addition, the technique can also detect forgery under wide range of double compression bitrates or quantization scale factors. We compare our experimental results against popular video forgery detection techniques and establish the effectiveness of the proposed technique.

Index Terms— Video Forgery Detection, Temporal Tampering, Estimation, Double Compression

Fig. 1. (a) Forged video frame (b) Original video frame

Fig. 2. Top row: forged video sequence. Bottom row: original video sequence

1. INTRODUCTION

The intelligent use of digital video editing techniques is constantly increasing the difficulty in distinguishing the authentic video from the tampered one. For example, Figures 1(a) and 1(b) show the frames from forged and original video respectively. In this forgery, the motive is to create an ambiguity in the entrance of the person.

Several forgery detection techniques have been proposed till date [1, 2, 3, 4, 5, 6, 7]. In the technique proposed in [1] 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. And this difference in statistics is used to detect double compression. In [2, 3], noise characteristics are used to detect forgery. In [4, 6] the authors present a technique to detect double compression by capturing empty bins exhibited in the distribution of quantized coefficients in a recompressed video. However, the technique proposed in [6] can only detect a double compressed I frame in variable bit rate mode only i.e. constant quantization scale factor. In [5], the authors use temporal and spatial correlation in order to detect duplications.

The techniques proposed in [1, 4, 6] cannot detect if one or more B or P frames are authentic or forged. This is particularly necessary, as in scenarios such as video surveillance, the frame rate may be low and the existence of an evidence may be captured in only a few frames. A forgery of this kind is shown in Figure 2, where initial frames of original video are edited to hide the entrance of the person. It is not clear from the forged sequences, if the person is entering from the door on his left or simply walking down the corridor. Here, the first three frames are edited using uncompressed background frames. And when the video undergoes a double compression, the edited frames get single compressed while the rest of the frames get double compressed. Thus, the techniques [1, 4, 6] which capture double compression cannot detect this kind of forgery. Further, the techniques presented in [2, 3, 5] suffer from the limitation that they can capture only copy paste related forgery or video inpainting tampering. And are practically not suitable for detecting other types of forgery such as mentioned in Figure 2. Some other forensic techniques can be obtained in [7, 8].

In this paper, we propose a forgery detection technique using the principles of estimation theory and double compression for videos captured from static cameras. The estimation part takes all the frames present in a GOP into account and therefore captures any artifacts occurring due to editing and double compression of one or more frames. Unlike the previously proposed algorithms which takes only I frame into account [4, 6] or collectively all the frames in a GOP [1], the proposed algorithm deals with a given frame under test against all the other frames in a GOP. Thus, it is capable of distinguishing between a single and double compressed frame or frames in a GOP rather than a less pronounced case of detecting only an I frame or a complete GOP as authentic or forged.

In the proposed algorithm, the pixels of a given I, P or B frame under test in a GOP are estimated from the spatially colocated pixels from all the other frames in that GOP. The percentage of error between the true value and estimated value lying inside a given interval is calculated. In order to detect double compression, the error percentage is calculated for bit rates ranging from slightly above the bit rate extracted from the video in question to the lower possible bitrates. Further, the difference between the error percentages for different bit rates are used as a measure, as explained in Section 2.3, to judge whether a given frame is authentic or forged. The rest of the paper is organized as follows. In Section 2, we explain the proposed algorithm using pixel estimation and double compression. The experimental results are given in Section 3 and Section 4 concludes the paper.

2. PROPOSED ALGORITHM

In this proposed scheme, we begin with the discussion of estimation of the pixels as described in Section 2.1. Then we explain the effect of double compression on the estimation of pixels which is presented in Section 2.2. The error between the true and estimated pixel values exhibits a lower variation in case of double compression compared to that of single compression under different recompression bitrates. And this particular property is used to detect the forged video which is explained in Section 2.3. Here, we focus on detection of authentic or forged video for the particular case when second compression bitrate is higher than the first compression bitrate.

2.1. Pixel Estimation

A pixel of a given frame in a GOP is estimated using all the other spatially colocated pixels in that GOP. A typical GOP is given by [9] I_1 B_2 B_3 P_4 B_5 B_6 P_7 B_8 B_9 P_{10} B_{11} B_{12} where the indexes indicate the time sequence. We use the luminance component for estimation purpose. Let $\mathcal{R}$ denotes the time index set, where $\mathcal{R} = \{1, 2, 3, ..., 12\}$. Let the frame $(I, P,$ or $B)$ whose pixels are to be estimated be located at time index R, where $R \in \mathcal{R}$. Let a set $\mathcal{S}$ be defined as $\mathcal{S} = \mathcal{R} - \{R\}$. Then the estimation model for R^{th} frame is given by eq (1)

$$\mathbf{Z} = \mathbf{A} + \mathbf{W} \tag{1}$$

In eq (1) all the variables are random vectors and, $\mathbf{Z} = \{\mathbf{z}_i\}\ \forall\ i = 1, 2, .., L$, where $\mathbf{z}_i = \{z_i(k)\}\ \forall\ k \in \mathcal{S}$ represents the spatially colocated pixel observations from the frames with time index $k \in \mathcal{S}$. $\mathbf{A} = \{a_i\}\ \forall\ i$ represents the true value which is the corresponding pixel value in R^{th} frame which is to be estimated. $\mathbf{W} = \{\mathbf{w}_i\}\ \forall\ i$, where $\mathbf{w}_i = \{w_i(k)\}\ \forall\ k \in \mathcal{S}$ denotes additive white Gaussian noise (AWGN) with zero mean and variance σ^2. Noise is modeled as AWGN with zero mean and variance to make the mathematical formulation easier. $L = M \times N$ denotes the product of the dimensions of the frame. In order to find an efficient estimator for $\mathbf{A}$, we first find the likelihood function and then apply the Cramer-Rao Lower Bound (CRLB) [10].

$$\begin{aligned} p(\mathbf{Z}; \mathbf{A}) &= \prod_{k \in \mathcal{S}} \frac{1}{\sqrt{2\pi}\sigma} exp\Big[-\frac{1}{2\sigma^2}(\mathbf{Z} - \mathbf{A})^2 \Big] \\ &= \frac{1}{(2\pi\sigma^2)^{|\mathcal{S}|/2}} exp\Big[-\frac{1}{2\sigma^2} \sum_{k \in \mathcal{S}} (\mathbf{Z} - \mathbf{A})^2 \Big] \end{aligned} \tag{2}$$

where $|.|$ denotes cardinality of the set $\mathcal{S}$. On taking the first derivative of $p(\mathbf{Z}; \mathbf{A})$ and solving, we can find that the estimate is given by $\hat{\mathbf{A}} = \frac{1}{|\mathcal{S}|} \sum_{k \in \mathcal{S}} \mathbf{Z}$.

Now, the pixelwise error $\mathbf{E}$ between $\mathbf{A}$ and $\hat{\mathbf{A}}$ is given by $\mathbf{E} = \mathbf{A} - \hat{\mathbf{A}}$. Then, the percentage of error values e_i lying in an interval $(-e_{th}, e_{th})$ is given by eq (3)

$$E_N = \sum_{i=1}^{L} \frac{\mathbb{I}_{[e_i \in (-e_{th}, e_{th})]}}{L} \times 100 \tag{3}$$

where $\mathbb{I}$ is an indicator function which gives output 1 if the input belongs to the given set, else the output is 0. E_N is used as a measure to detect double compression of an I, P or B frame in a GOP which is discussed in the Section 2.3. Next, we discuss the effect of double compression.

2.2. Double Compression

Let us consider a DCT coefficient x which gets quantized using a stepsize q_1. Then the reconstructed coefficient will be given by eq (4)

$$x_{q_1} = \left[\frac{x}{q_1} \right] \times q_1 \tag{4}$$

where $q_1 = qp_1 q_{mat}$ and qp_1 denotes the quantization scale factor while q_{mat} denotes the value from quantization matrix [9]. And [.] denotes rounding of function.

Let the second quantization occurs with stepsize q_2. We discuss the case $q_1 > q_2$ followed by the case $q_1 < q_2$. The reconstructed coefficient after second quantization is given by,

$$\begin{cases} y_{q_2} = \left[\frac{x_{q_1}}{q_2} \right] q_2, \text{and,} \\ x_{q_1} - \frac{q_2}{2} \le y_{q_2} \le x_{q_1} + \frac{q_2}{2} \end{cases} \tag{5}$$

Now, when $q_1 > q_2$, y_{q_2} can be written as, $x_{q_1} - \frac{q_1}{2} < y_{q_2} < x_{q_1} + \frac{q_1}{2}$

On the other hand, when $q_2 > q_1$,

$$\begin{cases} y_{q_2} \notin (x_{q_1} - \frac{q_1}{2}, x_{q_1} + \frac{q_1}{2}), \text{and,} \\ (x_{q_1} - \frac{q_1}{2}, x_{q_1} + \frac{q_1}{2}) \subset (x_{q_1} - \frac{q_2}{2}, x_{q_1} + \frac{q_2}{2}) \end{cases} \tag{6}$$

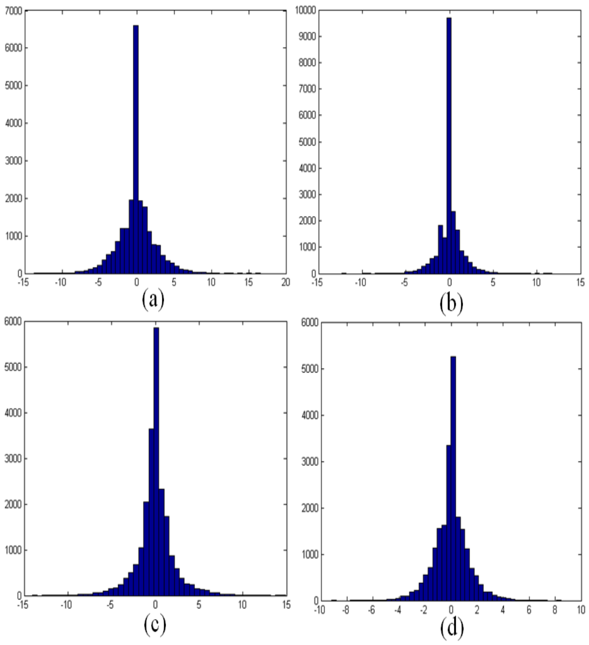

Fig. 3. Histogram of $\mathbf{E}$ (a) Authentic case (single compressed) - I, P or B frames (b) Forged case - I frame (c) Forged case - P frame (d) Forged case - B frame

Hence, we can see that the noise induced in the later case is more. We now use this observation and explain the forgery detection in the following section.

2.3. Forgery Detection

The double compression as explained in Section 2.2 causes E_N to decrease when the stepsize for double compression $q|q > q_1$, while E_N may remain similar for $q|q_2 \leq q < q_1$. This is because the higher amount of noise in case of $q > q_1$ increases the estimation error. This is also visible from the histogram of $\mathbf{E}$ as shown in Figure 3. Figure 3(a) shows the histogram for the authentic case. While Figure 3(b) shows the histogram when I frame is estimated from rest of the frames for forged case. Similarly, Figures 3(c) and 3(d) show the histogram for estimation of P and B frames respectively for forged case. In Figure 3(a), the spread of $\mathbf{E}$ is higher than that of Figure 3(b) suggesting higher error in the former case. However, this is not the case always. Let us consider a case when a coefficient gets singly compressed with q_{1s}, and in another case, the same coefficient gets doubly compressed with q_{1d} followed by q_2, such that $q_{1s} \ll q_{1d}$ and $q_{1d} > q_2$. Then, the noise induced in the single compression is less compared to double compression simply because higher number of bits are allocated for single compression. Towards this, we need to compare E_N obtained at different bitrates to detect double compression as explained below.

In order to capture double compression, the video is recompressed using bitrates $B_j = B + (1 - 2j)\triangle \ \forall \ j \geq 0$ such that $\forall \ j \ B_j > 0$, where B is the extracted bitrate from the video in question and $\triangle$ is the amount of change in bitrate. In other words, the quantization scale factor will be increasing with decreasing bitrate. And E_N is computed for each recompression. Let us define the difference between E_N for the different bitrates as given in eq (7).

$$T_j = E_{N_{B+(1-2j)\triangle}} - E_{N_{B+(1-2(j+1))\triangle}} \qquad (7)$$

In practice, sharp boundaries are difficult to find which can differentiate authentic from a forged frame. Therefore, to find an efficient threshold, first of all E_N is rounded of. Then, we experimentally set the threshold n_{th} in eq(8). If the following criteria is satisfied, the video is declared to be authentic.

$$\{T_j > n_{th}\} \bigcup \{T_j > 0 \bigcap T_{j+1} > n_{th}\} \text{ for } j = 0 \qquad (8)$$

If the criteria in eq (8) is unsatisfied then we check for double compression using eq (9)

$$\{T_j > n_{th}\} \bigcup \{T_{j=1} < -n_{th}\} \ \forall \ j > 0 \qquad (9)$$

In eq (9), the case $T_{j=1} < -n_{th}$ is taken for the case when a single B or P frame is considered for double compression.

We find that certain GOPs may not satisfy any of the criteria and hence they are discarded. This may happen when the statistics of the quantized coefficients exhibit a highly zero-centric distribution which is common in flat regions.

Table 1. False positives, average of 600kbps and 900kbps

e_{th}	FP	FP†	Discarded (%)	Discarded† (%)
1	.09	.02	26.6	24.8
2	0.06	0.01	21.7	22
3	.008	0	24.2	21.7
4	.03	0.008	33.2	33.15

3. EXPERIMENTAL RESULTS

In our experiments we generated a total of 7680 GOPs from the following videos - Salesman, Akiyo, Paris, Silent, Grandma and Hall Monitor [11]. In order to compute false positives (incorrectly detecting authentic as forged) 160 GOPs are first compressed using 900kbps and 500kbps using TM5 MPEG-2 codec. Both the compressed videos are recompressed using 1000kbps, 800kbps, 600kbps and 400kbps resulting in a total of 1240 GOPs. For computing true positives (correctly detecting forged as forged) single (or first) compression bitrates are 300kbps, 500kbps, 600kbps and 700kbps. And double compression bitrates for video originally compressed with 300kbps are 600kbps and 700kbps while for video originally compressed with 500kbps, 600kbps and 700kbps, the double compression is 900kbps. 6400 GOPs are generated for computing true positives. The value for $n_{th} = 1$ and $\triangle$ = 100kbps.

Table 2. True positives

e_{th}	TP	TP†	Discarded (%)	Discarded† (%)
1	.97	.99	24.8	23.9
2	.89	.92	14.4	13.2
3	.88	.90	12.2	9.2
4	.91	.92	15.3	14.6

Table 3. False positives for individual frames

Frame	B_2	P_4	B_5	P_7	B_8	P_{10}
FP	0.03	0.06	0.03	0.08	0.06	0.08
FP†	0.00	0.00	0.01	0.02	0.01	0.03

3.1. Detection Performance

Tables 1 and 2, give the false positive (FP) and true positive (TP) rates respectively. We mention the FP and TP rates for different thresholds e_{th} and also report the percentage of GOPs which can neither be detected as forged nor authentic. The symbol $'\dagger'$ denotes the result after 1x3 median filtering. Here, based on different experimental results we use threshold $e_{th} = 1$ for reporting further results for I or P frames while $e_{th} = 2$ for B frames. The threshold is set higher for B frames as they offer highest compression and more noise is induced in B frames. From Tables 1 and 2, it is clearly evident that the detection rates are good. The FP rate after median filtering is 0.02 for $e_{th} = 1$, while TP rate is 0.99. Further, we also observe that median filtering improves the result as the false positive and false negative (incorrectly detecting forged as authentic) cases may occur in isolation. Such isolated cases can easily be removed by median filtering to improve the detection performance. The percentage of discarded GOPs are 20.45% and 19.85% for authentic and forged cases respectively.

Table 3 gives the FP rates for frame types P and B for e_{th} = 1 and e_{th} = 2 respectively. We can see that the FP rates are very low ($\approx 10^{-2}$) while testing individual frames for authenticity.

3.2. Performance Evaluation

We perform an evaluation of our results against the existing video forgery detection algorithms based on detecting double compression [1, 4]. Table 4 gives the detection accuracy (DA) comparison, where DA = (TP + TN)/2 and TN (true negatives) refers to correctly detecting authentic as authentic.

From the Table 4, we can observe that our algorithm performs equally good in case of detection of a GOP as forged or authentic with DA $>$ .95 for all the algorithms. Here, we consider a GOP as inauthentic if the I frame is detected to be inauthentic. The results are also reported for detecting B or P frames forgery in case a GOP is single compressed except one double compressed B or P frame as shown in Figure 2. It is clear from Table 4 that a GOP and a P frame can be detected as authentic or forged with an accuracy of 0.98 and 0.94 respectively. While B frame can be detected forged or not with an accuracy equal to 0.9. The accuracy for B frame is lower as higher noise is induced compared to I and P frames. $'-'$ means that the algorithms are not suitable for detecting a single B or P double compressed frame.

Table 4. Detection accuracy comparison for a GOP/frame

Scheme	GOP	B	P
Proposed	.98	.9	.94
Chen et.al. [1]	.96	-	-
Farid et.al. [4]	.99	-	-

Table 5. Detection accuracy comparison for a given ratio (q_1/q_2)

Ratio (q_1/q_2)	Proposed	Farid et.al. [4]
1.2 - 1.3	.8	.51
1.3 - 1.7	.87	.71
>1.7	.96	.99

Further, we compare our results against [4] for different ratios of single to double quantization for detecting if an I frame is forged or not. From Table 5 we can find that the proposed algorithm performs better than the algorithm in [4] for the ratios lower than 1.7. The improvement for the case when ratio is less than 1.3 is 60.8%, while 33.8% when ratio is between 1.3 and 1.7. However, for ratio greater than 1.7 it slightly decreases by 1%.

4. CONCLUSIONS

In this paper, we proposed a novel forgery detection algorithm based on pixel estimation and double compression statistics for videos captured from static cameras. An efficient estimator is derived to estimate the pixel values for a given frame. We also show the effect of the noise added due to single and double compression on the estimation of the pixels. Further, the estimated error is used to detect double quantization. However, the algorithm can detect the forgery in case of double compression bitrate higher than single compression bitrate only.

Experimental results show that the proposed algorithm provides high detection accuracies. In addition, the comparison results show that the proposed algorithm can also detect a double compressed B or P frame with the rest of the frames in the GOP being single compressed. Also the proposed algorithm gives better detection accuracies when compared to the existing algorithms. In future, we would like to work on localization of the tampered frames as well as small forged regions, which is one of the challenging and interesting problems in video forgery.

5. REFERENCES

[1] W. Chen and Y. Shi. Detection of double mpeg compression based on first digit statistics. *LNCS, Digital Watermarking*, 5450:16–30, 2009.

[2] C.C. Hsu, T.Y. Hung, C.W. Lin, and C.T. Hsu. Video forgery detection using correlation of noise residue. In *Proc. 10th IEEE Workshop on Multimedia Signal Processing*, pages 170–174, 2008.

[3] M. Kobayashi, T. Okabe, and Y. Sato. Detecting video forgeries based on noise characteristics. *LNCS, Advances in Image and Video Technology*, 5414:306–317, 2009.

[4] W. Wang and H. Farid. Exposing digital forgeries in video by detecting double quantization. In *Proc. 11th ACM workshop on Multimedia and Security, 2009*, pages 39–48.

[5] W. Wang and H. Farid. Exposing digital forgeries in video by detecting duplication. In *Proc. 9th ACM Workshop on Multimedia & Security*, pages 35–42, 2007.

[6] W. Wang and H. Farid. Exposing digital forgeries in video by detecting double mpeg compression. In *Proc. 8th ACM Workshop on Multimedia and Security*, pages 37–47. ACM, 2006.

[7] W. Wang and H. Farid. Exposing digital forgeries in interlaced and deinterlaced video. *IEEE Transactions on Information Forensics and Security*, 2(3):438–449, Sep. 2007.

[8] T. Van Lanh, K.S. Chong, S. Emmanuel, and M.S. Kankanhalli. A survey on digital camera image forensic methods. In *Proc. IEEE International Conference on Multimedia and Expo*, pages 16–19, 2007.

[9] B.G. Haskell, A. Puri, and A.N. Netravali. *Digital video: an introduction to MPEG-2*. Kluwer Academic, 1997.

[10] SM Key. Fundamentals of statistical signal processing: Estimation theory, 1993.

[11] http://trace.eas.asu.edu/yuv/.