# COMPRESSION NOISE BASED VIDEO FORGERY DETECTION

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

[†] Electronics and Communication Engineering, [‡] Computer Science and Engineering,
Indraprastha Institute of 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

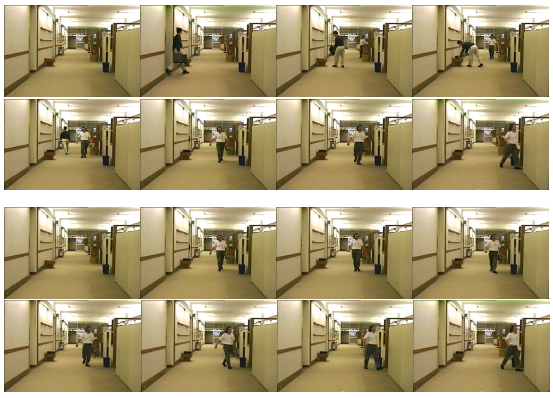

**Fig. 1**. Top two rows (in clockwise direction) : fames of an original video clearly showing two men on the hall. Bottom two rows: frames of the forged video where only one man is seen walking in the hall

## 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 recompressed, 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 MPEG videos [6], we model the extracted noise as a first

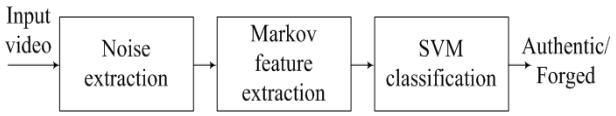


**Fig. 2**. Block diagram of authentication using the proposed scheme

order Markov process. Noise from each frame of a video is divided into $16 \times 16$ blocks and Transition probability matrix (TPM) for each block in 8 directions is obtained. The 8 TPMs are linearly combined to get a single TPM and the resulting 18-D feature is used for training and testing using Support Vector Machine (SVM). The detection unit is a single clip of a sequence of 10 clips.

The rest of the paper is organized in the following way. Section 2 gives the quantization process and related works while Section 3 gives the actual proposed scheme of forgery detection. In Section 4 the experimental setup, obtained results and comparison with other methods are given. Also, the advantage of this method over the others is discussed. Section 5 contains the conclusion and future work.

## 2. RELATED WORKS

Let $\mathbf{Z}$ be a frame of a video in spatial domain, $\mathbf{Y}$ (in frequency domain) be the transformed coefficients that we get after applying block based DCT matrix $\mathbf{H}$ for compressing frame $\mathbf{Z}$. Then,

$$\mathbf{Z} = \mathbf{HY} \Rightarrow \mathbf{Y} = \mathbf{H}^T\mathbf{Z} \tag{1}$$

If the quantization operator on the DCT coefficients is represented as $\mathbf{Q}[\cdot]$ then the quantized DCT coefficients are given by $\mathbf{Y}_q = Q[\mathbf{Y}]$. The quantized or compressed frame in spatial domain can be obtained by inverse-DCT of the quantized DCT coefficients as $\mathbf{Z}_q = \mathbf{H}^T\mathbf{Y}_q$ The quantization error in the spatial domain and frequency domain can generally be represented as,

$$\mathbf{e_Z} = \mathbf{Z} - \mathbf{Z_q} \text{ and } \mathbf{e_Y} = \mathbf{Y} - \mathbf{Y_q} \tag{2}$$

respectively. The 2D representation of the error in spatial domain is given as

$$\mathbf{e_Z} = \sum_{i=o}^{M-1}\sum_{j=o}^{N-1} \mathbf{H}_{i,j}(\mathbf{Y_q}[i,j] - \mathbf{Y}[i,j]) \tag{3}$$

The main parameter that is needed to model this error term or noise is the variances of individual frequency coefficients and the covariance matrix. Let the covariance matrix in frequency domain be represented as $\mathbf{K_{ey}}$, a diagonal matrix, whose diagonal elements are individual frequency domain coefficientsvariances given by $\sigma^2_{e_y(i,j)}$. The covariance matrix in the spatial domain will then be

$$\mathbf{K_{ez}} = E[(\mathbf{Z_q} - \mathbf{Z})(\mathbf{Z_q} - \mathbf{Z})^T] = \mathbf{H}^T\mathbf{K_{ey}}\mathbf{H} \tag{4}$$

This quantization error, and the variance-covariance matrix of the error are used to extract the noise from each frame.

## 3. PROPOSED SCHEME

### 3.1. Noise Extraction

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,

$$\hat{\mathbf{Z}} = \arg\max_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{Z_q}) \tag{5}$$

$$= \arg\max_{\mathbf{Z}} p(\mathbf{Z}) p(\mathbf{Z_q}|\mathbf{Z}) \tag{6}$$

where $\hat{\mathbf{Z}}$ 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 $\mathbf{Z_q} = \mathbf{Z} + \mathbf{e_Z}, \mathbf{Z}_q|\mathbf{Z}$ which is a Gaussian random variable with mean $\mathbf{Z}$ and auto covariance $\mathbf{K_{ez}}$. 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,

$$p(\mathbf{Z}) = \frac{1}{G} exp\big(-\lambda \sum_{c\in C} \rho_T(\mathbf{d}_c^t\mathbf{Z})\big) \tag{7}$$

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,

$$\rho_T(u) = \begin{cases} lu^2, & |u| <= T, \\ l(T^2 + 2T(|u| - T)), & |u| > T \end{cases} \tag{8}$$

where,

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

where, we introduce $l$ as weight in order to incorporate the effect of single compression on a frame. And $\mathcal{S}$ is the set of pixels which belong to the border pixels in each 8x8 block. $\mathbf{d}_c^t$ extracts the differences between a pixel and its neighbors so that,

$$p(\mathbf{Z}) = \frac{1}{G} exp\bigg(-\lambda \sum_{n=0}^{M-1} \sum_{m\in N_n} \rho_T(\mathbf{Z}[n] - \mathbf{Z}[m])\bigg) \tag{10}$$

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

$$\hat{\mathbf{Z}} = \Bigg\{ \frac{1}{G} exp\Bigg( -\lambda \sum_{n=0}^{M-1} \sum_{m \in N_n} \rho_T(\mathbf{Z}[n] - \mathbf{Z}[m]) \Bigg) \Bigg( \frac{1}{2\pi |\mathbf{K_{ez}}|^{1/2}} exp\Bigg( -\frac{1}{2} \mathbf{E_Z^T} {\mathbf{K_{ez}}}^{-1} \mathbf{E_Z} \Bigg) \Bigg) \Bigg\} \quad (11)$$

In order to maximize eq (6), the argument of $exp(.)$ in eq (11) should be minimized. This is performed using the method given in [11], and subsequently the noise is extracted. Let the resulting denoised frame be $\mathbf{Z_n}$ and the input frame be $\mathbf{Z}$, then the compression noise $\mathbf{C_n}$ present in the compressed frame is the difference between $\mathbf{Z_n}$ and $\mathbf{Z}$.

### 3.2. Markov Feature Extraction

The noise $\mathbf{C_n}$ can be modeled as a first order Markov Process such that, $Pr_{X_{t+1}} = Pr(X_{t+1}|X_t)$, where $X_{t+1}$ is the present state and $X_t$ is the previous state. The features that we use to represent this noise is the Transition Probability Matrix (TPM). $\mathbf{C_n}$ is divided into non overlapping blocks of $16 \times 16$ elements. Each block is used separately to extract TPM. The sign of each value in a block is obtained as,

$$\mathbf{C_n}(i,j) = \begin{cases} 0, \mathbf{C_n}(i,j) = Negative \\ 1, \mathbf{C_n}(i,j) = Zero \\ 2, \mathbf{C_n}(i,j) = Positive \end{cases} \quad (12)$$

This provides us with three different states to model a Markov chain. The transition probability between each of the three obtained states is calculated in each of the eight directions considering 8-connected neighbourhood. The probability along *right* direction for each element is obtained by the following condition,

$$P^{\rightarrow}_{u,v} = Pr(C_{n_{i,j+1}} = u | C_{n_{i,j}} = v) \quad (13)$$

where, $u, v \in [0, 2]$, and $u, v \in \mathbb{Z}$. Similarly, the probabilities can be obtained for other directions. The size of each TPM will be $3 \times 3$ since there are only three states and 9 possible transitions. Totally there will be 8 TPMs for each $16 \times 16$ block of a frame with $3 \times 3$ transition probabilities which is a large data.

In order to reduce the dimensionality of the obtained features, the TPMs along the top, bottom, left and right directions are averaged to get $\mathbf{F_1}$, as shown in eq (14). Similarly, the TPMs along all the diagonals are averaged to get $\mathbf{F_2}$, eq (15), resulting in only 2 TPMs per block per frame of a video. The two TPMs are concatenated to get the final feature which is 18-D, and for an $M \times N$ frame size, the dimension of the feature vector for the frame is $M/16 \times N/16 \times 18$. The direction of the arrows below show the direction along which the TPMs are calculated.

$$\mathbf{F_1} = \frac{1}{4}(\mathbf{F}^{\rightarrow} + \mathbf{F}^{\leftarrow} + \mathbf{F}^{\uparrow} + \mathbf{F}^{\downarrow}) \quad (14)$$

$$\mathbf{F_2} = \frac{1}{4}(\mathbf{F}^{\nearrow} + \mathbf{F}^{\searrow} + \mathbf{F}^{\swarrow} + \mathbf{F}^{\nwarrow}) \quad (15)$$

## 4. EXPERIMENTAL RESULTS

The video files are obtained from various open sources [13] in $4:2:0$ Common Intermediate Format (CIF) of resolution $352 \times 288$. Sixteen different sequences of 300 frames each are taken and are encoded using *'ffmpeg'* MPEG encoder. Details for MPEG-2 video detection are given here and that of MPEG-4 is discussed in Section 4.3. The encoding sequence is considered as "IPPPP" 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 $QF_1$ and double compression scale factor $QF_2$) resulting in 160*182 (total number of pairs) = 29120 clips. In Table1, 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* [14]. 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 $t_h$ = 0.5 ([6]), 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 $t_h$.

### 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

| $QF_1 \setminus QF_2$ | 2 | 3 | 4 | 6 | 9 | 10 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | x | 94 | 95 | 97 | 98.9 | 99.4 | 100 | 100 | 100 |
| 3 | 97 | x | 95.8 | 96.3 | 98.5 | 99.7 | 100 | 100 | 100 |
| 4 | 96 | 95.3 | x | 93 | 96.8 | 97.3 | 100 | 100 | 100 |
| 6 | 99.4 | 96.5 | 94 | x | 97.6 | 97.4 | 100 | 100 | 100 |
| 9 | 100 | 99.2 | 98.6 | 95.4 | x | 83.4 | 90.2 | 97.9 | 100 |
| 10 | 100 | 100 | 97.5 | 97.5 | 94.6 | x | 88 | 93.6 | 95 |
| 13 | 100 | 100 | 100 | 100 | 95.2 | 93.8 | x | 82.8 | 82.4 |
| 14 | 100 | 100 | 100 | 100 | 100 | 94.2 | 85.3 | x | 75.6 |
| 15 | 100 | 100 | 100 | 100 | 100 | 100 | 84.5 | 85.8 | x |

**Table 1**. Accuracy Rate for Various Compression Pairs

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.

In [2],[6],[7],[8], the authors point out that detection becomes harder when the double compression scale factor $QF_2$is a multiple of single compression scale factor $QF_1$. Proposed method is able to classify a given video sequence as authentic or forged irrespective of whether it was compressed with $QF_2$ that is an an *odd/even* multiple of $QF_1$. Comparison of classification accuracy between the proposed method for both MPEG-4 and MPEG-2 videos and the previous methods such as [2] for MPEG-2 and [6] for MPEG-4 is given in Table 2. It is clearly evident from Table 2, that the proposed scheme gives significant improvement in case of odd multiple case. In case of even multiple, the performance is better than that of [2, 6].

ROC curve for the proposed method of classification is given in Figure 3. Here, FPR is the ratio of classified authentic clips as forged to that of total number of forged. The plot shows that high TPR can be achieved while maintaining very low FPR. The better performance of our proposed scheme is because, the characteristics of the noise extracted from the single compressed frame differs from that of double compressed frame. This differences in spatial noise characteristics is effectively extracted by Markov modeling of the noise.

### 4.3. Discussion

The proposed scheme is also tested on YUV sequences when the encoding was done in the standard sequence i.e. *'IBBPBBPBBPBB'*. Also, apart from deleting a certain number of frames from a video sequence, forgeries such as 'copy paste', 'scaling', 'interchanging GOPs / certain frames randomly' is also considered. Since it is the double compression that is being detected, changes in the forgery type would theoretically give similar accuracy while detecting forgery.

In addition we tested our algorithm for videos that were compressed with MPEG-4 part-2 encoder and found that the detection accuracy is similar to that of Table 1. Further, as the proposed technique detects double compression based on compression noise, it can detect forgery in videos that were encoded using any of the MPEG coding techniques like MPEG-4 part-10.

| Method | Proposed (MPEG-2/4) | Markov for DCT coefficients [6] (MPEG-4) | First digit statistics [6, 2] (MPEG-2) |
|---|---|---|---|
| Odd Multiple | 98.98% | 51.53% | 50.32% |
| Even Multiple | 98.62% | 96.28% | 59.46% |

**Table 2**. Detection accuracy comparison

## 5. CONCLUSION

An efficient method to detect forgeries in video by detecting double compression is proposed. The effectiveness of this method is three fold. 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.

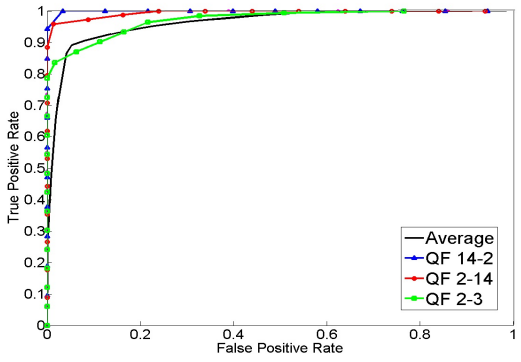


**Fig. 3**. ROC Curve showing True positive and False positive rate for three different scaling factors and the average.

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. REFERENCES


[1] http://mine.csie.ncu.edu.tw/core/en/group video, "Multimedia information networking laboratory," *online*.

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

[3] 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*, 2008, pp. 170–174.

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

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

[6] Xinghao Jiang, Wan Wang, Tanfeng Sun, and Yun Q. Shi, "Detection of double compression in mpeg-4 videos based on markov statistics," *IEEE Signal processing letters*, vol. 20, pp. 447–450, May 2013.

[7] T.-F. Sun, W.Wang, and X.-H. Jiang, "Exposing video forgeries by detecting mpeg double compression," in *Proc. IEEE International conference on Acoustic Speech Signal Processing*, 2012, pp. 1389–1392.

[8] Y.-Q. Shi B. Li and J.-W. Huang, "Detecting doubly compressed jpeg images by using mode based first digit features," in *Proc. IEEE International Workshop on Multimedia Signal Processing (MMSP)*, October 2008, pp. 730–735.

[9] Paolo Bestagini, Simone Milani, Marco Tagliasacchi, and Stefano Tubaro, "Local tampering detection in video sequences," in *Proc. IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*, 2013, pp. 488–493.

[10] AV Subramanyam and Sabu Emmanuel, "Video forgery detection using hog features and compression properties," in *Proc. IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, 2012, pp. 89–94.

[11] M.A Robertson and R.L. Stevenson, "Dct quantization in compressed images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 27–38, January 2005.

[12] C.-H. Chen, Y.-Q. Shi, and W. Su, "A machine learning based scheme for double jpeg compression detection," in *Proc. IEEE International Conference on Pattern Recognition*, 2008, pp. 1814–1817.

[13] www.xiph.org and www.trace.eas.asu.edu/yuv/, "video test media, derf's collection and yuv sequences," *online*.

[14] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.