Video Forensic Method Based on Temporal Noise Correlation

William D. Boxx II
525.759 Image Compression, Packet Video, and Video Processing
Department of Electrical and Computer Engineering
Johns Hopkins University, Whiting School of Engineering
3400 North Charles Street, Baltimore, MD 21218-2608
wboxx1@jhu.edu
10 December 2015

***Abstract*—In the arena of video evidence, authentication is key to determining use in the courtroom. This paper presents a method of using correlation values of temporal noise to determine tampering. A Gaussian mixture model (GMM) is proposed to model the distribution of a tampered video frame. Two classification methods are proposed for assigning clusters to the data based on this model. Finally two videos that demonstrate tampering due to in-painting techniques are used to test the proposed methods. Results are presented using three metrics: Detection Results (DR), False Alarm Rate (FAR) and False Rejection Rate (FRR). The results show promising accuracy but further research is needed in determining better classification methods.**

I. Introduction

In today's courtroom, the possibilities of digital media being introduced as forensic evidence in a trial are ever increasing. Before this media can be used as evidence, it must first be confirmed as authentic. This paper will develop the technique proposed in [1] to detect video tampering from temporal noise residue.

The method described in [1] has three main components: de-noise of frames, correlation of temporal noise blocks and classification. The de-noising algorithm used was first proposed in [2] and further developed in [3]. The proposed method uses wavelet decomposition to extract the noise from the original image. Once the noise as being isolated for each frame, correlation values are found between temporal frames. The correlation will be performed using a block based approach with block size of 8. Finally, a classification method is used to distinguish tampered and non-tampered blocks. An initial, less calculation intensive, classification is used to separate potential tampered frames from frames with non- potential of tampering. Next a fine classification is done. In the fine classification stage the correlation coefficients are first fit to a Gaussian Mixture Model (GMM). A GMM is useful for data sets with multiple clusters of highly correlated data. The correlation coefficients can be assumed to contain two Gaussian distributions if the frame has been tampered. Therefore a GMM with 2 clusters is used to fit the data. The Expectation Maximization (EM) algorithm is then used to find thresholds in order to distribute the data into the classifications. This method was proposed in [1]. It is in this way that the blocks in each frame can be designated as tampered or non-tampered blocks.

The rest of the paper is organized as follows. Section 2 will present the de-noising algorithm in greater detail. Section 3 will present the correlation method between adjacent frames. Section 4 will present the statistical models and classification methods used. In Section 5, the method will be tested on an experimental data set. Section 6 will conclude this paper.

II. De-noising Frames

The proposed detection method requires the extraction of noise from temporal frames in video sequence. The method used in this paper was proposed in [2] and further developed in [3]. The method has two main stages:

1. Estimate the local image variance
2. Use Weiner filter to obtain estimate of de-noised image in wavelet domain.

Color images are de-noised for each color channel separately. The high frequency wavelet coefficients will be modeled as a local stationary independent and identically distributed (i.i.d) signal with zero mean as well as zero mean Gaussian noise additive with variance, σ^2. The following step by step method is taken from [3].

Step 1) The first step calculates the fourth level wavelet decomposition of the noisy image using an 8 tap daubechies filter. The following steps are described for one high frequency level but are done for all four levels in practice. Denote the sub-bands as a(i,j), h(i,j), v(i,j) and d(i,j) where (i,j) runs through index set J that depends on the decomposition level.

Step 2) In each sub-band, estimate the local variance of original noise-free image for each wavelet coefficient using the MAP estimation for four sizes of a square W x W in the neighborhood N, for W ϵ {3, 5, 7, 9}.

$$\widehat{\sigma_W^2}(i,j) = \max(0, \frac{1}{W^2} \sum_{\{(i,j)\epsilon N\}} h^2(i,j) - \sigma_0^2), (i,j)\,\epsilon\, J$$

Take the minimum of the four variances as the final estimate.

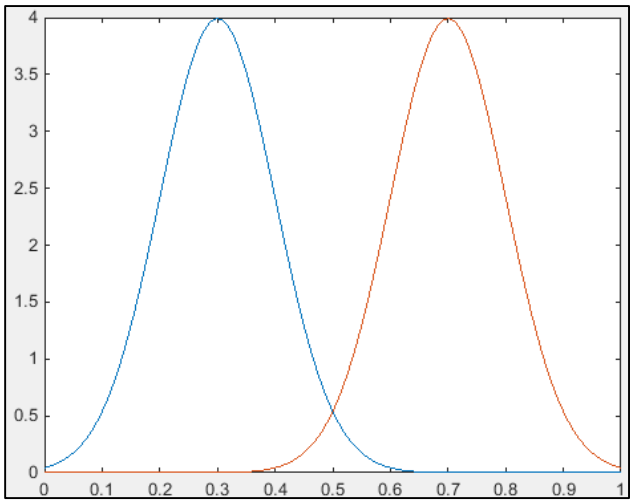

Figure 1 Example of 1 Dimensional GMM

$$\widehat{\sigma^2}(i,j) = \min\left(\sigma_3^2(i,j), \sigma_5^2(i,j), \sigma_7^2(i,j), \sigma_9^2(i,j)\right), (i,j)\epsilon$$

Step 3) The de-noised wavelet coefficients are obtained using the Weiner filter.

$$h_{den}(i,j) = h(i,j)\frac{\widehat{\sigma^2}(i,j)}{\widehat{\sigma^2}(i,j) + \sigma_0^2}$$

Step 4) Repeat steps 1-3 for each level and each color channel. The de-noised image is obtained by applying the inverse wavelet transform to the de-noised wavelet coefficients.

III. Correlation

Once the noise values are obtained for each frame. The next step is to find the correlation between adjacent frames. The idea behind this method is that the noise between adjacent frames will have high correlation. Therefore, if you have a frame that is tampered, you should have two different statistical sets in the correlation coefficients after correlation is applied. The correlation method used for this paper is the same as presented in [1].

Let n_{ij} denote the noise residual at pixel coordinate(i,j). The correlation value r between previous frame and current frame on each block can be defined as

$$r = \frac{\sum_i \sum_j (n_{i,j}^t - \overline{n^t})(n_{i,j}^{t-1} - \bar{n}^{t-1})}{\sqrt{(\sum_i \sum_j (n_{i,j}^t - \bar{n}^t)^2 \sum_i \sum_j (n_{i,j}^{t-1} - \bar{n}^{t-1})^2)}}$$

where t denotes the t-th frame and $\bar{t}^n$ is the mean value of the noise residual at t-th frame. The resulting correlation coefficients are used in the classification step.

IV. Classification methods

In [1], a passive forgery detection process is proposed. The proposed method uses the statistical properties of temporal noise residue, or noise between adjacent frames. The classification

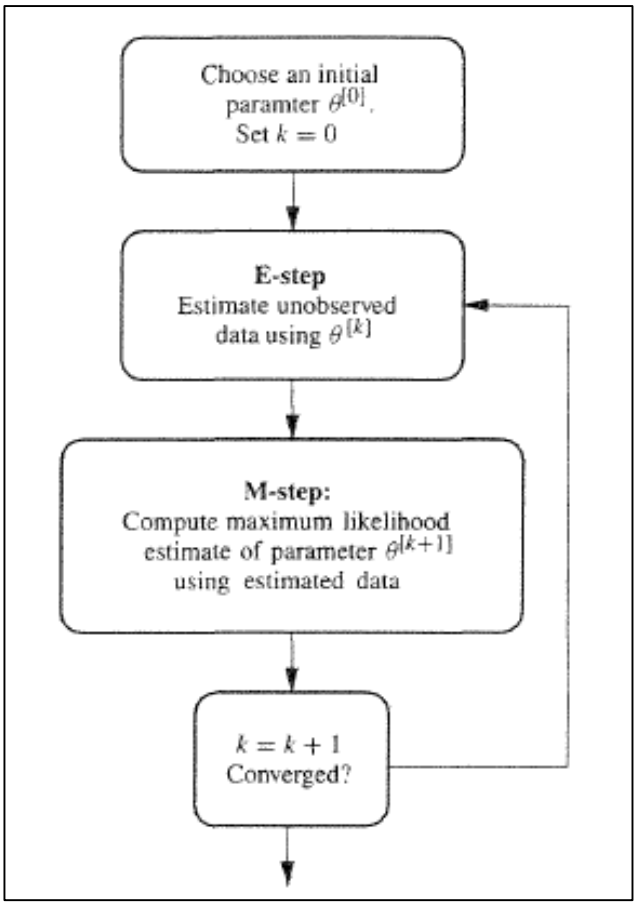

Figure 2 Block Diagram of the EM Algorithm

method is broken down into two steps: an initial classification and a fine classification. In a typical tampered frame the area of forged regions will be much smaller than the area of non-forged ones. Therefore an initial classification is needed to eliminate as any unnecessary calculations on frames with low probability of tampering. The initial classification is determined as follows:

$$Class_n = \begin{cases} 0 & |r_n - \bar{r}| < k\sigma_r \\ 1 & otherwise \end{cases}$$

Where $\bar{r}$ and σ_r are the mean and standard deviation values of the correlation coefficients and are found as follows:

$$\sigma_r^2 = \frac{1}{N}\sum_{\{n=1\}}^{N}(r_n - \bar{r})^2$$

$$\bar{r} = \sum_{n=1}^{N} r_n$$

Once an initial classification is complete the tampered frames are passed to a fine classification stage. A tampered frame will have statistical characteristics that can be modeled as two Gaussian distributions with separate means and variances.

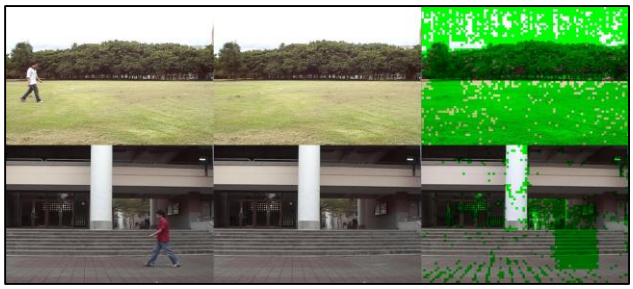

Figure 3 Test Results for Method 1

Therefore, a GMM with two clusters is used as a model to fit the data. A GMM is useful when modeling data that has multiple distribution clusters within the data set. Figure 1 shows an example of a 1 dimensional GMM such as we might see for a tampered frame. The correlation values for non-tampered blocks will have one mean, whereas the correlation of tampered blocks will have a different mean. The threshold value is where the two models interest and is used as a decision point to determine ownership. The threshold for the model is found using the EM algorithm as defined in [4]. The EM algorithm consists of two primary steps: an expectation step followed by a maximization step. The expectation is obtained with respect to unknown underlying variables, using the current estimate of the parameters and conditioned upon the observations. The maximization step then provides a new estimate of the parameters. These two steps are iterated until convergence. Figure 2 shows a block diagram of the EM algorithm. The general statement of the EM algorithm is as follows [4].

1. E Step
   Compute

$$Q\left(\theta\middle|\theta^{[k]}\right) = E[\log f(x|\theta)|y,\theta^{[k]}]$$

   Where $\theta^{[k]}$ is our estimate of the parameters at the kth iteration and θ is what we would like to find.
2. M Step
   Let $\theta^{[k+1]}$ be that value of θ that maximizes $Q(\theta|\theta^{[k]})$

$$\theta^{[k+1]} = argmaxQ(\theta|\theta^{[k]})$$

The code implementation of the EM algorithm for this paper uses the following MATLAB function from the statistics toolbox

GMModel = fitgmdist(X,k)

X is given as the correlation data for the frame and k is 2. This function fits a GMM to the data, X, using the number of clusters, k. It uses the EM Algorithm to converge on the mean values. The maximum allowable iterations is set to 1000. The function used to assign the posterior probabilities is the following

P = posterior(GMModel,X)

Figure 4 Test Results for Method 2

This function creates a k by length(X) array that holds probability values for each data point. A simple comparison between the two assigned probabilities for each data point is done to assign each data point to a model. The cluster set with the lowest mean is determined to be the untampered blocks.

V. Testing and Results

Two test videos were acquired from [5]. Each original video contains a person walking across the screen with a different background. The edited videos use in-painting techniques to remove the person from the frames. These edited videos were used as the test "evidence" and processed throughout the MATLAB code.

Three metrics were used in order to measure the performance of the detection method. The methods used were Detection Rate (DR), False Alarm Rate (FAR) and False Rejection Rate (FRR). DR, also known as sensitivity, was measured as

Detection_Rate = TP/(TP + TN).

TP are true positives and TN are true negatives.

The second metric used, FAR, tells us how many untampered blocks were falsely targeted. FAR was measured as

FAR = FP/(FP + FN).

FP are false positives and FN are false negatives.

The last metric used, FRR, tells us how many tampered blocks were undetected. This metric was calculated as

FRR = TN/(TN + TP)

Figure 3 shows the results for both videos when using the threshold determined by the EM algorithm. Table 1 shows the results of their DR, FAR and FRR metrics (method 1). Both videos had large false alarm rates due to poor threshold choices

TABLE I. METRIC RESULTS

	Vid 1 Method 1	Vid 1 Method 2	Vid 2 Method 1	Vid 2 Method 2
DR	80.21%	61.10%	91.48%	75.93%
FAR	10.61%	1.56%	79.42%	1.58%
FRR	19.52%	38.90%	8.52%	24.06%
Threshold	0.88	0.9621	0.79	0.9776

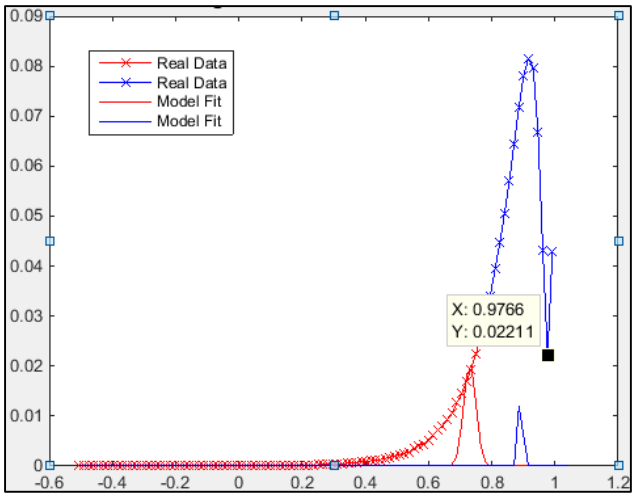

Figure 5 Histogram for Video Sequence 1

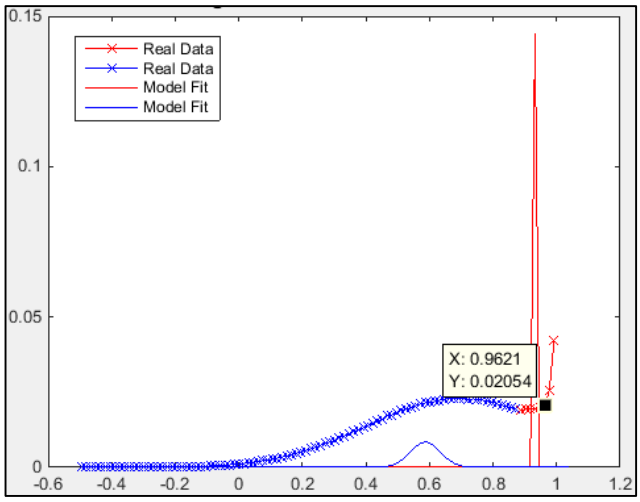

Figure 6 Histogram for Video Sequence 2

given from the EM algorithm. Figure 4 shows the results of both videos when thresholds were manually chosen by the user. In this method, the histogram for the entire video sequence was analyzed prior to classification and the user was allowed to manually enter a threshold level. The histograms for video sequence 1 and video sequence 2 are shown in Figures 5 and 6 respectively. Simple observation of the histograms tells the user that the EM algorithm choices are not perfect. Therefore user input is preferred for best results. The second part of Table 1 shows the metric results for the video sequences with user chosen thresholds. The false alarm rates were lowered considerably for each.

The results show that the EM algorithm alone is not suitable for classification. If no user interaction is to occur, a more suitable classification method must be used. It is however, rather simple to add a stopping point in the program that allows the user to do an analysis of the histogram and select a more appropriate threshold value. This need only be done once per video sequence

VI. Conclusion

In this paper, I have outlined a video forgery detection technique that uses the correlation between temporal noise residues and an EM classifier to extract tampered blocks. I presented test examples as well as metric for results. The EM method proved to be unsatisfactory as a sole determining factor for threshold values. An improvement in results was obtain through user determined threshold values. More research into classification methods may lead to better results in the future.

References

[1] C.-C. Hsu, T.-Y. Hung, C.-W. Lin and C.-T. Hsu, "Video Forgery Detection Using Correlation of Noise Residue," *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, pp. 170-174, 2008

[2] M. K. Mihcak, I. Kozintsev and K. Ramchandran, "Spatially Adaptive Statistical Modeling of Wavelet Image Coefficients and its Application to Denoising," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 6, pp. 3253-3256, Mar. 1999, Phoenix, AZ.

[3] J. Lukas, J. Fridrich and M. Goljan, "Digital Camera Identification From Sensor Pattern Noise," *IEEE Trans. Information Forensics Security*, vol. 1, pp. 205-214, June 2006.

[4] T. K. Moon and W. C. Stirling, Mathematical Methods and Algorithms for Signal Processing, Upper Saddle River, NJ: Prentice Hall, 1999.

[5] C.-C. Hsu, "Chih-Chung Hsu Downloads Page," 10 December 2015. [Online]. Available: https://sites.google.com/site/nthujesse/research/downloads