

# Video Inter-Frame Forgery Identification Based on Consistency of Correlation Coefficients of Gray Values

Qi Wang, Zhaohong Li, Zhenzhen Zhang, Qinglong Ma

School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China  
Email: [13120030@bjtu.edu.cn](mailto:13120030@bjtu.edu.cn), [zhhli2@bjtu.edu.cn](mailto:zhhli2@bjtu.edu.cn), [11111053@bjtu.edu.cn](mailto:11111053@bjtu.edu.cn), [qlma@bjtu.edu.cn](mailto:qlma@bjtu.edu.cn)

Received October 2013

---

## Abstract

Identifying inter-frame forgery is a hot topic in video forensics. In this paper, we propose a method based on the assumption that the correlation coefficients of gray values is consistent in an original video, while in forgeries the consistency will be destroyed. We first extract the consistency of correlation coefficients of gray values (CCCoGV for short) after normalization and quantization as distinguishing feature to identify inter-frame forgeries. Then we test the CCCoGV in a large database with the help of SVM (Support Vector Machine). Experimental results show that the proposed method is efficient in classifying original videos and forgeries. Furthermore, the proposed method performs also pretty well in classifying frame insertion and frame deletion forgeries.

## Keywords

Inter-Frame Forgeries; Content Consistency; Video Forensics

---

## 1. Introduction

Nowadays with the ongoing development of video editing techniques, it becomes increasingly easy to modify the digital videos. How to identify the authenticity of videos has become an important field in information security.

Video forensics aims to look for features that can distinguish video forgeries from original videos. Thus people can identify the authenticity of a given video. A kind of distinguishing method which is based on video content and composed of copy-move detection and inter-frame tampering detection becomes a hot topic in video forensics. Chih-Chung *et al.* [1] used correlation of noise residue to locate forged regions in a video. Leida *et al.* [2] presented a method to detect the removed object in video by employing the magnitude and orientation of motion vectors. A.V. *et al.* [3] detected the spatial and temporal copy-paste tampering based on Histogram of Oriented Gradients (HOG) feature matching and video compression properties. For inter-frame forgeries, such as frame insertion and frame deletion forgeries, some effective methods were proposed [4,5]. Juan *et al.* [4] presented a model based optical flow consistency to detect video forgery as inter-frame forgery will disturb the optical flow consistency. Tianqiang *et al.* [5] proposed a method using gray values to represent the video content

and detected the tampered video including frame deletion, frame insertion and frame substitution.

In this paper, we propose a simple yet efficient method based on consistency of correlation coefficients of gray values (CCCoGV for short in the following text) to distinguish inter-frame forgeries from original videos. The basic idea of our method is that differences of correlation coefficients of gray values of original sequences are consistent while differences of inter-frame forgeries will have abnormal points. We first calculate the differences of correlation coefficients of gray values between sequential frames of videos, and then use Support Vector Machine (SVM) [6] to classify original videos and inter-frame forgeries. We test our method on a large database, and results show that it is a method of high detection accuracy.

## 2. Review of Content Continuity

In a video which has not been tampered, the correlation of the content between the adjacent frames is high, while the content correlation between the far apart frames is relatively low. Gray value of frame can be a good representative of video content, which reflect the distribution and characteristics of the color and brightness level of an image. Based on this idea, a method was proposed to find out which frames has been tampered in a video by using gray value to depict the video content and the CCCoGV to describe whether video content is continuous or not [5].

The framework and details of the method [5] is as shown in **Figure 1**.

The main steps are as follows:

1) Decompose the video into orderly image sequence. Video is actually the orderly sequence of the 2-D image in the 1-D time domain. To extract the content features of a video, we can extract the features of the orderly image sequence.

2) Transform the RGB color image into gray image. The required storage space of a RGB color image is three times larger than that of a gray image. To reduce the amount of calculation, we use gray image to represent the content of the real color image. In the YUV model, the physical meaning of Y component is brightness of point. The brightness value represents the gray value of image. According to the relationship between RGB model and YUV model, brightness Y can be represented by the R\G\B components as

$$Y = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B. \quad (1)$$

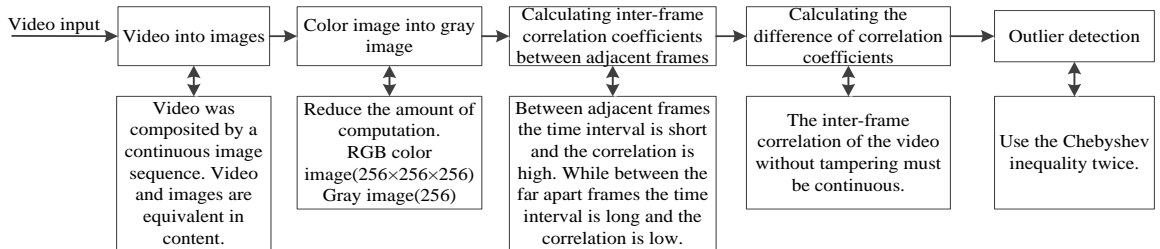
3) Calculate inter-frame correlation coefficient between adjacent frames. A gray image can be stored as a single matrix. Each element of the matrix corresponds to one image pixel and represents a different gray level. Equation (2) defines the correlation coefficient between two frames.

$$r_k = \frac{\sum_i \sum_j (F_{ra_k}(i, j) - \overline{F_{ra_k}})(F_{ra_{k+1}}(i, j) - \overline{F_{ra_{k+1}}})}{\sqrt{(\sum_i \sum_j (F_{ra_k}(i, j) - \overline{F_{ra_k}})^2)(\sum_i \sum_j (F_{ra_{k+1}}(i, j) - \overline{F_{ra_{k+1}}})^2)}}, \quad (2)$$

$$k = 1, 2, \dots, n-1$$

where,  $r_k$  is the correlation coefficient between frame  $k$  and frame  $k + 1$ .  $F_{ra_k}(i, j)$  represents the gray value of the pixel which is in the row  $i$  and column  $j$  of the frame  $k$ .  $\overline{F_{ra_k}}$  corresponds to the average value of the gray values  $F_{ra_k}(i, j)$ . Therefore, from a continuous frame sequence of which the length is  $n$ , we can obtain a correlation sequence of  $n - 1$ .

4) Calculate the difference between correlation coefficients of adjacent frames. Equation (3) defines the difference.



**Figure 1.** The framework and details of the method [5].

$$\Delta r_k = \begin{cases} |r_k - r_{k+1}|, & k \geq 2 \\ 0, & k = 1 \end{cases} \quad (3)$$

By the equation, we can obtain a sequence  $\Delta r_k = \{\Delta r_1, \Delta r_2, \Delta r_3, \dots, \Delta r_{n-4}, \Delta r_{n-3}, \Delta r_{n-2}, \Delta r_{n-1}\}$  to represent the variability of the content of a whole video. As we supposed, the variability of the video without tampering should be consistent, otherwise, the consistency will be destroyed by the deleted or inserted frames.

5) Finally, use the Chebyshev inequality twice to find out abnormal points which indicate where the tampered frames are.

Digital video forgeries detection based on content continuity is efficient. Based on this method, we can detect which frames were tampered in a video. But we cannot identify whether a video has been tampered by the method [5].

### 3. Proposed Method

#### 3.1. Analysis of CCCoGV for Original Videos and Inter-Frame Forged Videos

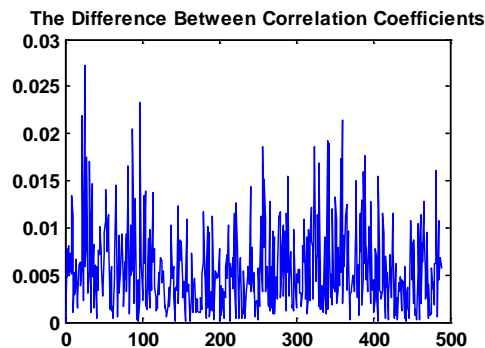
We proposed a method based on CCCoGV to classify original videos and inter-frame forgeries. Gray value is efficient in depicting video content. For a frame-tampered video, the gray values would be changed a lot at the tampered point.

In the original video database, every frame is non-tampered. The continuous frames in the original video are as shown in **Figure 2**. These three frames are continuous in a video. We can see the motion of the people is small and the background almost has no difference. In fact, a whole video is composed by hundreds of frames. The fluctuation of differences between correlation coefficients of adjacent frames is shown in **Figure 3**. From **Figure 3**, we can see that the differences of sequential frames are almost stable with values which are a bit larger than zero. That is to say, differences of correlation coefficients of original frames are consistent. The results are square with the consensus that for videos with fast motion, the correlation coefficient between continuous frames is low, but the change of correlation coefficients is stable.

Videos in the 25-frame-inserted video database and 100-frame-inserted video database are inserted 25 or 100 frames somewhere. **Figure 4** shows four frames gotten from a 100-frame-inserted video. The first and fourth images were the adjacent frames before frame insertion. We insert 100 frames between them then. And the inserted



**Figure 2.** Three continuous frames from an original video.



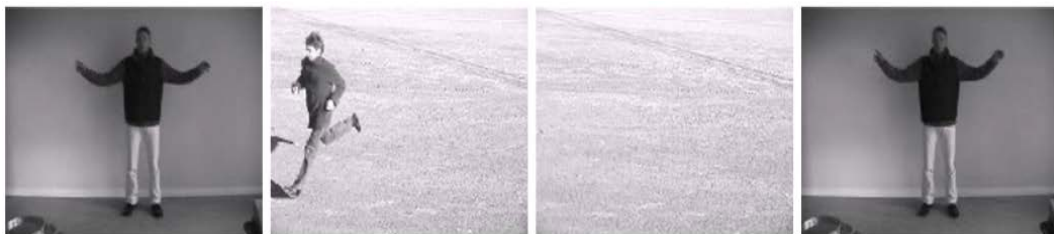
**Figure 3.** The CCCoGV of original video.

100 frames are continuous. The second and third images are the first and last frames of the video which was inserted to the original video. In a whole 100-frame-inserted video, the fluctuation of differences between correlation coefficients of adjacent frames is shown in **Figure 5**. From **Figure 5** we can see two pairs of peaks, which represent two tampered points. At every tampered point, a pair of peaks appears. The front pair of peaks denotes the start point of the insertion, while the latter pair of peaks represents the end point. The fluctuations in these two places are much larger than that of other points.

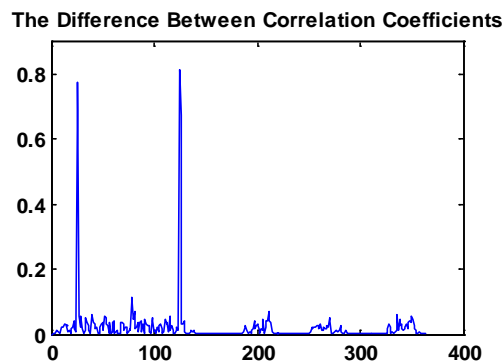
Similar to the frame-inserted video, in the 25-frame-deleted video database and 100-frame-deleted video database, every video is deleted 25 or 100 frames somewhere. **Figure 6** shows three frames gotten from a 100-frame-deleted video. The first two images are the frames before frame deletion and the third image is the frame after deleting 100 frames from the original video. In fact, we cannot find much difference between the last two images. In a whole 100-frame-deleted video, the fluctuation of differences between correlation coefficients of adjacent frames is shown in **Figure 7**. From **Figure 7** we can see one pair of peaks, which represent a tampered point. The pair of peaks denotes the deletion point, where the fluctuation of differences is much larger than that of other points.

### 3.2. Framework of Inter-Frame Forgery Identification Scheme

In this paper, a CCCoGV based inter-frame forgery identification scheme is proposed. The framework of our method is as shown in **Figure 8**.



**Figure 4.** Four continuous frames from a 100-frame-inserted video.



**Figure 5.** The CCCoGV of 100-frame-inserted video.



**Figure 6.** Three continuous frames from a 100-frame-deleted video.

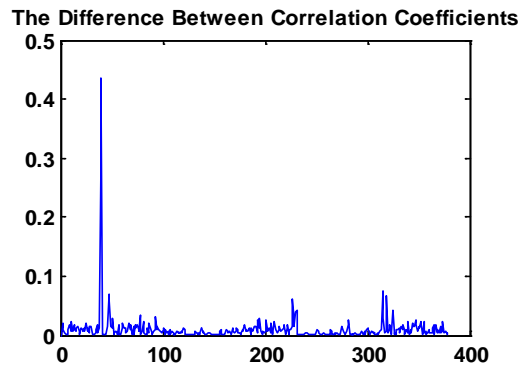


Figure 7. The CCCoGV of 100-frame-deleted video.

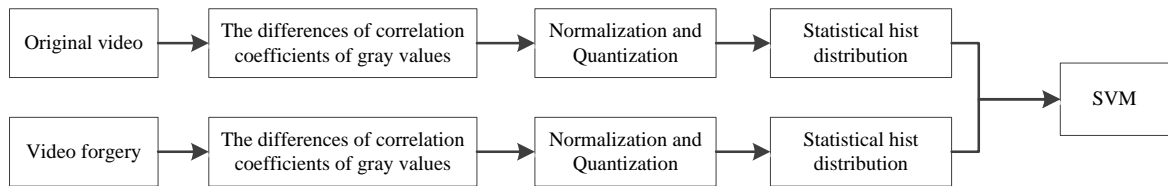


Figure 8. The framework of video inter-frame forgery identification.

The detailed scheme of our method is:

1) Calculate the differences between correlation coefficients of gray values of adjacent frames.

This step is same to the previous work [5]. Equation (1-3) is used to calculate gray value  $Y$ , correlation coefficients  $r_k$  and difference of correlation coefficients  $\Delta r_k$ . For a video with  $M$  frames, get a  $(M-1) \times 1$  vector to store the differences of the video.

2) Normalize and quantize the vector.

For each vector of the array, find the maximum  $m$  and divide every element in the matrix by  $m$ . Then quantify all the elements into  $D$  quantization levels. As all of them are distributed from 0 to 1, the quantization interval is  $1/D$ .

3) Count the statistical distribution of differences.

After the normalization and quantization step, the  $M-1$  elements are all discrete values from  $1/D$  to 1 ( $1/D, 2/D, \dots, 1$ ). Count the distribution of the discrete values into a  $D \times 1$  vector. For a video database with  $N$  videos, build an array contains  $N$  vectors of the statistical distribution to represent the CCCoGV features of all the videos in the database.

4) Distinguish original videos and video forgeries with Support Vector Machine (SVM) [6].

Put original video array and a video forgery array to the SVM, and distinguish them. Repeat 20 times and get the average classification accuracy.

The difference between our method and previous works [5] is the purpose. Their method is used for detecting where was tampered in a video. While our method is for the purpose to identify whether a video was tampered and classify original videos and video forgeries.

As their method cannot distinguish the original and forgeries, we improve their method by using normalization and quantization to obtain distinguishing feature of a fix length, and then use SVM to classify original videos and inter-frame forgeries.

## 4. Experimental Results

### 4.1. Experimental Object

In experiments, we use five video databases. One is original video database. The other four are tampered video databases, of which each separately contains the videos inserted 25 frames, inserted 100 frames, deleted 25 frames and deleted 100 frames. There are 598 videos with still background and a little camera shake in each video database ( $N = 598$ ).

**Table 1.** Classification accuracy.

Forgeries	Experiment Video Databases(With original video database)			
	<i>25-frame-insertion</i>	<i>100-frame-insertion</i>	<i>25-frame-deletion</i>	<i>100-frame-deletion</i>
Accuracy	99.22%	99.34%	94.19%	97.27%

**Table 2.** Classifying accuracy of two kinds of forgeries.

Forgeries	Experiment Video Databases	
	<i>25-frame-insertion and 25-frame-deletion</i>	<i>100-frame-insertion and 100-frame-deletion</i>
Accuracy	96.21%	95.83%

## 4.2. Experimental Setting

In our experiments, we use SVM of polynomial kernel [6] as the classifier.

To train the SVM classifier, about 4/5 of the videos are randomly selected as the training set (480 originals and 480 forgeries). The rest 1/5 form testing set (118 originals and 118 forgeries). The experiments are repeated for 20 times to secure reliable classification results.

## 4.3. Results and Discussion

We extract CCCoGV features of the five video databases, and classify original video database and each of the four tampered video databases separately. The accuracy was as shown in **Table 1**.

Obviously, this method is efficient to identify whether a video was tampered. As expected, the classifying accuracy between originals and frame-inserted forgeries is higher than that of frame-deleted forgeries. But even the classifying accuracy between originals and 25-frame-deleted forgeries is 94.19%, which is the lowest level of the four experiments.

We then try to mix the four kinds of forgeries and classify them from originals. The result is efficient, too. The accuracy is 96.75%.

Inspired by this, we further try to classify frame-inserted forgeries and frame-deleted forgeries. With the same method, we classify the 25-frame-inserted videos and 25-frame-deleted videos, as well as 100-frame-inserted videos and 100-frame-deleted videos. The result is as shown in **Table 2**.

We can see that the accuracy still stays high. The method is effective in classifying two kinds of forgeries. Then we mix the 25-frame-inserted videos and 100-frame-inserted videos, as well as 25-frame-deleted videos and 100-frame-deleted videos. And classify the two kinds of mixed forgeries. The accuracy is 98.79%.

To sum up, our method is efficient in classifying original videos and forgeries as well as different kinds of forgeries.

## 5. Conclusions

In this paper, we extract the consistency of correlation coefficients of gray values as distinguishing features to classify original videos and forgeries. The classifying accuracies are high in all of the experiments indicate that the proposed method is efficient in classifying original videos and forgeries yet different kinds of video forgeries.

In future work, we will find some new distinguishing features to identify the forgeries. And in view of the fact that the videos we use in the experiments are under still background, we will build a new video database with moved background.

## References

- [1] Hsu, C.C., Hung, T.Y., Lin, C.W. and Hsu, C.T. (2008) Video Forgery Detection Using Correlation of Noise Residue. 2008 *IEEE 10th Workshop on Multimedia Signal Processing*.
- [2] Li, L.D., Wang, X.W., Zhang, W., Yang, G.B. and Hu, G.Z. (2013) Detecting Removed Object from Video with Sta-

- tionary Background. *Digital Forensics and Watermarking. Lecture Notes in Computer Science*, **7809**, 242-252.
- [3] Subramanyam, A.V. and Emmanuel, S. (2012) Video Forgery Detection Using HOG Features and Compression Properties. *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, 89-94.
  - [4] Chao, J., Jiang, X.H. and Sun, T.F. (2013) A Novel Video Inter-Frame Forgery Model Detection Scheme Based on Optical Flow Consistency. *Digital Forensics and Watermarking*, Springer, Berlin, Heidelberg, 267-281.
  - [5] Huang, TQ., Chen, Z.W., Su, L.C., Zheng, Z. and Yuan, X.J. (2011) Digital Video Forgeries Detection Based on Content Continuity. *Journal of Nanjing University (Natural Science)*, **47**, 493-503.
  - [6] Chang, C.-C. and Lin, C.-J. (2011) LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, **27**, 1-27. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>