Digital Video Tampered Inter-frame Multi-scale Content Similarity Detection Method

Lan Wu(✉), Xiao-qiang Wu, Chunyou Zhang, and Hong-yan Shi

College of Mechanical Engineering, Inner Mongolia University for the Nationalities, Tongliao 028043, Inner Mongolia, China
wlimun@163.com

Abstract. With the popularity of the Internet and the increasing power of video editing software, digital video can easily be tampered with. The detection of the authenticity and integrity of digital video is very important. A video tampering detection method based on multi-scale normalized mutual information is proposed. Firstly, the mutual information is introduced into video tamper detection and the normalized mutual information content of the video frames is extracted. Then, based on the "scale invariance" feature of human vision, the mutual information between frames is analyzed from a multi-scale perspective. The multi-scale normalized mutual information is used to characterize the similarity of content between video frames. Finally, the LOF algorithm is used to calculate the degree of abnormality of the similarity coefficient sequence to achieve three kinds of tampering detection in the time domain: deletion, insertion, and replication. Experimental results show that the proposed method can effectively detect tampered video.

Keywords: Video tampering · Content continuity · Multi-scale · Content anomaly

1 Introduction

With the continuous maturity of digital multimedia technology and the increasing popularity of Internet technologies, the number of video files has increased dramatically. Video file access has become easier. The people have been able to enjoy the video feast brought by the Internet at any time. However, with the widespread using of various video editing soft wares such as Adobe photoshop and Adobe premier CSX, video files have been intentionally or unintentionally tampered with time [1, 2]. Once these tampered videos are used in official media, judicial proceedings, insurance, forensic evidence, etc., it is easy to cause misunderstandings or distort the facts of the truth and have a tremendous impact on others and the entire society. Especially at the current stage, the attention of the society to the security field has been increasing, and the authenticity and integrity of multimedia video content have also become the focus of public attention [3, 4]. Therefore, how to make scientific and reliable identification of the originality, authenticity, and integrity of digital video has become a branch of research in the field of modern multimedia information security.

G. Gui and L. Yun (Eds.): ADHIP 2019, LNICST 302, pp. 451–460, 2019.
https://doi.org/10.1007/978-3-030-36405-2_46

Compared with the tampering events of image and audio, the complexity of video tampering technology is relatively large. People have high recognition of the authenticity of video. Formally because of this, the video tamper is more harmful. At present, the world has made remarkable achievements in image modification and forensics research. However, digital video forensics technology is still in the preliminary research stage. There are few scientific and technological achievements that can truly detect various video materials [5]. At the present stage, most scientific research dissertations are inspired by digital image forgery and forensic detection techniques. They detect color changes by extracting color features, texture features, wavelet features, lighting features, and noise features [6–9]. In addition, due to the coding specificity of the video, extracting its GOP change characteristics or MPEG2 coding characteristics can also achieve video tamper detection.

Aiming at the problems of deletion, insertion, and replication in the time domain, a multi-scale normalized mutual information video tampering detection method is proposed. This method extracts the inter-frame mutual information amount of tampered video based on information theory, simulates the mutual information amount between video frames based on human visual multi-scale, and uses multi-scale mutual information to measure the similarity between video frames. Finally, the LOF algorithm is used to determine the degree of abnormality between tampered video frames. Experimental results show that this method can effectively detect three kinds of video tampering in time domain, such as deletion, insertion and replication.

2 Content Similarity Calculation

When humans identify an object, regardless of the object's distance, they can correctly determine the object's category. This is called "scale invariance." Therefore, by performing different scale analysis on the image, different details of the image can be obtained, thereby making the analysis of the image more accurate. Using a multi-scale normalized average mutual information model to measure the similarity between adjacent frames is closer to the perceptual characteristics of the human visual system.

2.1 Multi-scale Analysis

The spatial information of different scales of a two-dimensional image can be calculated by convolving the image with the Gaussian kernel, as follows:

$$L(x, y, \sigma) = I(x, y) \otimes G(x, y, \sigma) \tag{1}$$

where $I(x, y)$ is used to represent the gray value of the video frame image, $G(x, y, \sigma)$ is the Gaussian kernel function, and σ is the scale space factor, which is the variance of the normal distribution and reflects the degree to which the image is smoothed. Image Gaussian pyramid transform is used to build multi-scale spatial information of video frames.

After the video frame passes the Gaussian pyramid transform, the image resolution will decrease with the increase of the number of layers. According to the size of the initial video frame, the video frame is generally transformed by 3–4 layers.

$$G_k(x,y) = \sum_{m=-2}^{2} \sum_{n=-2}^{2} \bar{w}(m,n) G_{k-1}(2x+m, 2y+n) \tag{2}$$

where $1 \le k \le N, 0 \le x \le NR_k, 0 \le y \le NC_k$. NR_k and NC_k represent the number of rows and columns of the k-th Gaussian pyramid image, respectively. w is a 2D 5th order Gaussian window function

$$\bar{w} = \frac{1}{256} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} \tag{3}$$

2.2 Video Frame Mutual Information

Mutual information was first proposed by the American scientist Shannon, the father of information. Its purpose is mainly to measure the size of another random variable contained in the information of a random variable. Based on the technology of mutual information in image processing, we apply it to video frames. For a given video segment, each frame in the segment is viewed as a time sequence $\{F_1, F_2, \cdots, F_m\}$, and each video frame F_t is a grayscale image. The amount of information provided by video frame F_t can be measured by source information.

$$H(F_t) = -\sum_{i=0}^{L-1} p(l_i(F_t)) \log_2 p(l_i(F_t)) \tag{4}$$

where $p(l_i(F_t))$ represents the probability that the source F_t sends the symbol l_i, that is, the probability that the l_i gray level appears in the frame B. Therefore, by definition, $H(F_t)$ can be calculated using the grayscale histogram of video frame F_t.

Any two adjacent frames in a video sequence form a communication system. Then, the amount of information provided when two adjacent frames appear at the same time can be measured by the unity of the information theory.

$$H(F_t, F_{t+1}) = -\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p(l_i(F_t), l_i(F_{t+1})) \log_2 p(l_i(F_t), l_i(F_{t+1}))) \tag{5}$$

For the communication system with F_t as the source and F_{t+1} as the sink, the average amount of information the sink receives from the source can be measured by the average mutual information.

$$MI(F_t, F_{t+1}) = H(F_t) + H(F_{t+1}) - H(F_t, F_{t+1}) \tag{6}$$

The value of $MI(F_t, F_{t+1})$ can measure the visual content similarity between adjacent video frames $H(F_t)$ and $H(F_{t+1})$. It is called the single-scale content similarity operator.

2.3 Multi-scale Normalized Mutual Information

We combine the two methods of mutual information and multi-scale analysis between images, and introduce multi-scale normalized mutual information to measure visual content similarity between video frames. First, multi-scale analysis is performed on adjacent video frames to obtain Gaussian pyramid images for each layer. Then, the normalized average mutual information is calculated at each layer. Finally, the weighted sum of the normalized mutual information for each layer is normalized, and the result is a multi-scale normalized mutual information amount.

The formula for multi-scale normalized mutual information of adjacent video frames is as follows

$$\rho(t) = \sum_{k=0}^{n} w_k \cdot \frac{H(F_t(k)) + H(F_{t+1}(k))}{2H(F_t(k), F_{t+1}(k))} \tag{7}$$

where w_k represents the weight, $H(F_t(k))$ denotes the k-th layer Gaussian pyramid image in the t-th frame, $H(F_t(k), F_{t+1}(k))$ is the joint entropy of the kth Gaussian pyramid image between the t-th and $t + 1$-th video frames.

In this thesis, for a given video segment, it can be converted into a visual content similarity sequence $\{\rho(1), \rho(2), \cdots, \rho(M-1)\}$ by mutual information operator. The data sequence is one-dimensional data that describes the similarity of the content of the image sequence within the video segment.

2.4 Content Similarity Abnormality Measure

The degree of outliers of a data is not only related to its own data, but also related to the degree of outliers of the surrounding data. Therefore, the relative value of the average local density in the data point domain is used to describe the degree of data anomaly. For data point $\rho(i)$, the degree of abnormality is specifically defined as

$$L_{of}(i) = \frac{\bar{l}_{rd}(j)}{l_{rd}(j)} \tag{8}$$

where $\bar{l}_{rd}(j)$ represents the average of all points in the decentralized neighborhood with $\rho(i)$ as the center and $D_k(i)$ as the radius. $L_{of}(i)$ is called the abnormality of data B.

The value of the degree of abnormality $L_{of}(i)$ of data $\rho(i)$ measures the degree of anomaly of the data object $\rho(i)$ relative to the surrounding data points. The larger the value of C, the higher the degree of abnormality is. This shows that the visual content of the i-th frame differs greatly from the visual content of the video frames before and after it. Once the variability exceeds a pre-set threshold, it is reasonable to consider the data location as a tampered or stopped position of the video frame.

3 Algorithm Design and Implementation

Although the video encoding formats are different, they all have a high-speed frame rate, generally higher than 24 frames/second. Therefore, there is a high degree of similarity in visual content between video frame sequences. This paper measures the similarity between two adjacent frames by constructing multi-scale normalized mutual information descriptors between adjacent frames. The degree of abnormality of the similarity data sequence is established by means of the LOF algorithm. Once the video has been tampered with, the value of the sequence of the degree of similarity of the similarity sequence will change significantly. By setting a reasonable exposition to detect a relatively large coefficient of abnormality, the location of the video that has been tampered with is determined.

The basic steps of video tamper detection designed in this paper are described in detail as follows.

(1) Read the video segment to be detected, separate the video frames, audio and other elements in the video segment to be detected, and obtain the frame sequence.
(2) RGB color video frames are converted to grayscale frames.
(3) Calculate the Gaussian pyramid transform for each frame of image.
(4) Calculate the histogram of the Gauss pyramid image at each level and the joint histogram of adjacent frames.
(5) Calculate the multi-scale normalized mutual information operator of the t-th frame.
(6) Calculate the degree of abnormality sequence $\{L_{of}(1), L_{of}(2), \cdots, L_{of}(M-1)\}$ of the similarity sequence.
(7) Set the detection threshold β. For the i-th degree of abnormality $L_{of}(i)$, if $L_{of}(i) > \beta$, it is determined that the degree of similarity can be an outlier data point, thereby determining that the i-th frame is the starting point of the tampering position. Otherwise, it is considered that the i-th frame is not the starting point of the tampered position.

According to the basic steps of the algorithm, the three types of tamper detection flowcharts for deleting, inserting, and copying in the time domain are shown in Fig. 1.

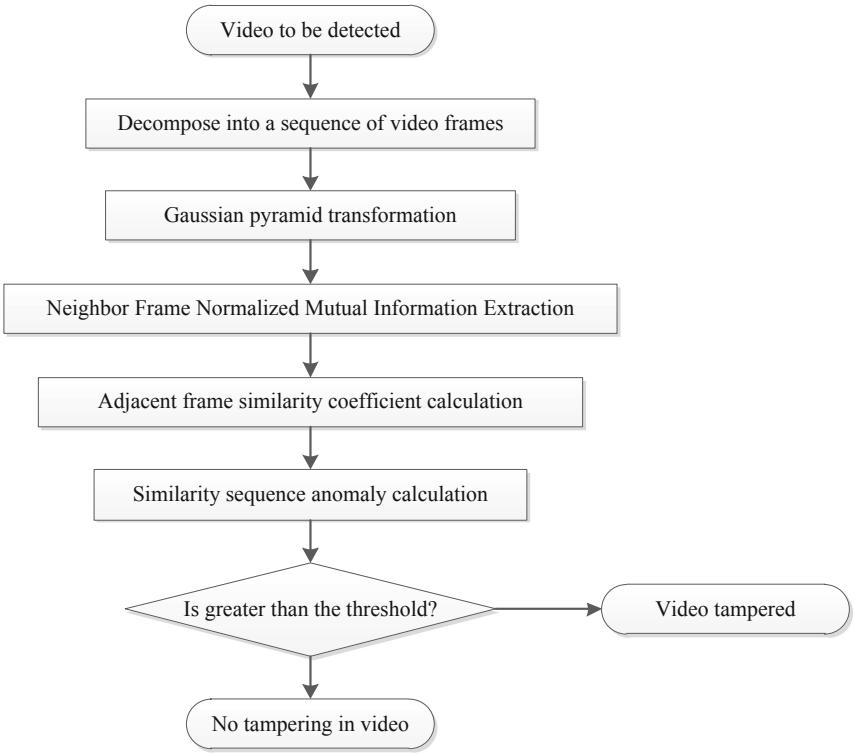

Fig. 1. Multi-scale normalized mutual information tamper detection process

4 Experimental Results and Analysis

4.1 Experimental Data

The existing video tamper detection technology does not have a unified video library. Various detection algorithms use self-captured video or some network resources for experimentation. Therefore, this article uses self-built test database. The video library downloads eight YUV format video clips from the literature [10]. The relevant parameters of the video are shown in Table 1.

Video library original video clips Video tamper is mainly to use some video editing software to modify the video content, time stamp, encoding format, etc., resulting in the destruction of the original video's authenticity and integrity. This paper mainly detects the time domain tampering in the video tampering method, including inter-frame deletion, frame replacement, and frame insertion.

Table 1. Video library original video clips

Video number	Name	Number of frames	Resolution
Video-1	hall_cif_yuv	300	352 × 288
Video-2	coastguard_cif_yuv	300	352 × 288
Video-3	coastguard_cif_yuv	300	352 × 288
Video-4	silent_cif_yuv	350	352 × 288
Video-5	pans_cif_yuv	150	352 × 288
Video-6	bus_cif_yuv	200	352 × 288
Video-7	flower_cif_yuv	300	352 × 288
Video-8	mobile_cif_vuv	300	352 × 288

4.2 Tamper Detection Results

The computer configuration used in the experiment was Intel 2.2 GHz CPU, 4G RAM, 750G hard disk, and Microsoft Windows 7 64 bit operating system.

For tampering with the clip Video-1 in the video library, the clip is obtained by deleting 80 to 123 video frames of Video-1. The calculated anomaly curve is shown in Fig. 2.

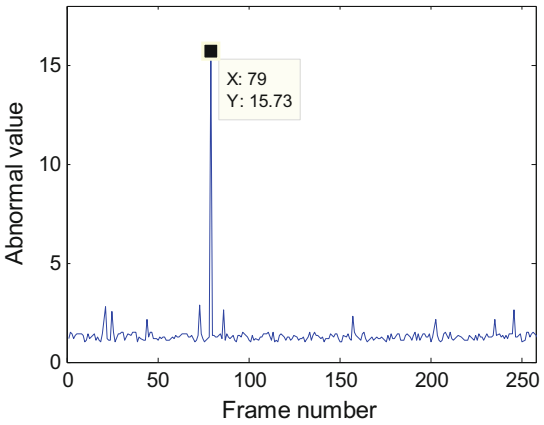

Fig. 2. The abnormal curve of the tampered video clip Video-1

It can be seen from Fig. 2 that the abnormal value of most frames is less than 3, while the abnormal value of the 79th frame far exceeds the abnormal values of other frames. Therefore, it can be considered that there is a tampering between the 79th and 80th frames. This position is consistent with the deletion of the 80th to 123th frames of the original video. Therefore, the 79th position detected by the algorithm is correct.

For tampering with the video-3 in the video library, the clip is formed by inserting a copy of the video frames 210 to 230 of Video-3 into frames 149 and 150. The calculated anomaly curve is shown in Fig. 3.

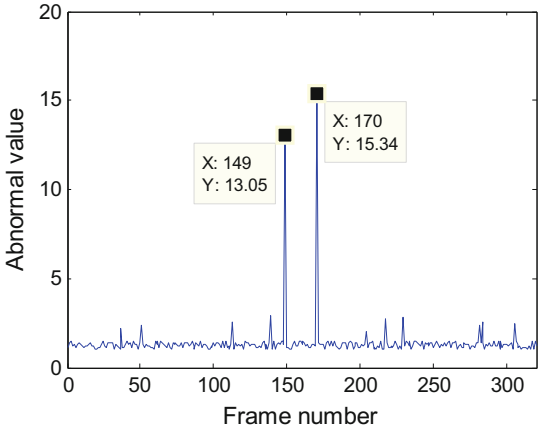

Fig. 3. The abnormal curve of the tampered video clip Video-3

It can be seen from Fig. 3 that there are two locations where the outlier exceeds the threshold 10, that is, 149 and 170, and the outliers at other locations do not exceed 3. This shows that there is a big difference in the visual content between the 149th and the 150th frames, and there is also a big difference in the visual content between the 170th and the 171th frames. Experimental results confirm that abnormality detection is correct.

4.3 Performance Comparison

In order to evaluate the performance of the algorithm, this paper uses the traditional detection accuracy rate R_p and the detection rate R_r. Two performance indicators are defined as follows

$$R_p = \frac{N_C}{N_C + N_F} \tag{9}$$

$$R_r = \frac{N_C}{N_C + N_M} \tag{10}$$

where N_C indicates the number of detected abnormal points, N_F indicates the number of abnormal points detected by mistake, and N_M indicates the number of abnormal points that are missed. In theory, the larger the R_p and R_r values corresponding to the tamper detection method, the better the detection performance is. Through statistics and calculations, the indicator values shown in Table 2 are obtained.

Table 2. Video tamper detection performance comparison

	Correct	Lost	Misdetection	R_p	R_r
Proposed method	28	1	2	90.32%	93.33%
Nonnegative tensor method [11]	19	7	5	73.07%	69.16%

The data in Table 2 shows that the tamper detection accuracy and detection overall rate using the method proposed in this paper are significantly higher than the Non-negative tensor method. The detection accuracy of this method is 90.32%, which is 17.25% higher than the Nonnegative tensor method. In terms of overall detection rate, this article reached 93.33%, an increase of 24.17% over the Nonnegative tensor method. Statistical indicators show that the proposed algorithm has better detection performance for video tampering.

5 Conclusion

This paper studies the detection of frame deletion, frame replacement and frame insertion in digital video. A digital video tamper detection method based on multi-scale content similarity between frames is proposed. This method can analyze the mutual information between adjacent frames in multiple scales from the perspective of human visual effects perception, and implement video tamper detection through multi-scale normalized mutual information and LOF algorithm. The experimental results verify the validity of the detection method. This method provides a new idea for digital video tamper detection.

Acknowledgements. Inner Mongolia National University Research Project (NMDYB1729).

References

1. Amanipour, V., Ghaemmaghami, S.: Video-tampering detection and content reconstruction via self-embedding. IEEE Trans. Instrum. Meas. **99**, 1–11 (2017)
2. Hu, W.C., Chen, W.H., Huang, D.Y., et al.: Effective image forgery detection of tampered foreground or background image based on image watermarking and alpha mattes. Multimed. Tools Appl. **75**(6), 3495–3516 (2016)
3. Wu, M.L., Fahn, C.S., Chen, Y.F.: Image-format-independent tampered image detection based on overlapping concurrent directional patterns and neural networks. Appl. Intell. **47** (2), 347–361 (2017)
4. Lin, J., Huang, T., Lai, Y., et al.: Detection of continuously and repeated copy-move forgery to single frame in videos by quantized DCT coefficients. J. Comput. Appl. (2016)
5. Fallahpour, M., Shirmohammadi, S., Semsarzadeh, M., et al.: Tampering detection in compressed digital video using watermarking. IEEE Trans. Instrum. Meas. **63**(5), 1057–1072 (2014)
6. Tang, Z., Wang, S., Zhang, X., et al.: Structural feature-based image hashing and similarity metric for tampering detection. Fundamenta Informaticae **106**(1), 75–91 (2011)
7. Huang, D.Y., Chen, C.H., Chen, T.Y., et al.: Rapid detection of camera tampering and abnormal disturbance for video surveillance system. J. Vis. Commun. Image Represent. **25** (8), 1865–1877 (2014)
8. Sitara, K., Mehtre, B.M.: Digital video tampering detection: An overview of passive techniques. Digit. Invest. **18**(8), 8–22 (2016)

9. Aghamaleki, J.A., Behrad, A.: Malicious inter-frame video tampering detection in MPEG videos using time and spatial domain analysis of quantization effects. Multimed. Tools Appl. **76**(20), 1–27 (2016)
10. http://trace.eas.ast.edu/yuv/index.html/.2013,7
11. Zhang, X., Huang, T., Lin, J., et al.: Video tamper detection method based on nonnegative tensor factorization. Chin. J. Netw. Inf. Secur. **3**(6), 1–8 (2017)