Inter-frame Tamper Forensic Algorithm Based on Structural Similarity Mean Value and Support Vector Machine

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

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

Abstract. With the development of network technology and multimedia technology, digital video is widely used in news, business, finance, and even appear in court as evidence. However, digital video editing software makes it easier to tamper with video. Digital video tamper detection has become a problem that video evidence must solve. Aiming at the common inter-frame tampering in video tampering, a tampered video detection method based on structural similarity mean value and support vector machine is proposed. First, the structural similarity mean value feature of the video to be detected is extracted, which has good classification characteristics for the original video and the tampered video. Then, the structural similarity mean value is input to the support vector machine, and the tampered video detection is implemented by using the good non-linear classification ability of the support vector machine. The comparison simulation results show that the detection performance of this method for tampered video is better than that based on optical flow characteristics.

Keywords: Video tampering · Inter-frame tampering · Structural similarity mean value · Support vector machine

1 Introduction

Digital information is flooding people's daily lives. Digital video as the mainstream of digital information is widely used in various fields such as news, justice, entertainment, military, and science. However, while people enjoy the enormous convenience of digital video, their negative effects gradually emerge. With the increasing versatility and ease of operation of video editing software, digital video is easily tampered with, resulting in the destruction of the integrity and authenticity of digital information. Since the tampering of the video after tampering is not easily perceived by the human eye, people cannot discern the authenticity of the video content [1, 2]. If lawless elements use falsified videos for gaining interests or as evidence in court, it will cause great confusion in society. Therefore, how to accurately judge whether a video has been tampered with has become an important topic in the field of information security.

At present, the main research methods for digital video forensics detection are divided into two categories: active forensics and passive forensics [3]. The active

S. Liu and G. Yang (Eds.): ADHIP 2018, LNICST 279, pp. 570–577, 2019.
https://doi.org/10.1007/978-3-030-19086-6_62

forensics technology refers to pre-embedded authentication information such as digital fingerprints or digital watermarks in the digital video to be forensic, and determines whether the video has been tampered by verifying whether the embedded authentication information is complete during the forensic process. Due to the need to embed verification information into the video in advance, the active forensics technology has great limitations. The passive forensics technology does not depend on external verification information [4]. It will leave tampering traces after the video content is tampered with, and will destroy the original statistical characteristics of the video content. It will use the statistical nature of the video content itself and verify the authenticity of the video. Passive forensics technology is more practical.

For the passive detection of video inter-frame falsification, scholars have proposed many methods. Stamm performs Fourier transform on the prediction error sequence based on the periodic peaks of the prediction error sequence, and detects the tampering video by searching for the peak value [5]. Dong uses the motion compensation edge effect to detect whether the video has been tampered with in the frame-delete mode [6]. Yuan uses the gray level co-occurrence matrix to extract the texture features of the video frames, and detects the heterogeneity frame insertion and frame replacement tampering according to the continuity of the feature [7]. Pandey proposes a passive forensic method for detecting tampered video using the characteristics of noise variation between the original frame and the tampered frame for the removal of dynamic objects and frame copying [8]. The method uses wavelet decomposition to extract the noise characteristics of the de-noised video frame, and then uses the Expectation Maximization algorithm to estimate the Gaussian Mixture Densitvl as the feature of the classification detection. Saxena uses optical flow inconsistency to detect and locate tampered video regions, but the method is not accurate [9]. Bagiwa proposed a new tampering detection algorithm for the falsification of video LOGO being removed [10]. The algorithm estimates the suspicious area by analyzing the spatial and temporal statistical characteristics of the LOGO area, and then uses the SVM to extract the features of the suspicious area and discriminate whether the suspicious area is a tampering area.

To solve the problem of digital inter-frame tamper detection, this paper proposes a video inter-frame falsification forensics algorithm based on structural similarity mean value. Structural similarity mean value is a measure of the similarity of two images, which is a combination of brightness, contrast, and structure in the image. It can more accurately express the similarity of the two images. First, the structural similarity mean value of the video to be detected is extracted, and then the tampered video detection is implemented using a support vector machine. The experimental results verify the validity of the detection method.

2 Passive Tamper Detection Fundamentals

Digital video has a great deal of relevance in the time and space domains. Spatio-temporal correlation detection of video can be used to detect whether it has been tampered with. The most important feature of digital video passive forensics is feature selection and extraction. Digital video information usually contains certain fixed

statistical characteristics due to the impact of video capture equipment and shooting scenes. By extracting and fusing these statistical features and analyzing the consistency between the features, video tamper detection can be implemented. Different researchers proposed different tamper detection features, including pattern noise, motion vectors, textures, and optical flow. Passive tamper detection uses multiple features to improve detection accuracy. Multiple features require fusion detection. Passive tamper detection uses the inherent nature of video for forensics and is universal for all types of video. Therefore, many research experts are dedicated to finding more effective features for passive tampering detection research.

Under normal circumstances, the basic flow of passive detection of video tampering is shown in Fig. 1.

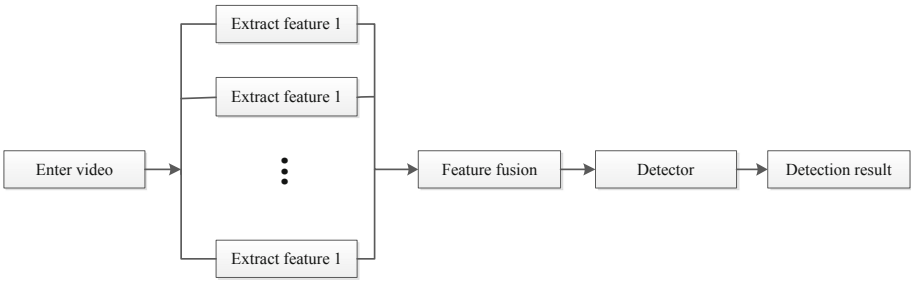

Fig. 1. Video passive forensics basic schematic

3 Structural Similarity Feature Extraction

Structural similarity is a new measure of the similarity of two images. Structural similarity theory holds that natural images are highly structured [11]. In other words, there is a strong correlation between adjacent pixels in the natural image, and this correlation carries important information of the object structure in the visual scene. The human visual system has been accustomed to extracting the structural information of the image from the visual field, and can use the measure of structural information as an approximation of the perceived quality of the image. Compared with the traditional methods of objective assessment of image quality, MSE and PSNR, structural similarity has been widely adopted because of its superiority in image similarity evaluation.

The structural similarity not only contains the brightness information of the image but also reflects the structure information of the object in the image. Therefore, it can more fully reflect the information in the image. For video inter-frame tamper detection, the use of features based on structural similarity results in better distinguishing characteristics. Firstly, a video is decomposed into a continuous image sequence, and then the images are divided into non-overlapping 8×8 sub-blocks, and the structural similarity values between the 8×8 sub-blocks corresponding to the adjacent two images A and B are calculated. Structural similarity consists of three parts: brightness, contrast and structure of similarity. Their definitions are as follows

$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \tag{1}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{2}$$

$$s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \tag{3}$$

where x and y represent the numbers of the corresponding 8×8 sub-blocks in images A and B, respectively, μ_x and μ_y represent the mean values of the corresponding 8×8 sub-blocks in images A and B, σ_x and σ_y represent the standard deviations of the corresponding sub-blocks, and $\sigma_x\sigma_y$ represents Correspond to the sub-block covariance. c_1, c_2, and c_3 are normal numbers that tend to 0, and are used to prevent the denominator of the three equations from showing 0.

Structural similarity is a combination of three different parts and is defined as

$$S_{SIM}(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \tag{4}$$

where α, β, and γ are used to adjust the relative weights of the three components. In general, set $\alpha = \beta = \gamma = 1$ and $c_3 = c_2/2$ to get a simplified version of the structural similarity

$$S_{SIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{5}$$

The mean value of structural similarity is calculated as

$$M_{SSIM}(A,B) = \frac{1}{M}\sum_{x=y=1}^{M} S_{SIM}(x,y) \tag{6}$$

where M is the total number of 8×8 sub-blocks in images A and B.

4 SVM Video Tamper Detection

Digital video consists of an ordered sequence of images in a one-dimensional time domain of two-dimensional images. Video can be first decomposed into a series of continuous images and then tampered with digital video. For a digital video, the content correlation between adjacent frames is high, and the content correlation between two frames that are far apart is smaller. Therefore, calculating the correlation between two adjacent frames in a video sequence of a video can describe the continuity of the content between the video frames.

If a digital video content changes quickly, the value of the correlation between adjacent frames is relatively small, that is, the structural similarity mean value is

smaller [12]. On the contrary, if the content changes slowly, the structural similarity mean value is relatively large. The structural similarity mean value between the video frames that have not been tampered with is not only high but close to the mean; however, the structural similarity mean value between the two frames at the tampered point in the tampered video is very low.

Support vector machine is a modern technology based on data machine learning. Support vector machine first maps linear inseparable data into a linear separable high-dimensional space. In the high-dimensional space, constructing the optimal classification surface based on the principle of minimizing structural risk is a method that can learn precision and learning based on finite samples. It is an intelligent learning method that seeks the best compromise between abilities.

According to the principle of minimizing the risk of high-dimensional space structure, the support vector machine attributes the detection problem to an optimization problem with constraints. The optimization function is

$$\min \frac{\|\omega\|^2}{2} \tag{7}$$

The constraint is

$$y_i\left[\omega^T\varphi(x_i)+b\right] \geq 1 \quad i=1,2,\cdots,K \tag{8}$$

where $\varphi(\cdot)$ is a kernel function mapped to a high-dimensional space, $\omega \in R_K$ is a weight vector, and $b \in R$ is an offset value.

The kernel function must be satisfied

$$\varphi(x_i)\cdot\varphi(x_j)=\kappa(x_i,x_j) \tag{9}$$

The kernel function satisfies the Mercer condition to meet the above requirements. To solve the optimization function, define the Lagrange function as

$$L(\omega,b,\alpha)=\frac{\|\omega\|^2}{2}-\sum_{i=1}^{K}\alpha_i(y_i\omega^T\varphi(x_i)+b-1) \tag{10}$$

where α_i is Lagrange multiplier. Calculate the partial derivative of A versus B and C respectively, and make the partial derivative equal to zero. Solve α_i values that satisfy the following conditions

$$\max H(\alpha)=\sum_{i=1}^{K}\alpha_i-\sum_{i,j}^{K}\alpha_i\alpha_j y_i y_j\varphi(x_i)\varphi(x_j) \tag{11}$$

There are the following constraints

$$\sum_{i=1}^{K}y_i\alpha_i=0 \quad \alpha_i\geq 0, i=1,2,\cdots,K \tag{12}$$

The optimal classification discriminant function is

$$y = \mathrm{sgn}\left\{\sum \alpha_i y_i \left[\omega^T \varphi(x_i) + b\right]\right\} \tag{13}$$

5 Results and Analysis

The experimental video library is divided into 5 sub-video libraries, including an original video library, a 25-frame deleted frames video library, a 25-frame inserted video library, a 100-frame deleted video library, and a 100-frame inserted video library. The video in the tampered four video banks is generated by inserting or deleting a certain number of video frames from the video in the original video library. In addition, the number of videos in each sub-video library is 598. The contents of the videos in the video library are the six kinds of human motion: wave, clapping, boxing, walking, jogging, and running.

In the experiment, polynomial kernel support vector machines were used to classify two types of video. In order to train the SVM classifier, 480 of the 598 videos were randomly selected as the training set, and the remaining videos were used as the test set. In order to ensure the reliability of the experimental results, the experiment was repeated 20 times and the average of the 20 experimental results was taken as the final classification accuracy. In the de-averaging process, k is 0.8 and the number of quantization bits is 30. All experimental videos have a resolution of 720 × 576. The frame rate of each video is 25 Fps.

Figure 2 is a structural similarity mean value curve of the original video and the tampered video after deletion of 25 frames. The experimental results show that the interval between the two curves is very large and the classification feature is very

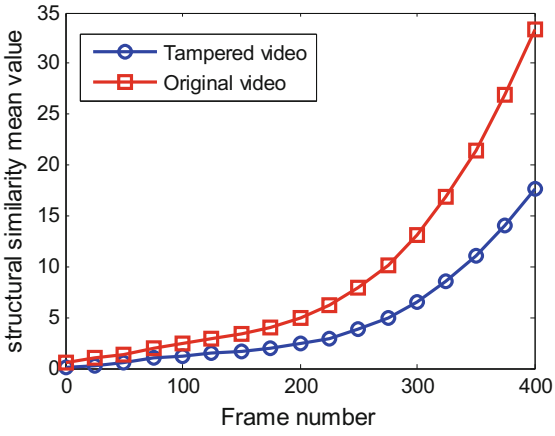

Fig. 2. Structural similarity mean value

obvious. This shows that the original video and tampered video's structural similarity mean value has a good separation, and its input to the support machine can achieve high probability detection of tampered video.

Table 1. Comparison of tampered video detection results

Tampered video	Proposed method	Optical flow
25-frame deleted library	91.45%	83.91%
25-frame inserted library	93.29%	89.34%
100-frame deleted library	97.25%	93.95%
100-frame inserted library	98.91%	94.28%

Table 1 is a comparison of the results of the proposed tampered video detection method and the optical flow tamper-based video method. Experimental results show that this method has high classification accuracy for original video and tampered video. Even with the 25-frame deletion video with the lowest classification accuracy, the accuracy rate reached 90.72%. In addition, as can be seen from Table 1, the detection accuracy of the inserted tampered video is higher than that of the deletion tampered video.

6 Conclusion

Video inter-frame falsification forensics algorithms are studied in this article. A tampered video detection method based on structural similarity mean value and support vector machine is proposed. The method utilizes structural similarity mean value difference of tampered video and original video to realize tampered video detection and uses support vector machine to improve detection performance. Experimental results show that the method can effectively detect tampered video and has high accuracy. However, the computational complexity of this detection method is still high, and the next step needs to solve this problem.

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

References

1. Arab, F., Abdullah, S.M., Hashim, S.Z., et al.: A robust video watermarking technique for the tamper detection of surveillance systems. Multimed. Tools Appl. **75**(18), 10855–10891 (2016)
2. Wei, W., Fan, X., Song, H., et al.: Video tamper detection based on multi-scale mutual information. Multimed. Tools Appl. **9**, 1–18 (2017)
3. Chen, X., Shi, D.: A new detection algorithm for video tamper. In: IEEE International Conference on Electronic Measurement & Instruments, pp. 1150–1153 (2016)

4. Martino, F.D., Sessa, S.: Fragile watermarking tamper detection via bilinear fuzzy relation equations. J. Ambient Intell. Humaniz. Comput. **5**, 1–21 (2018)
5. Stamm, M.C., Lin, W.S., Liu, K.J.R.: Temporal forensics and anti-forensics for motion compensated video. IEEE Trans. Inf. Forensics Secur. **7**(4), 1315–1329 (2012)
6. Dong, Q., Yang, G., Zhu, N.: A MCEA based passive forensics scheme for detecting frame-based video tampering. Digit. Investig. **9**(2), 151–159 (2012)
7. Yuan, X.J., Huang, T.Q., Chen, Z.W., et al.: Digital video forgeries detection based on textural features. Comput. Syst. Appl. **21**(6), 91–95 (2012)
8. Pandey, R.C., Singh, S.K., Shukla, K.K.: A passive forensic method for video: Exposing dynamic object removal and frame duplication in the digital video using sensor noise features. J. Intell. Fuzzy Syst. **32**(5), 3339–3353 (2017)
9. Saxena, S., Subramanyam, A.V., Ravi, H.: Video inpainting detection and localization using inconsistencies in optical flow. In: IEEE Region 10 Conference, pp. 1361–1365 (2017)
10. Bagiwa, M.A., Wahab, A.W.A., Idris, M.Y.I., et al.: Digital video inpainting detection using correlation of hessian matrix. Malays. J. Comput. Sci. **29**(3), 179–195 (2016)
11. 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)
12. Zhao, D.N., Wang, R.K., Lu, Z.M.: Inter-frame passive-blind forgery detection for video shot based on similarity analysis. Multimed. Tools Appl. 1–20 (2018)