Forensic Science International 236 (2014) 164–169

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Detection of object-based manipulation by the statistical features of object contour

Chen Richao, Yang Gaobo *, Zhu Ningbo

School of Information Science and Engineering, Hunan University, Changsha 410082, China

ARTICLE INFO

Article history:
Received 29 January 2012
Received in revised form 18 December 2013
Accepted 20 December 2013
Available online 7 January 2014

Keywords:
Video forensics
Passive forensics
Object-based forgery
Video in-painting
Object detection

ABSTRACT

Object-based manipulations, such as adding or removing objects for digital video, are usually malicious forgery operations. Compared with the conventional double MPEG compression or frame-based tampering, it makes more sense to detect these object-based manipulations because they might directly affect our understanding towards the video content. In this paper, a passive video forensics scheme is proposed for object-based forgery operations. After extracting the adjustable width areas around object boundary, several statistical features such as the moment features of detailed wavelet coefficients and the average gradient of each colour channel are obtained and input into support vector machine (SVM) as feature vectors for the classification of natural objects and forged ones. Experimental results on several videos sequence with static background show that the proposed approach can achieve an accuracy of correct detection from 70% to 95%.

1. Introduction

In the era of digital media, the proliferation of image and video editing tools makes the tampering or forgery of digital media much easier. Even ordinary users can produce forged digital media and spread them over Internet for malicious purposes. This leads to an increasing concern about the trustworthiness of public digital media [1]. To verify the authenticity, originality and integrity of digital media, digital media forensics arises to analyse, collect and preserve evidences from digital media. The existing techniques for digital media forensics can be divided into two categories: active and passive forensics [2]. Compared with active forensics, passive forensics does not need any data such as digital watermark or signatures. Thus, passive forensics is becoming a hot research topic in the field of information security.

Compared with digital image, the tampering of digital video is often more sophisticated and time-consuming. However, it is becoming easier with the popularity of video editing tools, such as Video Edit Magic. In the literature, there are many works about digital image forensics [3,4]. However, the research on digital video forensics is still in its infancy. The most representative works are summarized as follows: (1) forensics by the inconsistent trails during the imaging process such as PRNU [5], noise level functions [6]; and (2) forensics by the traces of video tampering, such as ghosting shadow [7], block artefacts [8], GOP periodicity [9] and motion compensated edge artefacts (MCEA) [10]. These methods are effective to detect traditional forgery operations, including copy–paste, double MPEG compression and frame-based tampering.

Object-based manipulations are usually malicious for digital video. For example, if an object is added into, or deleted from digital video, it might have direct influence on the content of digital video that it conveys [11]. Digital video is often believed to provide stronger forensic evidence than still images. As a consequence, the forensics of digital video is extremely important, especially when it used for legal evidence or news report. However, there is still few work reported in literatures about the passive forensics of object-based forgery in digital video to the best of our knowledge. In fact, object-based manipulations will inevitably leave some splicing traces [12], which are resulted from the limited accuracy of video object detection and extraction. Therefore, the statistical features within the boundary areas near video object will be inconsistent. This provides valuable clues for passive video forensic. In this paper, we are motivated to propose a passive video forensic method for object-based tampering. The statistical properties of video object and its variable-width boundary areas are fully utilized to determine the classification of natural objects and forged ones.

The rest paper is organized as follows. In Section 2, motion object is detected from static background-by-background subtraction technique, and then the object boundary is located. In Section 3, the statistical features of variable-width object area are

* Corresponding author at: School of Information Science and Engineering, Hunan University, Lushan South Road, Changsha 410082, Hunan, China.
Tel.: +86 0731 88821341.
E-mail addresses: yanggaobo@hnu.edu.cn, gbyang_hunu@hotmail.cm (Y. Gaobo).

0379-0738/$ – see front matter © 2014 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.forsciint.2013.12.022

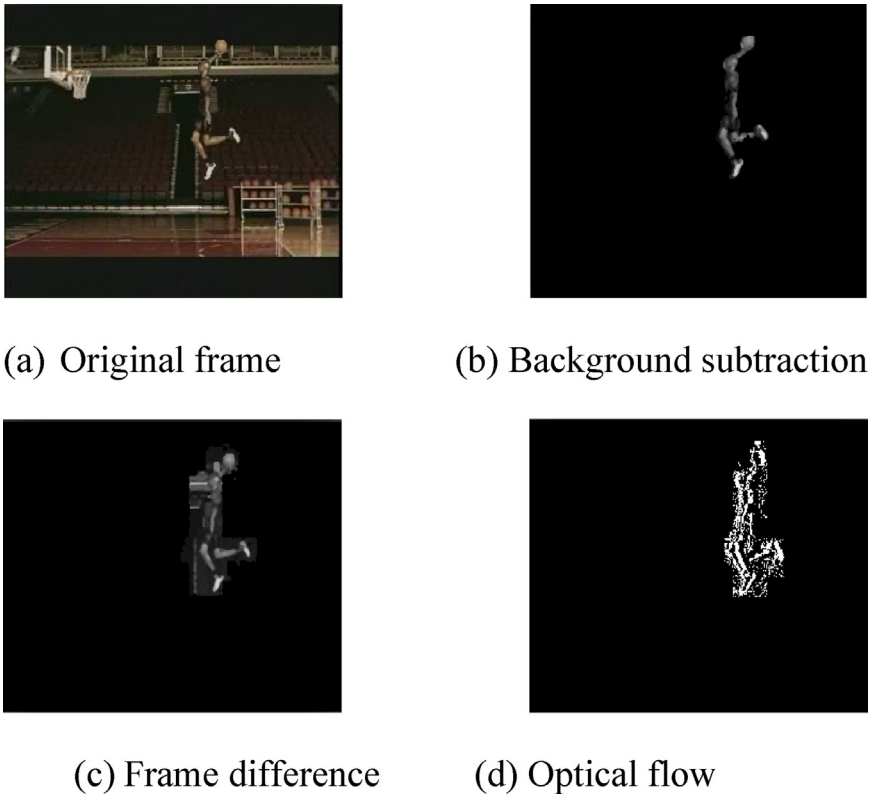

Fig. 1. Comparison of object detection methods: (a) original frame, (b) background subtraction, (c) frame difference, and (d) optical flow.

extracted and input into support vector machine (SVM) for pattern classification. Experimental results are reported and discussed in Section 4. We conclude the paper in Section 5.

2. Video object detection

Object-based video tampering refers to the generation of faked videos by adding, deleting or altering new video object. It usually consists of object detection/tracking, object manipulation, video in-painting and video layer fusion [13]. Therefore, object detection is the first step for digital video forensics to locate the object contour and its bounding areas. Then, the statistical features are extracted from the bounding areas around the object contour. With the help of pattern classifier, the originality and integrity of digital video is verified.

For motion object detection, the most conventional methods are optical flow, frame difference and background subtraction [14]. Optical flow method can obtain accurate diction results when tracking fast-moving object, but with intensive computation. Frame difference method is computationally efficient but very sensitive to scene change such as illumination. Therefore, it is relatively reasonable to choose background subtraction for motion object detection, especially for those video with static background. By establishing appropriate background model, the cumulative average of background frame can be obtained. Thus, motion object can be detected by making the difference between current frame and background frame. Apparently, the key issue for background subtraction technique is the background modelling and updating to adapt the external environment change. Among these background models, Gaussian Mixture Model (GMM) is most widely used [15]. It is a probabilistic approach that uses a mixture of normal distributions to model a multimodal background. For each pixel, each normal distribution in its background mixture corresponds to the probability of observing a particular intensity or colour in the pixel.

Fig. 1 shows the experimental results of *Jordan* sequence by the above-mentioned three object detection methods. Apparently, background subtraction method achieves the best object detection because the obtained object contour is more smooth and accurate. This will be beneficial to the successive statistical feature extraction from object contour and its bounding areas, and then the final classification result for passive forensics will be greatly improved.

After object-based tampering such as object removal, the structure in-painting, texture in-painting or combined structural and textural in-painting are usually performed to remove the motion artefacts. However, there are still some left traces for object-based video forgery, which always exist near the object boundary and its boundary areas. In our earlier work of object extraction, a new concept of adjustable width object boundary (AWOB) is introduced by mathematical morphology [16]. Let l be the extracted binary object, $\oplus$ be the dilation operation, and δ_s be the symmetric structure element, AWOB is defined as follows:

$$AWOB = \underbrace{\delta_s \oplus (\ldots \delta_s}_{n \text{ times}} \oplus (l)) \quad (1)$$

From Eq. (1), it is obvious that the object area gets larger with the increase of times n for dilation. In Fig. 2, an example is given for the AWOB generation of the detected object by background subtraction (in Fig. 1), where n equals 2.

3. The statistical features extraction

Because the gap between semantic object and low features used in object detection and extraction, there are always some irregularities near object boundary in the process of object-based tampering. Especially, when no dedicated video in-painting is performed after the object-based manipulation, there will be some subtle tampering artefacts near the object boundary and its

Fig. 2. AWOB: (a) original binary object, (b) AWOB, and (c) contour of AWOB.

bounding area. In the following, some statistical features are extracted to reflect and represent the forgery traces. In general, these statistical features should be stable, distinguishable and independent to each other.

Wavelet transform can provide us with the frequency of the signals and the time associated to those frequencies, making it very convenient for its application in numerous fields. The intrinsic properties of multi-scale analysis in wavelet transform make it useful in precisely locating the discontinuity (that is, high frequency details) of object contour in tampered video. In other words, the left tampering traces near the object boundary and its bounding area will be reflected in the sub-band coefficients. Especially, the sub-band coefficients within the AWOB will be bigger. Moreover, wavelet transform can be combined with moment invariants to further improve its robustness. The wavelet invariant moment can reflect the global and local information of image, and is also invariant to rotation, shift and scaling.

In this paper, the first and second order moments of wavelet detail sub-bands are utilized as the statistical features. Let $y = (y_1, \ldots, y_k)$ be the wavelet sub-band coefficients, where k is the number of coefficients, the definitions of absolute 1st moment and 2nd moment are as follows:

$$\text{Feature 1}: \quad m_1 = \frac{1}{k}\sum_{i=1}^{k}|y_i| \tag{2}$$

$$\text{Feature 2}: \quad m_2 = \sqrt{\frac{1}{k}\sum_{i=1}^{k}y_i{}^2} \tag{3}$$

Apparently, the first-order moment is actually the mathematical expectation of wavelet coefficients. It describes the centre of probability distribution. The second-order moment is in fact the variance, which reflects the dispersion degree of coefficients' distribution. In addition, several higher moments can be utilized to summarize the shape of distribution function and represent the regularities of distribution. Among them, the normalized third central moment, i.e., Skewness, describes the degree of symmetry in the variable distribution, whereas the fourth central moment, i.e., Kurtosis is a measure of the relative peakness or flatness of coefficient distribution [17]. Their definitions are as follows.

$$\text{Feature 3}: \quad m_3 = \frac{\mu_3}{\sigma^3} = \frac{(1/k)\sum_{i=1}^{k}(y_i-\bar{y})^3}{\left((1/k)\sum_{i=1}^{k}(y_i-\bar{y})^2\right)^{3/2}} \tag{4}$$

where $\bar{y}$ is the average of samples, μ_3 represents the 3rd central moment, σ^3 represents the variance.

$$\text{Feature 4}: \quad m_4 = \frac{\mu_4}{\sigma^4} - 3 = \frac{(1/k)\sum_{i=1}^{k}(y_i-\bar{y})^4}{\left((1/k)\sum_{i=1}^{k}(y_i-\bar{y})^2\right)^2} - 3 \tag{5}$$

where μ_4 is the 4th central moment.

To improve the representation of image detail, average gradient is introduced, which is the accumulation of local luminance contrast. In general, a bigger average gradient implies a stronger contrast of in local image details. Let *row* and *col* be the numbers of rows and columns in an image $f(x, y)$, respectively. The average gradient is defined as follows:(6)

$$\text{Feature 5}: \quad m_5 = \frac{1}{(row-1)(col-1)} \times \sum_{x=1}^{row-1}\sum_{y=1}^{col-1}\sqrt{\frac{(\partial f(x,y)^2/\partial x)+(\partial f(x,y)^2/\partial y)}{2}}$$

Moreover, edge intensity gradient can reflect the intensity of object edges. A small edge intensity gradient corresponds to a blurred edge appearance. For the R, G and B colour channels, their edge intensity gradients are computed, respectively, to avoid the loss of edge details. Then, the probability density function of Rayleigh distribution is utilized to model the intensity histogram of AWOB. Its probability function is defined as follows:

$$f(z) = \frac{z}{\sigma^2}\exp\left(-\frac{z^2}{2\sigma^2}\right), \quad z \geq 0 \tag{7}$$

For a sample frame, its edge intensity histograms of the R, G, and B channels are shown in Fig. 3.

Since Rayleigh distribution fits the edge intensity, its model parameters are estimated by MLE (maximum likelihood estimate) [17].

$$\text{Feature 6}: \quad m_6 = \sqrt{\frac{\sum_{i=1}^{k}y_i{}^2}{2k}} \tag{8}$$

Since different types of edge features have different representation capability, their influences on forensics are different. Though there are correlations among these features, they should be combined to improve the detection accuracy in passive forensics. Different weighting coefficients are assigned to these features, and then they are serialized to form a high dimensional vector. The weights are obtained by training with a gradual increase of step size from 0.5 to 0.9. Each group of weights is utilized for feature combination and pattern classification. The group with the highest classification accuracy is selected after all the training.

4. The classification based on SVM

Support vector machine (SVM) is a machine-learning algorithm based on statistical learning theory. It is a supervised learning model with associated learning algorithms that analyse data and recognize patterns for classification purpose. The basic idea behind SVM is to take a set of input data and predict, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. In this paper, SVM is chosen as the classifier for the authenticity of digital video.

Assuming that those videos with natural objects are positive samples, and video sequences after object-based tampering are negative samples, there will be four categories of possible

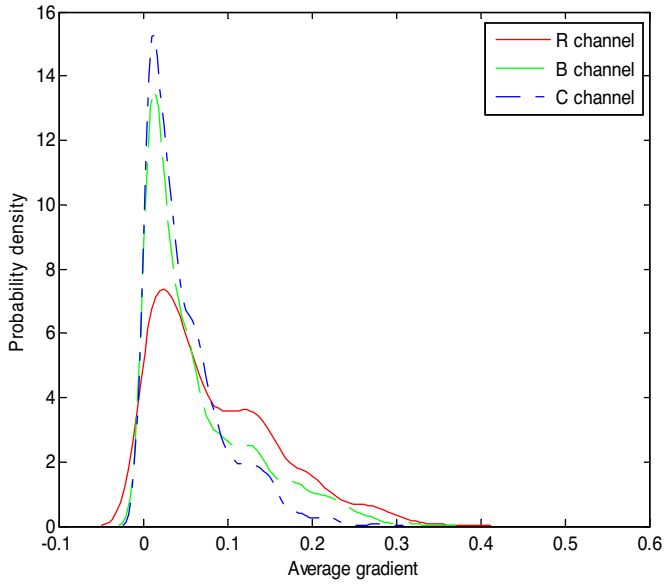

Fig. 3. Average gradient of RGB colour channel.

judgments for the binary (positive and negative) classification problem: (1) true positive (TP): positive samples are predicted as positive; (2) false negative (FN): positive samples are predicted as negative; (3) false positive (FP): negative samples are predicted as positive; (4) true negative (TN): negative samples are predicted as negative. Then, the following metrics, such as true positive rate (TPR), false positive rate (FPR), and accuracy are defined to evaluate the performance of classification.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{9}$$

$$TPR = \frac{TP}{TP + FP} \tag{10}$$

$$FPR = \frac{FP}{TN + FP} \tag{11}$$

The receiver operating characteristic (ROC), or simply ROC curve, is also known as a relative operating characteristic curve. In essence, ROC is a comparison of two operating characteristics TPR and FPR. Moreover, the area under curve (AUC) can also reflect how well a classifier works. The larger the AUC is, the better performance the classifier achieves. These metrics are used for performance evaluation of the proposed forensics approach.

5. Experimental results and discussions

To demonstrate the effectiveness of the proposed forensics approach, several videos are used for experiments. Unluckily, though there are several image libraries for forgery detection, no large-scale tampered video library is publically available [21]. To the best of our knowledge, the Surrey University Library for Forensic Analysis (SULFA) database is the only one available [18]. However, it is still at an early stage because the original videos are used to test algorithms for camera identification and device linking, whereas the forged videos for tampering detection are very limited. Therefore, we have tried our best to build a test video set. In total, there are 20 test videos for our experiments. The number of natural videos is 12, whereas the rest 8 are forged one with static background. Among the forged video, six are obtained from [19] without any processing, and the rest two are tampered video by us. The file formats of these test videos are AVI and WMV, respectively. The spatial resolution of video frame is 320×240. The experiment is performed with Matlab R2009a. The hardware configurations of PC are as follows: Inter(R) Core(TM) i3 CPU, 2.53 GHz, RAM 2 GB, Windows XP. For every video sequence, fifty video frames are selected, and the feature vectors are extracted and input SVM for classification. Before classification, every component in the feature vector is normalized as the following mapping. Thus, every component will play a balanced role:

$$f: \; x \to x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{12}$$

where x and x' represent the sample features before and after normalization, respectively. $x_{\min}$ and $x_{\max}$ are the minimum and maximum of x, respectively. That is, $x_{\min} = \min(x)$ and $x_{\max} = \max(x)$. Thus, the range of original feature vector x is normalized to [0, 1]. That is, $x' \in [0, 1]$, $i = 1, 2, \ldots, 6$.

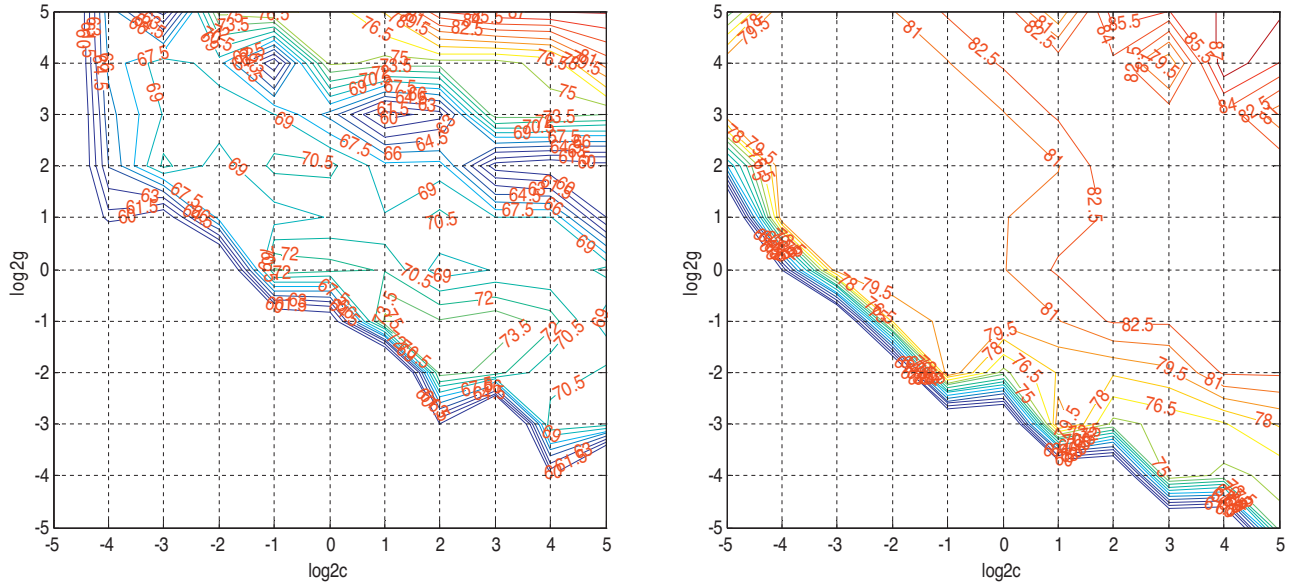

Fig. 4. Parameters optimization process.

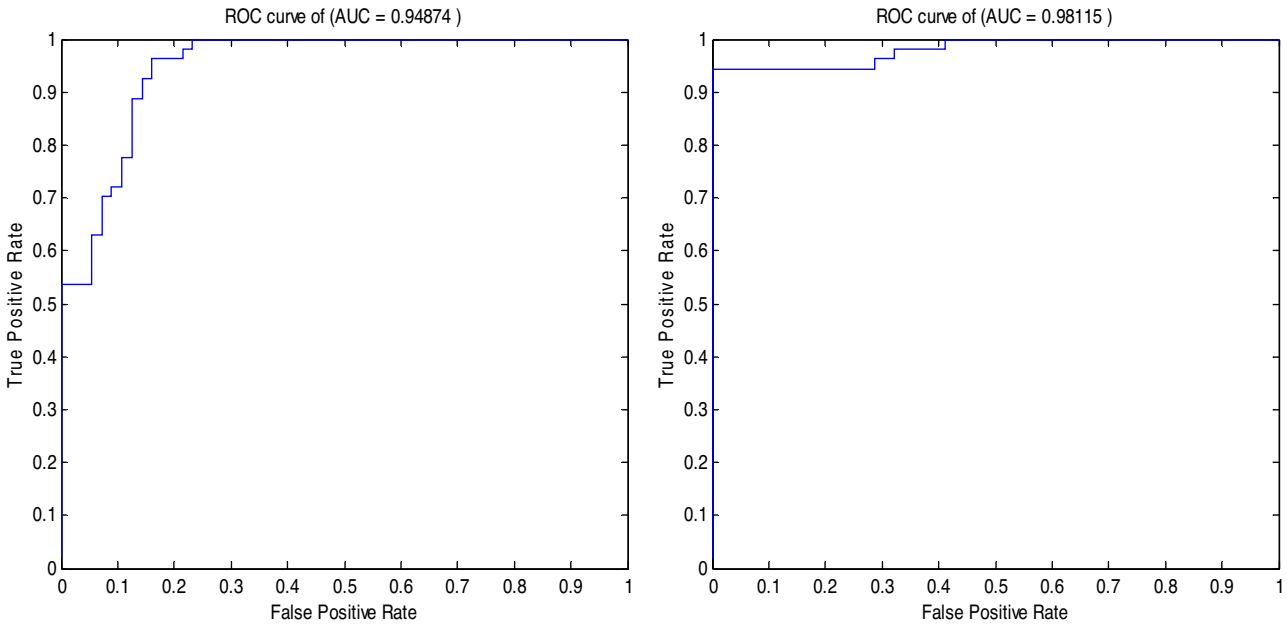

Fig. 5. The ROC curves under different circumstances.

The library LIBSVM is utilized to simplify the SVM implementation [20]. The radial basis function (RBF) is adopted as the kernel function. By cross validation based on leave-one-out, the penalty parameter c and RBF kernel function parameter g are obtained. To achieve better classification, the above processes are repeated for 10 times to obtain the average results. Fig. 4 shows the parameter optimization process. The x and y axes are the two parameters c and g of SVM, respectively. Lines of different colours represent the accuracy by contours when these two parameters are used in the process of cross validation. When $c = 8$ and $g = 32$, the accuracy is the highest (87.27%) in the left picture. When $c = 16$ and $g = 16$, an accuracy of 88.18% is achieved in the right picture.

The classification results are summarized in Table 1. It is apparent that different features have different effects on the classification of natural videos and forged ones, thought they all have accuracies above 70%. Moreover, the gradient features are relatively more effective than those moment features. When all the features are combined together, an accuracy of about 95% is achieved for classification.

Fig. 5 shows the ROC curves where the classification accuracies are 85.45% and 91.82%, respectively. Since its x axis is the FPR and the y axis is the TPR, ROC actually reflects the trade-off between true positive and false positive for the classifier. Moreover, the AUC are 0.948 and 0.981, respectively. These are satisfactory results for pattern classification.

Table 1
Classification results of extracted features.

Statistical features	TNR	TPR	Accuracy	AUC
The 1st moment (mean)	87.50%	87.04%	87.27%	0.8905
The 2nd moment (variance)	71.43%	68.52%	72.73%	0.7745
The 3rd moment (skewness)	67.64%	71.93%	68.18%	0.7523
The 4th moment (kurtosis)	78.57%	64.81%	73.64%	0.7576
R channel average gradient	92.29%	95.78%	94.55%	0.9901
R channel average gradient	92.86%	79.63%	90.91%	0.9894
R channel average gradient	94.64%	77.78%	85.45%	0.9487
R channel average gradient	99.32%	96.30%	99.09%	0.9947
R channel average gradient	98.87%	96.15%	90.63%	0.9812
R channel average gradient	83.43%	86.57%	88.18%	0.9547
The fused features	98.21%	94.44%	97.36%	0.9950

6. Conclusion

In this paper, a passive forensic method is proposed for object-based video forgery. The statistical features are extracted from AWOB for pattern classification. Specifically, the moment features of wavelet coefficients and gradient intensity of each colour channel are input into SVM for the classification of natural videos and forged videos. The experimental results show that for videos with static background, the proposed approach achieves desirable forensics results. However, due to the diversity and content complexity of digital videos, it still needs the support of sample database in the training process. Future investigation will be developing more robust features such as the motion trajectory. Moreover, a comprehensive dataset, which is similar to Columbia Image Splicing Detection Evaluation Dataset, is also needed for video forensics evaluation.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (Nos. 61072122 and 61379143), the Program for New Century Excellent Talents in University (No. NCET-11-0134), the Specialized Research Fund for the Doctoral Programme of Higher Education (No. 20120161110014) and the Key Project of Hunan Provincial Natural Science Foundation (No. 11JJ2053).

References

[1] A. Rocha, W. Scheirer, T. Boult, Vision of the unseen: current trends and challenges in digital image and video forensics, ACM Computing Surveys 43 (4) (2011) 26–40.
[2] S. Milani, M. Fontani, P. Bestagini, et al., An overview on video forensics, APSIPA Transactions on Signal and Information Processing 1 (e1) (2012) 1–18.
[3] J. Redi, W. Taktak, Digital image forensics: a booklet for beginners, Multimedia Tools and Applications 51 (2) (2011) 133–162.
[4] B. Mahdian, R. Nedbal, S. Saic, Blind verification of digital image originality: a statistical approach, IEEE Transactions on Information Forensics and Security 8 (9) (2013) 1531–1539.
[5] W. van Houten, Z. Geradts, Using sensor noise to identify low resolution compressed videos from YouTube, Digital Investigation 6 (1) (2009) 48–60.
[6] M. Kobayashi, T. Okabe, Y. Sato, Detecting forgery from static-scene video based on inconsistency in noise level functions, IEEE Transactions on Information Forensics and Security 5 (4) (2010) 883–892.
[7] Z. Jing, S. Yuting, Z. Mingyu, Exposing digital video forgery by ghost shadow artifact, in: ACM Workshop on Multimedia in Forensics, Beijing, China, (2009), pp. 49–54.
[8] W. Weihong, Digital video forensics, (Ph.D. thesis), Dartmouth College, 2009.

[9] D. Vazquez-Padın, M. Fontani, T. Bianchiy, et al., Detection of video double encoding with GOP size estimation, in: IEEE International Workshop on Information Forensics and Security (WIFS), 2012.
[10] D. Qiong, Y. Gaobo, Z. Ningbo, A MCEA based passive forensics scheme for detecting frame-based video tampering, Digital Investigation 9 (2) (2012) 151–159.
[11] T.K. Shih, N.C. Tang, J.C. Tsai, et al., Video motion interpolation for special effect applications, IEEE Transactions on Systems, Man, and Cybernetics (Part C) 41 (5) (2011) 720–732.
[12] R.W. Ciptasari, K.-H. Rhee, K. Sakurai, An image splicing detection based on interpolation analysis, Lecture Notes in Computer Science (PCM 2012) 7674 (2012) 390–401.
[13] T.K. Shih, N.C. Tang, W.-S. Yeh, T.-J. Chen, Video inpainting and implant via diversified temporal continuations, IEEE Transactions on Multimedia 13 (4) (2011) 602–614.
[14] R.J. Radke, S. Andra, O. Al-Kofahi, et al., Image change detection algorithms: a systematic survey, IEEE Transactions on Image Processing 14 (3) (2005) 294–307.
[15] J. Pierre-Marc, M. Max, K. Janusz, Statistical background subtraction methods using spatial cues, IEEE Transactions on Circuits and Systems for Video Technology 17 (12) (2007) 1758–1764.
[16] Y. Gaobo, Y. Shengfa, Modified intelligent scissors and adaptive frame skipping for semi-automatic video object segmentation, Real-Time Imaging 11 (4) (2005) 310–322.
[17] J. Weilan, W. Ronghua, X. Xiaoling, The statistical analysis of Rayleigh distribution under the full sample, Science Technology and Engineering 11 (11) (2011) 2551–2553.
[18] Surrey University Library for Forensic Analysis. http://sulfa.cs.surrey.ac.uk/.
[19] Multimedia Information Networking Laboratory of National Central University. http://mine.csie.ncu.edu.tw/core/en/group-video.
[20] LIBSVM – A Library for Support Vector Machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm/.
[21] Image Splicing Detection Evaluation Dataset. http://www.ee.columbia.edu/ln/dvmm/downloads/AuthSpliced-DataSet/dlform.html.