IMAGE TAMPERING DETECTION BASED ON STATIONARY DISTRIBUTION OF MARKOV CHAIN

Wei Wang, Jing Dong and Tieniu Tan

National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
P.O. Box 2728, Beijing, P.R. China, 100190
{wwang, jdong, tnt}@nlpr.ia.ac.cn

ABSTRACT

In this paper, we propose a passive image tampering detection method based on modeling edge information. We model the edge image of image chroma component as a finite-state Markov chain and extract low dimensional feature vector from its stationary distribution for tampering detection. The support vector machine (SVM) is utilized as classifier to evaluate the effectiveness of the proposed algorithm. The experimental results in a large scale of evaluation database illustrates that our proposed method is promising.

Index Terms— tampering detection, image chroma, Markov chain, stationary distribution

1. INTRODUCTION

Traditionally, a photograph implies the truth of what was happening. However, people in the digital world now sometimes can not trust image media since (maliciously) tampered images are often found in the Internet even published in newspapers. With the development of image editing software such as Adobe Photoshop, digital image can easily be manipulated and hardly detected by human eyes. If we take this issue for granted, it may eventually be harmful for our digital world, especially for the credibility of news coverage. Many researchers have worked on image forensics and a number of image tampering detection techniques have been proposed in recent years.

Generally speaking, there are two types of approaches of image tampering detection: active [1] and passive [2, 3] approaches. Active approaches often require pre-processing (for example, watermark embedding) to assist the authentication of digital images. However, active approaches are not desired for practical use in daily life since the image capture devices are not usually integrate with watermarking embedding module. Passive approaches, which gather evidence of tampering from images themselves, has more potential for practical use and gains more attention among researches in image forensics. In this paper we focus on passive image tampering detection based on supervised learning techniques.

There are several techniques for passive image tampering detection proposed in the recent literature [4–11]. In [5] and [6], Johnson and Farid developed a technique of tampering detection by analyzing the inconsistency of lighting in image. But it may fail when source images used for tampering are taken under similar lighting conditions. Popescu and Farid [7] argued that color interpolation (demosaicing) introduced specific correlations between neighboring pixels of a color image, while image tampering might destroy or alter them and based on this they proposed an image tampering detection algorithm to check the periodicity of these correlations. Besides, in [8], Dirik and Memon utilized artifacts created by Color Filter Array (CFA) to detect image tampering. They proposed two features. One is based on CFA pattern estimation and the other is based on the fact that sensor noise power in CFA interpolated pixels should be significantly lower than non-interpolated pixels. Actually, CFA artifacts are hardly detected for many images with heavy JPEG compression. Lukáš et al. [9] proposed a digital image tampering detection method to detect camera pattern noise which is considered as an unique stochastic characteristic of imaging sensors. The tampered region is determined when it is detected as lacking of the pattern noise. However, this method is only applicable when the tampered image is claimed to have been taken by a known camera or at least we have images taken by the camera before. Shi et al. [10] proposed a splicing detection method using effective features extracted from image Markov transfer matrices. Experiments were carried on Columbia image splicing detection evaluation dataset [12] and the results were satisfying. Aiming at color image splicing detection, we proposed an effective color image splicing detection approach based on image chroma [11]. We found that the analysis on chroma of color image was more reasonable for image splicing detection than on illuminance because chroma could reflect more information left by splicing which human eyes might not observe.

In this paper, we still apply our algorithm in image chroma channel for tampering detection. In our proposed approach, we model the edge information of image chroma as Markov chain (MC) and extract low dimensional feature vector from its stationary distribution. The dimension of feature vector of proposed method is much smaller than that we used in [11]. Similar to [10, 11], we train a SVM classifier with these features for tampering detection task.

The rest of this paper is organized as follows. Our proposed features for tampering detection are introduced in Section 2. In Section 3, the experimental results are reported and some analysis are given. Conclusions are drawn in Section 4.

2. FEATURE EXTRACTION

It is well known that tampering often changes the composition of an image. Even if the change cannot be perceived by human eyes, some inherent statistical dependencies (especially the higher order statistics) of image itself will be altered by tampering operation. In this section, we will introduce how to extract effective features for image tampering detection.

2.1. Thresholded Edge Image

Since image content is too strong to cover up tampering clues, in our approach we use edge information of image instead of image itself to extract features. We employ the mask M to convolute with an image chroma (Cb or Cr component) to get its edge information (see Equation 1) which we name as edge image in the rest of this paper. YCbCr is a family of color spaces just like RGB. Y is the luminance component and Cb and Cr are the blue-difference and red-difference chroma components. Cb or Cr component has little image content while most of image content is preserved in Y component. We find the image chroma is very useful for color image tampering detection. Since human are more sensitive to luminance than to chroma, even if tampered image looks natural to human, some unnatural clues will be left in chroma channel. Therefore, we could make use of the chroma information of the image for tampering detection.

$$E = |M \otimes I|, \tag{1}$$

where E is the edge image of an image chroma I and $|\bullet|$ is the operation of takeing absolute value. In our experiments, we set

$$M = \begin{bmatrix} -1 & -2 & -1 \\ -2 & 12 & -2 \\ -1 & -2 & -1 \end{bmatrix}.$$

Of course, other masks can be used to get edge images like Sobel, LoG etc.

We find that 90% of edge image's pixel values are below fifteen in our experiments as shown in Figure 1. Therefore, we can choose a reasonable value T to threshold edge image, and meanwhile, do not change the statistical regularity of it so much. Thresholding is according to the following rule:

$$e(i,j) = \begin{cases} e(i,j) & e(i,j) < T \\ T & e(i,j) \geq T \end{cases}, \tag{2}$$

where $e(i,j)$ is the value of an edge image at location (i,j).

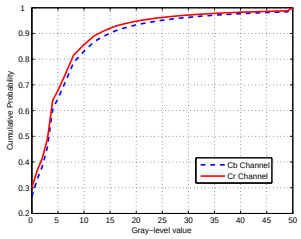


Fig. 1. The average gray-level value cumulative probability distributions of 10,246 edge images of Cb and Cr components of color images including authentic and tampered images, respectively.

2.2. Markov Chain

From above we know that the thresholded edge image's pixel values are integers from 0 to T. Hence, we can further model the edge image as a finite-state Markov chain (MC) to capture its interpixel dependencies. Markov chain is a well known statistical tool to model adjacent pixels' statistical dependency for steganalysis and splicing detection [10]; and transition probability matrix (TPM) can be used to characterize it. In this paper, we use a finite-state MC to model thresholded edge image, and further use one-step TPM to characterize it. The horizontal (0°) direction one-step TPM P^hcan be calculated in following way:

$$\begin{aligned} P^h_{m,n} &= P\{\nu(i,j+1) = n | \nu(i,j) = m\} \\ &= \frac{\sum_{i=1}^{H}\sum_{j=1}^{W-1}\delta_{m,n}(\nu(i,j),\nu(i,j+1))}{\sum_{i=1}^{H}\sum_{j=1}^{W-1}\delta_m(\nu(i,j))}, \end{aligned} \tag{3}$$

where $m,n \in \{0,1,\ldots,T\}$ are states of Markov chain. $\nu(i,j)$ is gray-level value of a $W \times H$ image at location (i,j) and

$$\delta_{m,n}(A,B) = \begin{cases} 1 & \text{if } A = m \text{ and } B = n \\ 0 & \text{Otherwise} \end{cases}.$$

In the same way, we can get other directions $(45^\circ, 90^\circ, 135^\circ)$ one-step TPMs which are named P^d, P^v and P^{-d}, respectively.

For a finite-state MC, there always exist a stationary distribution π which is a vector with entries sum up to 1 and satisfies the equation $\pi = \pi P$, where P is a TPM. We use the stationary distribution as feature vector, thus the dimension of the feature vector is $T+1$.

As we know, the stationary distribution of finite-state MC is always exists, but may not be unique. However, if all elements of its TPM are positive, which means the chain is irreducible and ergodic, the stationary distribution is unique. This can be proved by the following theorem [13]:

Theorem 1 *For an irreducible ergodic Markov chain,* $\lim_{n\to\infty} P^n_{ij}$ *exists and is independent of i. Furthermore, letting*

$$\pi_j = \lim_{n\to\infty} P^n_{ij}, \quad j \geq 0,$$

then π_j *is the unique nonnegative solution of*

$$\begin{cases} \pi_j = \sum_j \pi_i P_{ij}, & j \geq 0, \\ \sum_j \pi_j = 1. \end{cases}$$

π_j *is called the limiting probability that the chain is in state* j.

For the irreducible ergodic Markov chain, its limiting probability distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_j, \ldots)$ is exactly its stationary distribution. Thus, its stationary distribution is unique.

In our experiments, we found that almost all TPMs have their elements positive and satisfy Theorem 1. However, we can not guarantee other cases outside our experiments. In order to make our proposed method more general, we added a positive perturbation to TPM and then scaled each row of the matrix to make its L1 norm be one. Hence, we can make sure the stationary distribution of our modeled Markov Chain of the thresholded edge image to be unique.

2.3. Frame Work of Proposed Method

In [11], we find that the chroma channel of a color image is more suitable for image splicing detection than illuminance channel. Inspired by this observation, we again applied our proposed method in chroma channel for tampering detection. The frame work of our proposed method is shown in Figure 2. A RGB color image is first converted to YCbCr color space and chroma component (Cb or Cr channel) is used. Then, we calculate the stationary distribution of the thresholded edge image of chroma to server as a feature vector. The dimension of the feature vector is $T+1$ where T is the predefined value used for thresholding edge image. In Section 3, effectiveness of proposed low dimension feature vector for tampering detection will be testified.

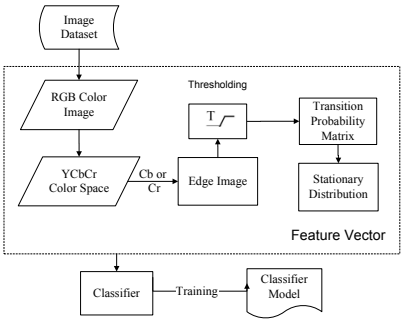


Fig. 2. A frame work of our proposed method

Fig. 3. Examples of authentic images (all in the top row) and their forgery counter parts (all in the bottom row)

3. EXPERIMENTAL RESULTS

3.1. Image Database

The only two available public image databases for tampering detection, specially for splicing detection, are provided by DVMM, Columbia University [12]. They are simple and small scale. In order to provide a more realistic and more challenging evaluation database for image tampering detection, we constructed a color image database to test our proposed approach. The CASIA tampered image detection evaluation database (CASIA TIDE) v2.0 [14] consists of 7,491 authentic and 5,123 sophisticatedly tampered color images of different sizes varying from 240×160 to 900×600. This database is with larger size and more realistic and challenging tampered images with complex splicing as well as blurring. The authentic images are collected from the *Corel* image dataset, websites and our own images captured from digital cameras. The tampered images are generated by the following ways:

- we randomly cut-and-paste image region(s);
- the cut image region(s) can be processed with resizing, rotation or other distortion and then be pasted to generate a spliced image;
- the post-processing (such as blurring) is considered after cut-and-past operation to finish the tampered image generation;
- difference sizes (small, medium and large) of tampered regions are concerned.

Some examples of CASAI TIDE v2.0 are shown in Figure 3.

3.2. Classifier

Support Vector Machine (SVM) is an optimal and efficient classifier which is commonly used for machine learning systems. Since our work in this paper only focuses on feature extraction rather than the design of classifier, we utilize the LIBSVM [15] as the classifier in our experiment and a RBF kernel is chosen. Five-fold grid searching is used to select parameters for the classifier.

3.3. Detection Performance

In our experiment, we set $T = 8$ and 15 to threshold edge images, respectively. Instead of using four directions TPMs of thresholded edge image, i.e., P^h, P^d, P^v and P^{-d}, only P^{-d} was involved. This is because the stationary distribution of the four matrices are almost the same and experimental results implies that we use P^{-d} can improve the detection accuracy compare with other three matrices.

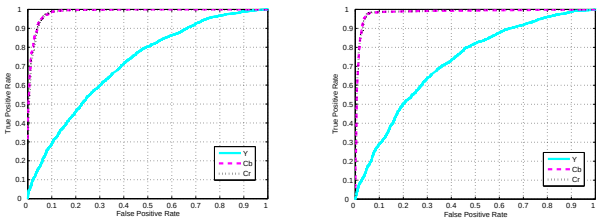
As we know, in classifier training and testing stages, we should keep balance of the number of authentic and tampered images. Hence, 5,123 authentic (randomly select from 7,491 authentic images) and 5,123 tampered images were selected to construct experimental database. The training samples (3,000 authentic and 3,000 tampered image) were randomly selected from the database. The remaining images were used in testing. We ran RBF kernel SVM classifier with the parameters C and γ which were determined by grid searching. Experiments with different thresholds Ts on same train and test database were carried out. The detection results using different image channels (Y, Cb and Cr) and different thresholds are shown in Table 1.

Table 1. Experiment results of proposed method

	$T = 8$			$T = 15$		
	Y	C_b	C_r	Y	C_b	C_r
AR	65.4%	**94.7%**	**94.9%**	66.5%	**95.6%**	**95.5%**
FPR	35.5%	**7.3%**	**7.3%**	36.3%	**6.7%**	**6.7%**
FNR	33.7%	**3.2%**	**3.0%**	30.9%	**2.1%**	**2.2%**

AR is detection accuracy rate. FPR and FNR are false positive rate and false negative rate, respectively. We make tampered images as positive samples and authentic images as negative samples in our experiments. The dimensions of feature vectors with $T = 8$ and $T = 15$ are 9 and 16 respectively. The effectiveness of our proposed feature is testified by the experiment with high detection accuracy and low FPR and FNR.

From Table 1 we can find that features extract from chroma (Cb or Cr) component perform much better than that from Y component, which coincide with what we have found in [11]. Though the detection accuracies of $T = 8$ is little lower than those of $T = 15$, the complexity of classifier training is much lower. Figure 4(a) shows



(a) The proposed method (b) The method proposed in [11]

Fig. 4. ROC curve of each component using proposed method (a) and the method proposed in [11] (b) with $T = 8$ on CASAI TIDE v2.0

the corresponding ROC curve of each component with $T = 8$. Table 2 shows the detection results of comparison experiment using the method proposed in [11] and Figure 4(b) are corresponding ROC curves. From these tables and figures, we find that performances of

Table 2. Experiment results of the method proposed in [11]

	$T = 8$		
	Y	C_b	C_r
AR	66.9%	**96.7%**	**96.8%**
FPR	32.0%	**4.3%**	**4.6%**
FNR	34.1%	**2.4%**	**1.8%**

these two methods are close, but the method in [11] needs boosting feature selection and dimension of the feature vector is much higher than our new approach. Figure 5 shows ROC curves of the experiment using proposed method on Columbia Uncompressed Image Splicing Detection Evaluation Dataset [12].

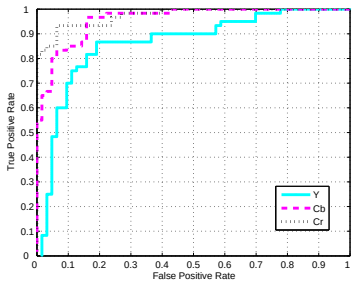


Fig. 5. ROC curve of each component using the proposed method with $T = 8$ on Columbia Uncompressed Image Splicing Detection Evaluation Dataset

4. CONCLUSIONS

In this paper, we have proposed a low dimension feature vector extraction method for color image tampering detection. We modeled the thresholded edge image of image chroma as a Markov chain and considered its stationary distribution as features for tampering detection. The experimental results have illustrated that the proposed 9-D feature vector is very effective for tampering detection.

In our approach, we only modeled the statistical dependency between two adjacent pixels of an image which can be characterized by one-step transition probability matrix of Markov chain. In fact, the actual dependency of image is not limited to just two adjacent pixel, hence, in our future work, the dependency among more than two adjacent pixels would be considered for further analysis. Though the work in this paper is only focus on image tampering detection, we are looking forward our proposed method can be useful at other similar forensics task like device source classification and steganalysis.

Acknowledgments.

This work is funded by research grants from the National Fundamental Research Program of China (Grant No.60603011) and the National Laboratory of Pattern Recognition. The authors also thank the anonymous reviewers for their valuable comments.

5. REFERENCES

[1] C. Rey and J.L. Dugelay, "A survey of watermarking algorithms for image authentication," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 6, pp. 613–621, March 2002.

[2] T.T. Ng, S.F. Chang, C.Y. Lin, and Q.Sun, "Passive-blind image forensics," in *Multimedia Security Technologies for Digital Rights*, chapter 06. Elsvier, 2006.

[3] T.T. Ng, S.F. Chang, and Q. Sun, "Blind detection of photomontage using higher order statistics," in *IEEE International Symposium on Circuits and Systems*, 2004.

[4] W. Wang, J. Dong, and T. Tan, "A Survey of Passive Image Tampering Detection," in *Proceedings of the 8th International Workshop on Digital Watermarking*. Springer-Verlag, 2009, pp. 308–322.

[5] M.K. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," in *ACM Multimedia and Security Workshop*, 2005, pp. 1–10.

[6] M.K. Johnson and H. Farid, "Exposing digital forgeries in complex lighting environments," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 450–461, 2007.

[7] A.C. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3948–3959, 2005.

[8] A.E. Dirik and N. Memon, "Image tamper detection based on demosaicing artifacts," in *IEEE International Conference onImage Processing (ICIP)*, 2009, pp. 1497 –1500.

[9] J. Lukáš, J. Fridrich, and M. Goljan, "Detecting digital image forgeries using sensor pattern noise," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Feb 2006, vol. 6072, pp. 362–372.

[10] Y.Q. Shi, C. Chen, and G. Xuan, "Steganalysis versus splicing detection," in *International Workshop on Digital Watermarking*, December 2007.

[11] W. Wang, J. Dong, and T. Tan, "Effective image splicing detection based on image chroma," in *IEEE International Conference on Image Processing*, 2009, pp. 1257–1260.

[12] T.T. Ng, S.F. Chang, and Q. Sun, "A data set of authentic and spliced image blocks," Tech. Rep., DVMM, Columbia University, 2004, Dataset: http://www.ee.columbia.edu/ln/dvmm/downloads/AuthSplicedDataSet/photographers.htm.

[13] S.M. Ross, *Introduction to Probability Models*, Elsevier, 9th edition, 2007.

[14] *CASIA Image Tampering Detection Evaluation Database*, 2010, http://forensics.idealtest.org.

[15] C.C. Chang and C.J. Lin, *LIBSVM: a library for support vector machines*, 2001, http://www.csie.ntu.edu.tw/∼cjlin/libsvm.