International Conference on Smart Computing and Electronic Enterprise. (ICSCEE2018) ©2018 IEEE

Copy-move Image Forgery Detection Based on Gabor Descriptors and K-Means Clustering

H M Shahriar Parvez
Faculty of computer science and information technology
University of Malaya
Kuala Lumpur, Malaysia
shahriar.parvez@siswa.um.edu.my

Hamid A. Jalab
Faculty of computer science and information technology
University of Malaya
Kuala Lumpur, Malaysia
hamidjalab@um.edu.my

Ala'a R. Al-Shamasneh
Faculty of computer science and information technology
University of Malaya
Kuala Lumpur, Malaysia
shamasneh@siswa.um.edu.my

Somayeh Sadeghi
Faculty of computer science and information technology
University of Malaya
Kuala Lumpur, Malaysia
sasomayeh1@gmail.com

Diaa M. Uliyan
Faculty of Information Technology
Middle East University
Amman, Jordan
duliyan@meu.edu.jo

***Abstract*—At present, popularity of using image as the fundamental media of information is growing. Rapid development of technology brings effective image processing tools available and makes image forgery very easy. As an outcome, it turns into a complicated issue in late time. In that case, validating the legitimacy and integrity of digital images is ending up progressively vital issue. The most challenging region-duplication forgery is made by copying some portion of an image and pasting on different region of the same image. This study proposes an efficient region-duplication forgery detection technique. This research is categorized into segment-based region duplication forgery detection method. The design of the algorithm based on image segmentation and using Gabor descriptors and K-Means clustering. Initially, the image is segmented using normalized cut (NCut) segmentation technique. Then, applied Gabor Filters to extract image features and cluster similar features using K-Means clustering algorithm. Finally, comparing the clustering regions with the given threshold value will decide image authenticity. Experiment results proves the strength of the proposed method against various post-processing attacks such as- rotation, scale, blurring and JPEG compression. A comparison with existing image forgery detection algorithms demonstrates that the proposed algorithm gives better performance.**

Keywords*—image forensics, image forgery detection, passive authentication, region-duplication forgery, copy-move forgery

I. Introduction

There is a significant role of digital images in our everyday life. However, image manipulation has become very easy by using powerful software. Without any doubt image authenticity now is a big matter of concern. Two main types of image forensic techniques are available to verify the integrity and authenticity of manipulated image. One is active forensic method [1] and another is passive forensic method [2][3] and they have been explained in the literature precisely. Watermarking and steganography are two techniques which are used in active methods to insert authentic information into the image. When question arise about the authenticity of an image, then prior embedded authentication information is recalled to prove the authenticity of that image. However, embedding authentication information to an image is very confidential. Only an authorized individual allows to do it or at the time of creating the image, authentic information could be embedded as well. Requirement of special cameras and multiple steps processing of the digital image are two main limitations which made this technique less popular. To avoid these limitations, the researchers are more interested in developing passive forensic technique as these methods utilize image forgery without requiring detailed previous information. The most popular method to construct forged image is copy-move forgery. It refers to copy one part from image, and paste it inside the same image. Sometime before pasting the copied regions, various post-processing operations like scale, rotation, blurring, intensifying or JPEG compression may be applied. The similarity regions between image clusters is used in this proposed method [4]. Firstly, the image is segmented using color regions by Normalized cut (NCut) segmentation technique. Secondly, centroids are identified from each region. Thirdly, image features for each region from the image are extracted by using Gabor descriptors [5]. In the next step, K-Means clustering is applied in order to cluster similar features from all regions. These clusters of features be compared to each other to identify which region has the most similar feature. Here, Euclidean distance calculation be used between each cluster for checking the threshold distance. If the similar feature cluster distance lies between the experimented threshold value, it is defined as tampered region otherwise authentic image. Finally, duplicated region will be localized. The paper is prepared as follows: Section 2 affords a discussion on the existing literature, while Section 3 details the proposed copy-move forgery algorithm.

Section 4 describes the experimented result. Lastly, the paper concludes with section 5.

II. Literature Review

Various number of region-duplication authentication techniques have been implemented to handle the issues identified with image authentication. All copy-move forgery detection techniques are developed with either block-based [9-13] or keypoint-based methods [14-17].

A. Block-based Image Forgery Detection

In block-based method, input image size of M x N is segmented into overlapping blocks size of z x z resulting into overlapping blocks, L = (M-z+1) x (N-z+1). A few features are extricated from each block. Distinctive features are extracted by applying different feature extraction technique such as DCT (Discrete Cosine Transform) [8], DWT (Discrete Wavelet Transform) [9], DFT(Discrete Fourier Transform) [10], PCA (Principal Component Analysis) [11] [12], SVD (Singular Value Decomposition) [13][14], and ZMs (Zernike Moments) [15]. Finally, a comparison is done based on blocks features similarity and distance. After finding the most matched features blocks, copy-move region is identified and localized. Sheng et al. [8] proposed forgery detection algorithm using block-based method. Proposed algorithm used DCT and circle blocking technique for extracting features from the image. Finally, the image which contains singularities within lines is presented by computing ridgelet transformation. Robustness against JPEG compression is the most significant feature of this method. Cao et al. [16] followed block-based method to detect tampered region where DCT feature extraction technique is applied. Initially, DCT is applied onto the divided sub blocks to extract key features by producing quantized coefficients. Threshold value was set prior to match features between closest similar image blocks. This method shows less computational complexity compared to existing methods [24-[17] because of reducing dimension of feature vector. Later, similar method and feature extraction technique used by Huang et al. [18]. The big difference in the result with Cao et al.'s DCT-based method [16], because of reducing the false matches rate. Due to low false matches rate, this method becomes powerful against noise and blurring. However, it is not robust for rotation attack and cannot detect multiple forgery. M. Bashar et al.,[19] developed more efficient forgery detection method based on DWT and kernel PCA (KPCA) features. Natural images have been used as a dataset in this method. As a consequence of quantitative analysis considering noiseless and uncompressed factor, it is found that the DWT performs well than KPCA in terms of features. On the other hand, in noise and JPEG compression domain, KPCA-based features perform better than DWT. The method shows robustness against noise and JPEG compression attack. Though the method takes too much time and not robust against scaling. It cannot detect multiple forgery.

B. Keypoint-based Image Forgery Detection Method

Different from block-based methods, features are extracted in keypoints-based method from the image without any type segmentation. Extracted features from every keypoints are compared to find similarities between them. Finally, based on the calculation of matched features, image forgery is detected. SIFT and SURF (Speeded Up Robust Features) are two main keypoints-based feature extraction methods. Somayeh Sadeghi et al., [20] and Diaa M. uliyan et al., [21] worked on keypoints-based technique (e.g. SIFT). Sadeghi et al., proposed SIFT to extract features and searched for similar features based on their Euclidean distance. Both methods are robust against several post-processing attacks; including scale, noise, rotation and JPEG compression. However, inability to detect small forged areas and performance of detection and localization for those forged areas are also questionable. In [21], primary approach of Uliyan et al., was to detect image regions by using Statistical Region Merging (SRM) Segmentation algorithm. Then, the experiment proceeded with applying Angular Radial Partitioning (ARP) and Harris Corner detection method on the image region. Finally, forged regions were detected based on matched keypoints. The method showed less robustness against forged regions with blurring and illumination attacks. Moreover, it shows different result for same image with different resolution. The major drawbacks of the previously mentioned conventional techniques are either not powerful against all post-processing attacks or have high computation time. Therefore, keeping up the low computational time is the most important robustness challenge. To tackle this issue, a new copy-move forgery detection method is proposed where region wise image segmentation is done. Gabor filters are used to extract image features. Afterwards, K-Means clustering, and Euclidean distance calculation facilitated to detect forged region from the suspicious image. Reducing the false matching rate is the most significant task to exhibit the proposed method as more viable compared to conventional methods.

III. Proposed Method

In this section, proposed blind copy-move image forgery detection algorithm based on Gabor descriptors and K-Means clustering is explained. Image features can be extracted by using many techniques such as edge detection, and corner detection [22] etc. Though, Gabor filter is chosen as feature extractor in the proposed method and image textures are extracted from each image region. Obviously, texture is the most preferred feature can be get after manipulating image regions every time. Compared to other feature extraction techniques [20][23][24], Gabor filter shows more robustness against scale, rotation, noise and JPEG Compression. It is faster for feature extraction as well. Several steps are followed in the proposed method: Image segmentation based on Ncut (normalized cut) segmentation technique and localizing the centroids of every region; extracting features by using Gabor descriptors and cluster similar features region based on K-Mean clustering, and lastly the detection of duplicated region in forged image. Proposed method is given as a block diagram in Fig. 1. Brief explanation of each steps is given in the following subsections.

A. Image Segmentation Using Normalized Cut and Centroids Detection

The proposed method employs normalized cut (NCut) segmentation to the input image I of size r x c. Image segmentation is done based on color and spatial threshold value with a minimum area of segmentation and smallest NCut value subject to the choice of image as a weighted graph. A weight matrix is calculated based on equation (1).

$$W_{ij} = e^{\frac{\|F(i)-F(j)\|_2^2}{\sigma_I^2}} * \begin{cases} e^{\frac{\|X(i)-X(j)\|_2^2}{\sigma_I^2}} & \text{if } \|X(i)-X(j)\|_2^2 < r \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where X denotes the spatial values and F is the feature in this case. Color values corresponding to point i, j and σ (sigma) refers to the specified threshold. A N x N diagonal matrix D is formed as $d_i = (\sum W_{ij})$ with subject to node I and a generalized Eigen system $(D-W)_y = \lambda D_y$ is solved to obtain two discreet eigenvectors such that we can bipartition the graph recursively based on the specified minimum threshold value [1][25][26]. Part of the experiment with tampered images and its NCut segmentations are displayed in Fig. 2. After image segmentation, the centroid of each segment is computed and a 7x7 block around the centroids for feature extraction and clustering have been proposed.

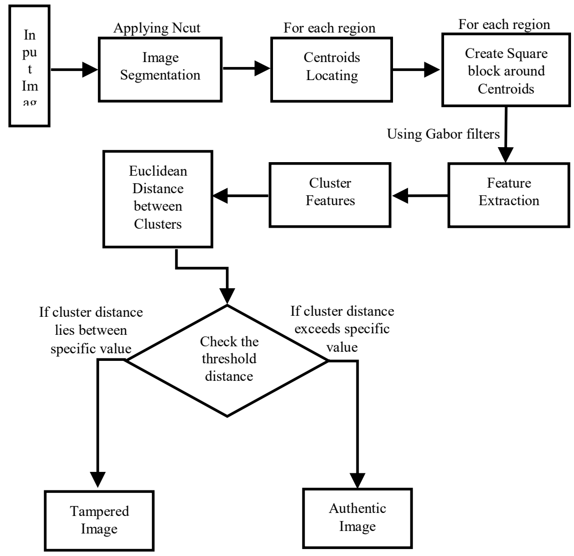


Fig. 1. Block diagram

In view of the proposed research, centroids hold the best significance of an image when compared with other methods [1][18][27][28][29]. The reason is, centroids bring into effective calculation of Euclidean distance between two clusters, also help to find the most similar clusters. Moreover, connection between centroids are utilized in order to calculate their Euclidean distance. For instance, when e and f represent two clusters respectively, then the linked centroids can be calculated as:

$$d(e, f) = P\overline{X_e} - \overline{X_f}P_2, \quad (2)$$

where $\overline{X_e}$ and $\overline{X_f}$ are defined as:

$$\overline{X_e} = \frac{1}{n_e}\sum_{i=1}^{n_e} X_i \, , \overline{X_f} = \frac{1}{n_f}\sum_{i=1}^{n_f} X_j, \quad (3)$$

From the above-mentioned equation, the variables $\overline{X_e}$ and $\overline{X_f}$ represent the weighted centroids of two different clusters i and j [20]. In the proposed method, the block size is experimentally fixed to 7x7 around the centroids. Finally, image segmentation using NCut and centroids calculation of each segment is shows as in Fig. 3.

(a) (b)

Fig. 2. (a) Copy-move forgery image (b) NCut segmentation

(a) (b)

Fig. 3. (a) NCut segmented image (b) Locating centroids

B. K-Means Clustering of Image Regions Based on Gabor Descriptors

Texture refers as one of the image features which is the normal repetition of an element or pattern on a surface. In Gabor filters, texture features are extracted by analyzing the frequency of the image utilizing frequencies and orientations. In the proposed method Gabor descriptors are applied in each image block to extract texture of each regions [29]. Experiment employed 2-D Gabor filtering technique, as the following:

$$f(x, y, \omega, \theta, \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x \sigma_y}$$

$$\exp\left[\frac{-1}{2}\left(\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2 + j\omega(x\cos\theta + y\sin\theta)\right)\right] \quad (4)$$

Where x, y denotes pixel position, ω is the centre frequency to the x and y directions. θ is for orientation and σ_x, σ_y represent standard deviation of the Gaussian along both x and y directions.

In the next step of experiment after feature extraction, K-Means clustering algorithm is proposed. It is an unsupervised technique and work as a classifier, which categorizes the input points into various clusters utilizing their in-between principle distance. Random intensities are used to cluster various information around centroids. As the following equation is used:

$$V = \sum_{i=1}^{k} \sum_{x_j} \in S_i \ (h(x) - \mu_i)^2 \quad (5)$$

Where k denotes as clusters. Proposed experiment has clusters value of k=2, S_i (where i = 1, 2, …, k) and μ_i= Centroid of all pixels x includes the histogram h(x) $\in Si$. At that point, similar clusters and collection of connected pixels are associated each other to create segmentation map as given below:

$$D = \arg\min j = ||h(x) - \mu_j||^2 \quad (6)$$

The repetition of j applies over every centroids [28].

C. Duplicate Region Detection

After examining the feature F_i of S_i, the Euclidean distance can be represented as the following:

$$D(f_i, f_j) = \sqrt{\sum_{i=1}^{n}(f_i - f_j)^2}, \ i \neq j \quad (7)$$

$D(f_i, f_j)$ represents the most relevant features of matched regions. The forged areas are to be found alongside their matched features [1]. Detection of duplicate region has been illustrated in Fig. 4.

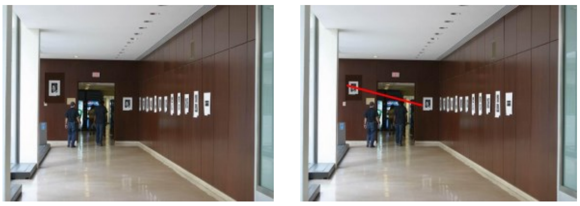
IV. EXPERIMENTAL RESULTS

The experiment is tested on MICC-F220 dataset [7] which is a popular dataset for the researchers who work on copy-move forgery detection. The dataset contains 220 images where 110 are original and 100 are tampered images. All tampered images in this dataset are post-processed by different attacks like scale, rotation, JPEG compression. The results of this proposed technique are illustrated in Fig. 5, 6, and 7 with sample images, and Table 1. Where Fig. 5 shows that the method can identify region duplication forgery accurately. Fig. 6 shows robustness of the proposed method against combined attack of rotation and JPEG compression. Lastly, Fig. 7 represents the ability of this forgery detection method against scale attack.

The TPR (true positive rate), and FPR (false positive rate) are used to measure the performance of proposed method.

$$T_{PR} = \frac{\text{No.of images detected as forged being forged}}{\text{No.of Forged Images}} \quad (8)$$

$$F_{PR} = \frac{\text{No.of images detected as forged being original}}{\text{No.of Original Images}} \quad (9)$$



(a) (b)

Fig. 4. (a) Forged image (b) Locating forged region


(a) (b)

Fig. 5. (a) Sample forged image (b) Detection result


(a) (b)

(c)

Fig. 6. (a) Forged image (b) Forged image with rotation attack ($\theta=45^0$) and recompress (JPEG Q Factor= 60) (c) Locating forged region


(a) (b)

Fig. 7. (a) Forged image with scale attack (factor 1.2) (b) Matched centroids between duplicate regions

TABLE I. AVERSGE DETECTION RATE

Operations	**Quality Factors**	**90**	**80**	**70**	**60**	**50**
Normal Copy Move	**TPR**	0.93	0.93	0.92	0.90	0.90
	FPR	0.02	0.02	0.03	0.04	0.05
Rotation	**TPR**	0.91	0.90	0.89	0.88	0.87
	FPR	0.09	0.10	0.11	0.11	0.21

Table 2 shows that the proposed method succeeded with high TPR value (93%) better than Mishra et al [30] and S.Lyu [31]. Moreover, the average FPS of the method is 2.89%, which is lower than Diaa Uliyan, et al., method [17].

TABLE II. DETECTION RESULTS OF PROPOSED METHOD

Methods	T_{PR} **%**	F_{PR} **%**
Segmented-based forgery detection method [17]	96	2.89
Keypoints-based forgery detection based on SURF feature extraction method [30]	73.6	3.64
Forgery detection based on SIFT feature extraction method [31]	89.96	1.25
Proposed Method	93.18	2.81

V. CONCLUSION

This study presented copy-move image forgery detection approach using Gabor descriptors and K-Means clustering. The experiment results demonstrate the ability of the proposed approach against post-processing operations like JPEG compression, scaling and rotation. Moreover, the proposed method gives better results compared with the existing methods in terms of detection rate. TPR is increased to 93.18%, while the FPR is 2.81%. However, the proposed method struggles for detecting multiple forged location in an image and also shows inaccurate result with noise attack. In conclusion, it is to believe that this proposed novel method would be beneficial to some extent in various areas of forensic science.

REFERENCES

[1] D. M. Uliyan, H. A. Jalab, and A. W. Abdul Wahab, “Copy move image forgery detection using Hessian and center symmetric local binary pattern,” in *ICOS 2015 - 2015 IEEE Conference on Open Systems*, 2016, pp. 7–11.

[2] R. Poisel and S. Tjoa, “Forensics investigations of multimedia data: A review of the state-of-the-art,” in *Proceedings - 6th International*

Conference on IT Security Incident Management and IT Forensics, IMF 2011, 2011, pp. 48–61.

[3] G. K. Birajdar and V. H. Mankar, "Digital image forgery detection using passive techniques: A survey," *Digit. Investig.*, vol. 10, no. 3, pp. 226–245, 2013.

[4] G. Lynch, F. Y. Shih, and H. Y. M. Liao, "An efficient expanding block algorithm for image copy-move forgery detection," *Inf. Sci. (Ny).*, vol. 239, pp. 253–265, 2013.

[5] H. C. Hsu and M. S. Wang, "Detection of copy-move forgery image using Gabor descriptor," in *Proceedings of the International Conference on Anti-Counterfeiting, Security and Identification, ASID*, 2012.

[6] D. M. Uliyan, H. A. Jalab, A. Abuarqoub, and M. Abu-Hashem, "Segmented-Based Region Duplication Forgery Detection Using MOD Keypoints and Texture Descriptor," in *Proceedings of the International Conference on Future Networks and Distributed Systems*, 2017, p. 6:1--6:6.

[7] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, "A SIFT-based forensic method for copy-move attack detection and transformation recovery," *IEEE Trans. Inf. Forensics Secur.*, vol. 6, no. 3 PART 2, pp. 1099–1110, 2011.

[8] G. Sheng, T. Gao, Y. Cao, L. Gao, and L. Fan, "Robust algorithm for detection of copy-move forgery in digital images based on ridgelet transform," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7530 LNAI, pp. 317–323.

[9] G. Muhammad, M. Hussain, and G. Bebis, "Passive copy move image forgery detection using undecimated dyadic wavelet transform," *Digit. Investig.*, vol. 9, no. 1, pp. 49–57, 2012.

[10] B. Mahdian and S. Saic, "Blind authentication using periodic properties of interpolation," *IEEE Trans. Inf. Forensics Secur.*, vol. 3, no. 3, pp. 529–538, 2008.

[11] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting duplicated image regions," *Dept. Comput. Sci., Dartmouth Coll. Tech. Rep. TR2004-515*, no. 2000, pp. 1–11, 2004.

[12] Z. Moghaddasi, H. A. Jalab, R. Md Noor, and S. Aghabozorgi, "Improving RLRN Image Splicing Detection with the Use of PCA and Kernel PCA," *Sci. World J.*, vol. 2014, 2014.

[13] D. Y. Huang, T. W. Lin, W. C. Hu, and C. H. Chou, "Boosting scheme for detecting region duplication forgery in digital images," in *Advances in Intelligent Systems and Computing*, 2014, vol. 238, pp. 125–133.

[14] W. Luo, Z. Qu, J. Huango, and G. Qiu, "A novel method for detecting cropped and recompressed image block," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2007, vol. 2.

[15] S. J. Ryu, M. Kirchner, M. J. Lee, and H. K. Lee, "Rotation invariant localization of duplicated image regions based on zernike moments," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 8, pp. 1355–1370, 2013.

[16] Y. Cao, T. Gao, L. Fan, and Q. Yang, "A robust detection algorithm for copy-move forgery in digital images," *Forensic Sci. Int.*, vol. 214, no. 1–3, pp. 33–43, 2012.

[17] D. M. Uliyan, H. A. Jalab, A. W. A. Wahab, P. Shivakumara, and S. Sadeghi, "A novel forged blurred region detection system for image forensic applications," *Expert Syst. Appl.*, vol. 64, pp. 1–10, 2016.

[18] Y. Huang, W. Lu, W. Sun, and D. Long, "Improved DCT-based detection of copy-move forgery in images," *Forensic Sci. Int.*, vol. 206, no. 1–3, pp. 178–184, 2011.

[19] M. Bashar, K. Noda, N. Ohnishi, and K. Mori, "Exploring duplicated regions in natural images," *IEEE Trans. Image Process.*, no. 99, 2010.

[20] S. Sadeghi, H. A. Jalab, K. S. Wong, D. Uliyan, and S. Dadkhah, "Keypoint based authentication and localization of copy-move forgery in digital image," *Malaysian J. Comput. Sci.*, vol. 30, no. 2, pp. 117–133, 2017.

[21] D. M. Uliyan, H. A. Jalab, A. W. A. Wahab, and S. Sadeghi, "Image region duplication forgery detection based on angular radial partitioning and harris key-points," *Symmetry (Basel).*, vol. 8, no. 7, 2016.

[22] H. A. Jalab and R. W. Ibrahim, "Texture enhancement based on the Savitzky-Golay fractional differential operator," *Math. Probl. Eng.*, vol. 2013, 2013.

[23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[24] R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, 2004, vol. 2, pp. 506–513.

[25] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.

[26] S. E. Chew and N. D. Cahill, "Semi-supervised normalized cuts for image segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, vol. 2015 Inter, pp. 1716–1723.

[27] L. Jing and C. Shao, "Image copy-move forgery detecting based on local invariant feature," *J. Multimed.*, vol. 7, no. 1, pp. 90–97, 2012.

[28] R. V. Patil and K. C. Jondhale, "Edge based technique to estimate number of clusters in k-means color image segmentation," *2010 3rd Int. Conf. Comput. Sci. Inf. Technol.*, vol. 2, pp. 117–121, 2010.

[29] R. Roslan and N. Jamil, "Texture feature extraction using 2-D Gabor Filters," in *Computer Applications and Industrial Electronics (ISCAIE), 2012 IEEE Symposium on*, 2012, pp. 173–178.

[30] P. Mishra, N. Mishra, S. Sharma, and R. Patel, "Region duplication forgery detection technique based on SURF and HAC," *Sci. World J.*, vol. 2013, 2013.

[31] X. Pan and S. Lyu, "Region duplication detection using image feature matching," *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 4, pp. 857–867, 2010.