Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 201 (2017) 436–444

Procedia Engineering

www.elsevier.com/locate/procedia

3rd International Conference "Information Technology and Nanotechnology", ITNT-2017, 25-27 April 2017, Samara, Russia

A new copy-move forgery detection algorithm using image preprocessing procedure

Andrey Kuznetsov[a,b]*, Vladislav Myasnikov[a,b]

[a]*Samara National Research University, Moskovskoye shosse 34, Samara, 443086, Russia*
[b]*Image Processing Systems Institute, Branch of the Federal Scientific Research Centre Crystallography and Photonics, Russian Academy of Sciences, Molodogvardeyskaya st. 151, Samara, 443001, Russia*

Abstract

One of the simplest ways of digital image forgery is a copy-move attack. The copy-move process consists of three stages: copy selected image fragment from one place, transform it using some image processing algorithm and paste it to another place of the same image. Two main approaches to copy-move detection exist nowadays: feature-based and hash-based. Most of the algorithms are developed according to the feature-based approach, whereas the hash-based approach is used only for plain copy-move detection (when the copied image fragment is not transformed). However, the main advantage of hash-based algorithms is low computational complexity. In this paper, we propose a new hash-based copy-move detection algorithm that can be applied to transformed duplicates detection due to a special preprocessing procedure. This procedure implements initial image transformation to incorporate the changes (affine transforms are not considered in this paper), produced by a transform algorithm on the second stage. Several preprocessing procedures are compared during experiments: image intensity range reduction, gradient calculation, expansion in orthonormal basis, adaptive linear contrast enhancement and local binary pattern.

Peer-review under responsibility of the scientific committee of the 3rd International Conference "Information Technology and Nanotechnology".

Keywords: Forgery; copy-move; detection algorithm; preliminary processing; hash value

* Corresponding author. Tel.: +7-927-759-6008.
E-mail address: kuznetsoff.andrey@gmal.com

1877-7058
Peer-review under responsibility of the scientific committee of the 3rd International Conference "Information Technology and Nanotechnology".
10.1016/j.proeng.2017.09.671

1. Introduction

One of the most common ways of image forgery is to embed a duplicate. In other words, this attack is called a copy-move. The embedding process consists of three stages: copying a fragment, adding changes (intensity or geometric) to this fragment, and inserting a fragment into that area of the image whose contents are supposed to be hidden from the end user. In this paper, we mean only intensity distortions when speaking about copy-move changes. These changes include linear contrast enhancement, noise addition, and filtering. An example of image forgery is shown in Fig. 1.

Fig. 1. Copy-move forgery image sample.

Nowadays there are a lot of works on copy-move detection algorithms development. Most of them correspond to the unified scheme of the algorithm presented in 2005 by H. Farid and A. Popescu [1], and, as a rule, differ in feature extraction step. This scheme is shown in Fig. 2. From the figure, you can distinguish two main types of algorithms, depending on the method used for feature extraction: based on keypoint detection (e.g., SIFT) and based on calculation some characteristics in a sliding window. As a preprocessing step, there are applied noise filtering methods, translation to other color spaces, etc. Feature vectors for analysis are calculated either using keypoints, or for every position of a sliding window [2-4]. The disadvantage of this scheme is high computational complexity, which limits the use of a sliding window for calculating features on large images ($>1000\times1000$). The reason for this is the comparison stage, when the search for nearest feature vectors is performed using some predefined metric. Moreover, this constraint leads to the increase in missed duplicates rate. Existing solutions [2-5] do not allow to achieve significant increase in performance. However there exist several techniques like [6] for features dimensionality reduction to decrease computational complexity.

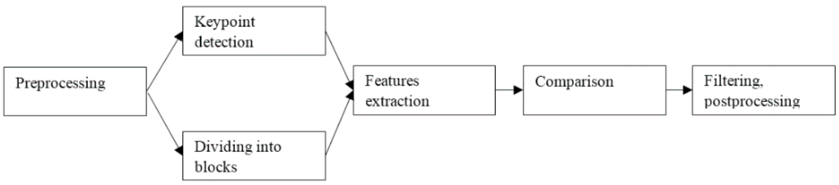

Fig. 2. Existing copy-move forgery detection scheme.

In the simplest case, when there are no duplicate transforms applied, we have pro-posed the algorithm for plain copy-move detection in 2014 [5]. The algorithm was based on hash values calculation in the sliding window mode. The proposed method demonstrated the absence of missed duplicates and a very low collision rate – false detected copy-move blocks (10^{-5}%). Moreover, the proposed solution had low computational complexity (in comparison with the existing features based scheme), which allows it to be used for analysis of remote sensing data and other large digital images.

In the case when duplicates are transformed (intensity or geometric), the developed hash-based approach is not applicable. For this reason, the problem arises of converting an input image with transformed duplicates to an undistorted state and therefore the previously developed hash-based solution can be easily applied. For this purpose, we need to transform the initial image to analyze it according to the hash-based approach.

In this paper, we propose a comparison of preliminary processing methods: image intensity range reduction, gradient calculation, expansion in orthonormal basis, adaptive linear contrast enhancement and local binary pattern. The comparison is carried out using true positive and false positive rates. Duplicate changes investigated in this paper include: intensity shift, linear contrast enhancement and additive white noise.

2. General scheme of the proposed solution

The core of the proposed solution, as mentioned above, is the plain copy-move detection algorithm, previously developed by the authors [7,9]. The algorithm is based on Rabin-Karp hash function [7,9], which values are calculated in a sliding window and are stored in a hash table to evaluate the frequency of hash values. Thus, to use this algorithm, it is necessary to incorporate changes of duplicates to the undistorted form. To solve this problem, we propose a new preliminary processing stage (see Fig. 3). The applied preliminary transformations are the following:

- image intensity range reducing;
- gradient calculation;
- expansion in orthonormal basis;
- adaptive linear contrast enhancement;
- local binary pattern.

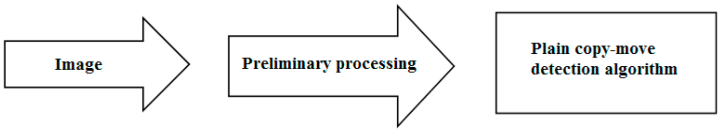

Fig. 3. General scheme of the proposed copy-move detection algorithm.

In the next section, we will briefly describe all the operations used on the preliminary processing stage.

3. Preliminary transformations

3.1. Image intensity range reduction

The most obvious way to reduce intensity distortion is to reduce intensity range. Therefore, the number of quantization levels is reduced. Let the value of each pixel of an 8-bit image be $f(m,n) \in [0, 2^8 - 1]$, which corresponds to 256 levels of quantization. We will reduce the number of quantization levels according to the following rule:

$$g(m,n) = \frac{f(m,n)}{2^{\varepsilon}}, \tag{1}$$

where $\varepsilon = 1..7$ characterizes the reduce of quantization levels to $2^{8-\varepsilon}$.

In addition to the image obtained using (1), there are created 2ε images according to the following equation:

$$g_i(m,n) = \frac{f(m,n)+i}{2^{\varepsilon}}, i \in [-\varepsilon, \varepsilon] / \{0\}. \tag{2}$$

After that, $2\varepsilon+1$ images are processed by the plain copy-move detection algorithm, and processing results are merged.

3.2. Gradient calculation

One of the simplest ways to detect edges is spatial differentiation of the intensity function $f(x,y)$. We will calculate the gradient for each image point - a 2D vector, the components of which are derivatives of the image intensity function along horizontal and vertical axes:

$$\Delta f(x,y)=\left(\frac{\partial f(x,y)}{\partial x},\frac{\partial f(x,y)}{\partial y}\right). \tag{3}$$

For each image pixel, the gradient vector is oriented in the direction of the greatest intensity increase, and its length corresponds to the magnitude of the intensity change. Since the direction can be arbitrary, the gradient value is determined as the length of the gradient vector and stored in a temporary image $\tilde{f}(m,n)$:

$$\tilde{f}(m,n)=\sqrt{(f(m,n)-f(m+1,n))^2+(f(m,n)-f(m,n+1))^2}\ . \tag{4}$$

The image $\tilde{f}(m,n)$ is then binarized using the following equation:

$$\hat{f}(m,n)=\begin{cases}1, \tilde{f}(m,n)>\tilde{f}(m,n+1) \,|\, \tilde{f}(m,n)>\tilde{f}(m+1,n) \,|\, \tilde{f}(m,n)>\tilde{f}(m+1,n+1)\\ 0, \mathrm{e}lse\end{cases}. \tag{5}$$

After that the following operations are applied in a loop to the image $\hat{f}(m,n)\in\mathbf{B}$ for 7 times:

- image smooth using Gauss filter 3×3;
- calculate gradient magnitude (4);
- binarization (5).

Eventually we generate 8 binary images $\hat{f}_i(m,n), i\in[0,7]$, which are then merged in a single 8-bit image:

$$g(m,n)=\sum_{i\in[0,7]}\hat{f}_i(m,n)\cdot 2^i. \tag{6}$$

3.3. Expansion in orthonormal basis

This method is based on the expansion of intensity values in the local neighborhood 5×5 of the current pixel in a basis. As a basis, we take linearly separable polynomials representable in the following form:

$$e_{ij}(x,y)=e_i(x)\cdot e_j(y), \tag{7}$$

where $e_i(x)$ - is an orthonormal basis in the range $[-2,2]$ explicitly defined as polynomials:

$$e_1 = \frac{1}{\sqrt{5}};$$
$$e_2 = \frac{1}{\sqrt{10}}x; \tag{8}$$
$$e_3 = \frac{1}{\sqrt{14}}x^2 - \frac{2}{\sqrt{14}}.$$

Using equation (7) and (8) we evaluate 9 2D filters $e_{ij}(x, y)$ with size 5×5. However, we do not use e_{11}, because it is a median filter. The other 8 filters are used to create the following images:

$$\widetilde{f}_{ij}(m,n) = \sum_{k=-2}^{2} \sum_{l=-2}^{2} f(m+k, n+l) \cdot e_{ij}(k,l), i, j \in [1,3], (i,j) \neq (1,1). \tag{9}$$

We can also modify this method by adding a normalization step when applying the filter $e_{ij}(k,l)$:

$$\widetilde{f}_{ij}(m,n) = \frac{\widetilde{f}_{ij}(m,n)}{\sum_{k=-2}^{2} \sum_{l=-2}^{2} f^2(m+k, n+l)}. \tag{10}$$

The result 8 images are binarized as follows:

$$\hat{f}_{ij}(m,n) = \begin{cases} 1, \widetilde{f}_{ij}(m,n) > median(\widetilde{f}_{ij}(m,n)) \\ 0, else \end{cases}. \tag{11}$$

3.4. Expansion in orthonormal basis

Intensity changes of a duplicate lead to changes in its local characteristics. To restore them, we use the method of adaptive linear contrast enhancement (ALC), which converts the local characteristics $\mu_f(m,n), \sigma_f(m,n)$ in a pixel $f(m,n)$ neighborhood $l \times l$ to the required μ_g, σ_g. To calculate parameters of linear contrast, we use the following equations:

$$a_l(m,n) = \frac{\sigma_g}{\sigma_f(m,n)},$$
$$b_l(m,n) = \mu_g - \mu_f(m,n)\frac{\sigma_g}{\sigma_f(m,n)}. \tag{11}$$

After that using (12) we transform pixels in the neighborhood $l \times l$ in a sliding window:

$$g_l(m,n) = a_l(m,n) \cdot f(m,n) + b_l(m,n). \tag{12}$$

The overall intensity of the image is adjusted to the specified characteristics μ_g, σ_g.

If it is necessary to detect copy-move with intensity shift, it is possible to use a simplified version of ALC, which is based on subtraction of the local mean in a sliding window. At the same time for other types of distortions, this method will show obviously worse results (for example, with linear contrast of a duplicate).
The other way of ALC is mapping pixels' values using *softmax* function:

$$g(m,n)=\frac{e^{f(m,n)}}{\sum_{(m',n')\in D} e^{f(m',n')}}. \quad (13)$$

3.5. Local binary pattern

Local binary pattern (LBP) is a well-known and commonly used local feature in pattern recognition. In practice, LBP operator combines statistical and structural properties of texture analysis, and enables to construct descriptors of digital images. To calculate LBP, a 3×3 window is usually used: 8 values are compared with the central pixel and a sequence of 8 bits is assigned to the center pixel. Several more complex forms of local patterns exist, and they were compared in [8] for copy-move detection problem solution using the feature-based approach. The LBP preliminary processing function is used to create the following result image:

$$g(m,n)=\sum_{\substack{i,j\in[-1,1]\\ k\in[0,7]}} I\big(f(m+i,n+j)\geq f(m,n)\big)\cdot 2^k, (i,j)\neq(0,0). \quad (14)$$

4. Experiments

We used a standard PC (Intel Core i5-3470 3.2 GHz, 8 GB RAM) to conduct experiments. We selected 10 8-bit images with sizes 512×512 for research. To create forgeries, we used the previously developed procedure for automatic copy-move generation [8-10] with controlling the size of duplicates, their quantity, as well as the algorithm and parameters of transformations. The following distortions of duplicates were generated:

1. Intensity shift (additive value $b\in[-20,20]$).
2. Linear contrast (multiplicative $a\in[0.6,0.9]$ and additive $b\in[-20,20]$ coefficients).
3. Additive white noise ($SNR\in[10,300]$).

To estimate the quality of the proposed solution we used true positive rate – *tp* and false positive rate – *fp*: *tp* is the number of correctly detected duplicates, *fp* is the number of missed duplicates.

Using the developed copy-move embedding procedure, 90 forgeries (30 images for each kind of distortion) were generated for each of the 10 images, which were further processed by the proposed copy-move detection algorithm, based on 8 different preliminary transformations: image intensity range reduction, gradient calculation, expansion in orthonormal basis (with normalization and without it), ALC, ALC simplified version, ALC based on the *softmax* function and LBP.

In Fig. 4 and Fig.5 the dependency of *tp* and *fp* values from the value of additive coefficient b of intensity shift is shown. Most of the preliminary transformations lead to high *tp* rate. The lowest quality is gained for image intensity range reduction transformation.

If the duplicate is distorted with linear contrast enhancement, then the best preliminary transformations leading to high detection quality are LBP and ALC. It is clearly seen that for all the transformations lower a values correspond to lower detection quality (see Fig. 6 and Fig. 7).

The main characteristic of additive white noise is signal-to-noise ratio (SNR), $SNR=\frac{\sigma_{signal}}{\sigma_{noise}}$. We added noise with $SNR\in[10,300]$ to change the pasted duplicates. The results of experiments are shown in Fig. 8 and Fig. 9.

As can be seen from conducted experiments, we obtained good results in the sense of the chosen quality criterion *tp* and *fp*. Most of the transformations lead to stable detection of copy-move forgeries for a wide range of distortions

parameters. At the same time, the computational complexity of feature-based solution considerably exceeds the computational complexity of the proposed method.
proposed method.

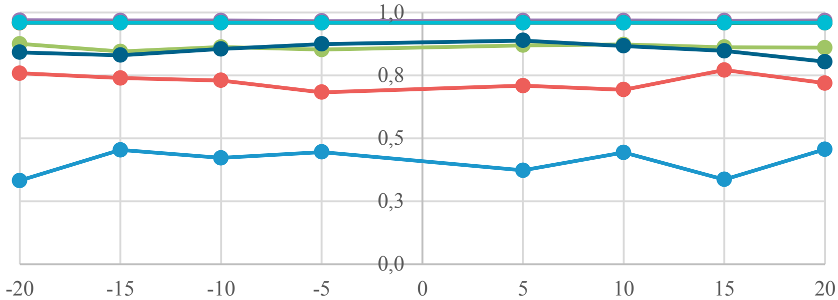

Fig. 4. The dependency of *tp* from the additive coefficient *b* of the duplicate intensity shift.

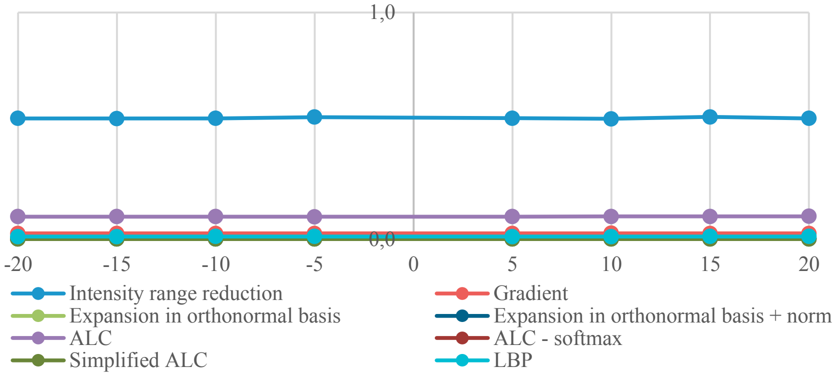

Fig. 5. The dependency of *fp* from the additive coefficient *b* of the duplicate intensity shift.

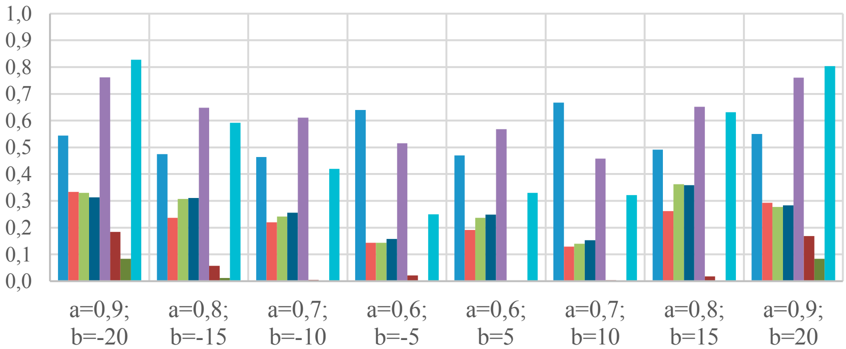

Fig. 6. The dependency of *tp* from *a* and *b* coefficients of the duplicate contrast enhancement.

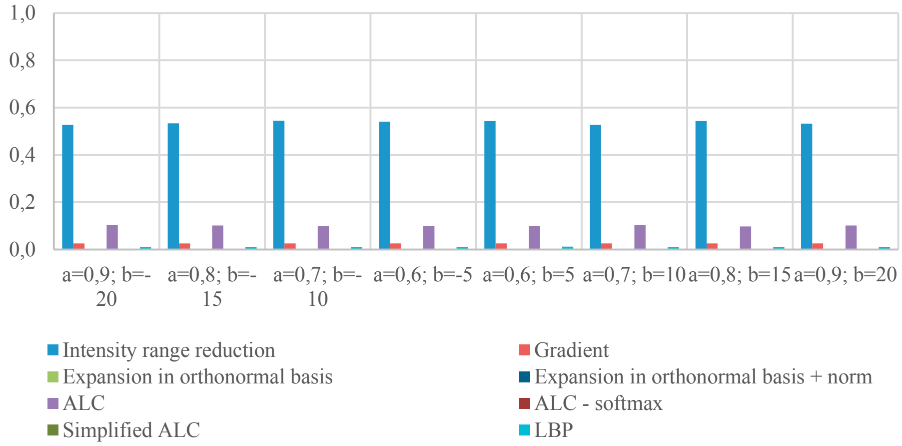

Fig. 7. The dependency of *fp* from *a* and *b* coefficients of the duplicate contrast enhancement.

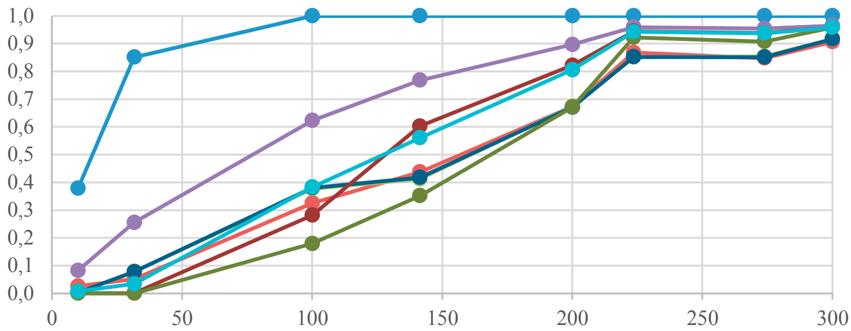

Fig. 8. The dependency of *tp* from *SNR*.

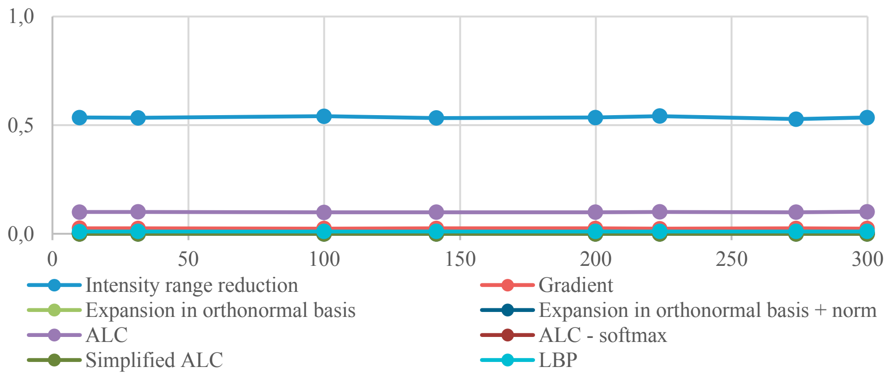

Fig. 9. The dependency of *fp* from *SNR*.

Considering the results of all experiments, the research in the field of image preliminary processing methods for copy-move detection is promising. With low computational complexity, it is possible to achieve high detection quality in comparison with existing feature-based solutions that obtain good results with significantly higher complexity.

5. Conclusion

In this paper, we showed the application of preliminary image processing methods to solve the problem of transformed copy-move detection. The following methods of image preliminary processing were taken for research: image intensity range reduction, gradient calculation, expansion in orthonormal basis, ALC and LBP. The carried out research showed high quality of copy-move detection with intensity shift distortions ($tp>0.7$, $fp<0.1$). In the case of linear contrast enhancement used to transform duplicates, calculation of LBP and ALC resulted in high values of detection quality in comparison with other preliminary processing methods. When duplicates were distorted by additive white noise, most of the transformations allowed to detected copy-move fragments for $SNR > 100$. Further research concerns continuing research in developing new preliminary processing methods to detect more complex forms of distortions and to expand the range of distortion parameters.

References

[1] A. Popescu, H. Farid, Exposing digital forgeries by detecting duplicated image regions. http://www.ists.dartmouth.edu/library/102.pdf, 2005.

[2] V. Christlein, C. Riess, J. Jordan, E. Angelopoulou, An evaluation of popular copy-move forgery detection approaches, IEEE Trans. Inf. Forensics Secur. 7 (2012) 1841–1854.

[3] Y. Cao, T. Gao, L. Fan, Q. Yang, A robust detection algorithm for copy-move forgery in digital images, Forensic Sci. Int. 214 (2012) 33–43.

[4] S. Bayram, H.T. Sencar, N. Memon, An efficient and robust method for detecting copy-move forgery, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 1053–1056.

[5] Y. Huang, W. Lu, W. Sun, D. Long, Improved DCT-based detection of copy-move forgery in images, Forensic Sci. Int. 206 (2011) 178–184.

[6] E.V. Myasnikov, The choice of a method for feature space decomposition for non-linear dimensionality reduction, Computer Optics. 38 (2014) 790–793.

[7] A. Kuznetsov, V. Myasnikov, A fast plain copy-move detection algorithm based on structural pattern and 2D Rabin-Karp rolling hash, Proceedings of International Conference on Image Analysis and Recognition 2014 in Lecture Notes in Computer Science. 8814 (2014) 461–468.

[8] A. Kuznetsov, V. Myasnikov, A copy-move detection algorithm using binary gradient contours, Proceedings of International Conference on Image Analysis and Recognition 2016 in Lecture Notes in Computer Science. 9730 (2016) 349–357.

[9] N.I. Glumov, A.V. Kuznetsov, V.V. Myasnikov, The algorithm for copy-move detection on digital images, Computer Optics. 37 (2013) 360–367.

[10] A.V. Kuznetsov, V.V. Myasnikov, A copy-move detection algorithm based on binary gradient contours, Computer Optics. 40 (2016) 284–293.