New scheme for fast copy-move detection

A V Kuznetsov\(^1,2\) and V V Myasnikov\(^1,2\)

\(^1\)Samara National Research University, Moskovskoe Shosse 34, Samara, Russia, 443086
\(^2\)Image Processing Systems Institute - Branch of the Federal Scientific Research Centre “Crystallography and Photonics” of Russian Academy of Sciences, Molodogvardeyskaya str. 151, Samara, Russia, 443001

e-mail: kuznetsoff.andrey@gmail.com

Abstract. Image forgery detection is a challenging and important problem for many years. One of the most frequently used types of forgery (copy-move) is copying and pasting image fragment within the same image. There are exist two main types of copy-move forgeries: on the one hand, they are created without applying transformations (plain copy-move) and on the other – different digital transforms can be used to hide the traces of copy-move forgery (transformed copy-move). To solve the task of copy-move forgery detection in a complex way we propose several novel techniques, described further. The carried out research showed high quality of copy-move detection with intensity shift, contrast enhancement and additive noise distortions. In conclusion part we provide limitations of the scheme and the ranges of distortion parameters that can be detected.

1. Introduction
Digital image forgery detection is a complex and important problem for many years. One of the most commonly used types of image forensics is copying and pasting an image fragment within the same image (copy-move embedding). There are two basic approaches for copy-move embedding (creating duplicates):

- fragments before pasting are not changed (plain copy-move);
- to hide embedding traces copy-move fragments are usually transformed by means of geometric, intensity or other transformations.

To solve the problem of copy-move detection we propose several new methods described below.

First, we developed a fast algorithm for plain copy-move detection in a sliding window mode using a special structural pattern and Rabin-Karp hash function. The novelty of the proposed approach is the absence of missed duplicates and low computational complexity when processing large images (for example, remote sensing data of the Earth). The main disadvantage of this solution is its inability to use it for distorted duplicates detection.
As for the development of the algorithm for detecting distorted duplicates, we propose a new solution, which is based on the use of a special pre-processing step, which eliminates distortions introduced into the duplicate. The conducted experiments showed high quality of distorted with local contrast enhancement and additive noise duplicates detection.

It should be noted that the existing duplicate detection scheme is widely used at present. Nevertheless, scientists conduct research on methods that have low computational complexity, which can be used to find distorted duplicates, transformed in a wide range of distortion parameters. In this area of research, we propose a distorted duplicate detection algorithm that uses functions based on binary gradient contours that are resistant to linear contrast, additive noise and JPEG compression. Comparing the quality of detection using algorithms based on binary gradient contours and based on various forms of local binary patterns showed a significant advantage in the accuracy of duplicates detection - 20-30%.

2. Literature review

Situational analysis is impossible without the use of digital images. In the domestic sense, digital images are transmitted through social networks to exchange information about some important events. The total number of digital images transmitted per day via popular web applications is enormous: almost 350 million images via Facebook [1] and 80 million images via Instagram [2]. This data is not protected, no one protects their authorship.

At present, protection against falsification of digital images is a complex problem. Everyone can make changes to the digital image with the help of modern software tools on a personal computer and on any mobile device in a few minutes. But no one can predict the consequences of these changes.

One of the most frequently used attacks on digital images is duplicate embedding. Duplicates in the image are fragments with similar content, obtained by copying and pasting the local fragment of this image. The copied areas are called duplicates. To hide the traces of copying and embedding the fragment, transformations such as local contrasting, changing lighting conditions, adding noise, applying compression and affine transformations can be applied.

By the moment, four review articles on the development of duplicate detection schemes have been published. In a review by Christlein et al. [3] the authors discussed the effectiveness of popular methods of extracting traits in the task of detecting duplicates. Then, the effectiveness of the approach based on various characteristics was analyzed on their own copy-move dataset. Lin et al. [4] classified the methods of matching areas on the image by direct comparison of blocks. Meanwhile, Al-Qershi and Khoo in 2013 classified [5] existing methods of extracting traits and analyzed their advantages and disadvantages. Warif N.B.A. and others (2016) in [6] presented a new look at the current research of duplicate detection algorithms and described duplicate detection schemes.

The latest review, published in 2017, by Salleh S.F.M. et al. [7] contains the analysis of computational complexity problem in copy-move detection algorithms and ways to solve this problem. The authors compared the scheme of direct block comparison, as well as more complex parallel methods.

Forgery detection methods can be classified to active and passive approaches, depending on the availability of a priori information. An active approach is based on embedding additional information in a digital image to detect unauthorized access, such as digital watermarks and digital signatures. Such information is used to confirm the originality of the data. On the other hand, the passive approach is able to detect changes in images without using additional information. An overview of methods of active and passive protection of digital images from counterfeits is shown in Figure 1 [6].
3. Copy-move detection schemes. Block-based and keypoint-based methods

Currently, there are many works on the development of copy-move detection algorithms. Most of them correspond to the unified scheme of the algorithm presented in 2005 by Farid and Popescu [8], and, as a rule, differ by the method of choosing the comparison scheme: regular partitioning into (possibly overlapping) blocks and keypoints based. Historically, the first approach appeared earlier [8]. Its significant disadvantage is the high computational complexity - the number of blocks obtained by dividing the image on a regular grid is proportional to the number of sample counts, and therefore the number of performed comparisons depends on the number of blocks quadratically. The immediate consequence of this fact is the need to select blocks as close as possible to each other, which leads to an increase in the probability false duplicates detection. The computational complexity of the second approach in practice turns out to be lower because of the relatively small number (in comparison with the image size) of the detected keypoints. Unfortunately, the same fact is the reason for not being able to guarantee the detection of all duplicates, especially for fragments of an image with a low dispersion or regular structure (for example, textures). Both schemes are combined and presented in Figure 2.

\[ F1 = \frac{2 \cdot tp}{2 \cdot tp + fp + fn}, \]

where \( tp \) is the number of true detected duplicates, \( fp \) is the number of missed duplicates, \( fn \) is the number of false detected duplicates.

As it was stated above, the main goal of our research is the development of a computationally efficient algorithm that does not miss duplicates.

Figure 1. Active and passive forgery detection methods.

Figure 2. Copy-move detection scheme.
4. Copy-move detection algorithm based on hash functions

Let there be given a structural pattern \( S = \{(0,0), (a,b)\} \) and an image \( f \), where \( f(m,n) \in [0,2^s - 1] \).

Then a hash value for an image fragment will be calculated as a 2D generalization of Rabin-Karp rolling hash:

\[
H(m,n,f) = f(m,n) \cdot 2^{q(ab-1)} + f(m,n+1) \cdot 2^{q(ab-2)} + \ldots \\
+ f(m,n+b-1) \cdot 2^{q(ab-b)} + \ldots + f(m+a-1,n) \cdot 2^{q(b-1)} + \\
+ f(m+a-1,n+1) \cdot 2^{q(b-2)} + \ldots + f(m+a-1,n+b-1) \cdot 2^0.
\]  

(1)

Let us also consider the following simplifications:

- \( f(m,n) \mod b_i = f(m,n) \) due to assumptions, that \( f(m,n) \in [0,2^s - 1] \) and \( b_i >> 2^s - 1 \);
- \( p_i = 2^s \mod b_i \), \( i \in [1, b - 1] \) can be calculated once before image analysis (no need to calculate these values for every sliding window position);
- \( \mod b_i \) calculation for every summand can be discarded because the computable sum will not exceed a bit grid dimension (let us consider, that \( b_i = 2^{31} - 1 \), then the order equals to \( ab \cdot 2^{31} \), so the value of (1) can be stored in LONG numeric type if \( ab < 2^{24} \)).

In consideration of these simplifications, expression (1) is changed to the following:

\[
H_r(m,n,f) = \left( f(m,n) p_i^{ab-1} + \\
+ f(m,n+1) p_i^{ab-2} + \ldots \\
+ f(m+a-1,n+b-1) \right) \mod b_i.
\]  

(2)

It is convenient to use \( b_0 = 2^{31} - 1 \). For \( b_i, i > 0 \) we suggest to select prime numbers less than \( b_0 \).

Hash value (2) can be calculated recursively for a structural pattern \( S = \{(0,0), (1,b)\} \) both for initial and modular representations as follows:

\[
H(m,n,f) = 2^s \left( H(m,n-1,f) - 2^{q(b-1)} f(m,n) \right) + f(m,n+b-1).
\]

For a 2D structural pattern the recursive algorithm will be as following (for rows and columns correspondingly):

\[
H(m,n+1,f) = 2^s H(m,n,f) - \\
-2^q f(m,n) + \sum_{i=1}^{b-1} 2^{q(ab-i)} \left( f(m+i,n+b-1) - f(m+i,n) \right) + \\
+ f(m+a-1,n+b-1),
\]

\[
H(m+1,n,f) = 2^q \left( H(m,n,f) - \sum_{i=0}^{b-1} 2^{q(ab-i-1)} f(m,n+i) \right) + \sum_{i=0}^{b} 2^{q(b-i)} f(m+a,n+i).
\]

Multiplication by powers of 2 is computed effectively using register shift operations. This simplification reduces hash value calculation time.

The scheme of plain copy-move detection algorithm is presented in Figure 3.

Figure 3. Plain copy-move detection algorithm scheme.
5. General approach to copy-move detection using hash functions

The core of the proposed solution, as mentioned above, is the algorithm for detecting plain duplicates, previously developed by the authors [9, 10]. The algorithm is based on the Rabin-Karp hash function [9], which values are calculated in a sliding window and stored in a hash table to estimate the frequency of the hash value occurrence. Thus, to use this algorithm to detect distorted duplicates, it is necessary to "convert" the transformed local areas in some way to an undistorted view. To solve this problem, we propose to use a new stage of preliminary processing (Figure 4).

Figure 4. Fast copy-move detection algorithm scheme.

An obvious requirement for the preprocessing procedure is the coincidence of the processing results for duplicate blocks. In general, this requirement can not be guaranteed. At the same time, for certain brightness transformations and some preprocessing parameters, this requirement can be fulfilled.

In the framework of this work, we used such transformations as reduction of the range of luminance levels of the image, calculation of the gradient, adaptive linear contrast, decomposition with respect to the orthonormal basis, various forms of local binary patterns.

6. Experimental research

We used a standard PC (Intel Core i5-3470 3.2 GHz, 8 GB of RAM) to conduct experiments. We selected 10 8-bit images with size $512 \times 512$ for research. To create forgeries we used the previously developed procedure for automatic generation of duplicates [9-11], which supports the control of the sizes of embedded duplicates, their number, as well as the type and parameters of the transformations. During the generation of a sample set of duplicates, the following distortions of local duplicate fragments were used:

- Brightness shift;
- Linear contrast (multiplicative and additive coefficients);
- Additive white noise.

Using the developed procedure for automatic embedding of duplicates, 90 forgeries (30 images for each type of distortion) were created for each of the 10 images that were further processed by the proposed copy-move detection algorithm based on 8 different preliminary transformations: contrast range decrease, gradient calculation, orthonormal decomposition basis (with and without normalization), adaptive linear contrast (ALC), its simplified version of ALC and ALC based on the softmax function.

Figures 5 and 6 show the dependence of the values of $tp$ and $fp$ on the value of the additive coefficient $b$ of the brightness shift. Most preliminary transformations lead to an increase in the value of $tp$. The lowest quality is achieved by converting the reduction in the intensity of the image.

If the duplicate is distorted by linear contrast, then the best preliminary transformations leading to a high quality of detection are local binary pattern (LBP) and ALC. It is clear that for all transformations below the value of $a$ correspond to a lower detection quality (Figures 7 and 8).

The main characteristic of additive white noise is the signal to noise ratio (SNR). We added noise as part of the procedure for embedding duplicates. The results of the experiments are shown in Figures 9 and 10.

As can be seen from the conducted experiments, we obtained good results in terms of the chosen quality criterion $tp$ and $fp$. Most transformations lead to a stable detection of duplicates for a wide range of distortion parameters. At the same time, the computational complexity of the proposed method is significantly lower than for block approaches that use invariant transformations of features.
The disadvantage is still a narrow spectrum of distortions, which can be eliminated by means of a preliminary processing procedure.

**Figure 5.** Dependency of $t_p$ from $b$ coefficient value of intensity shift.

**Figure 6.** Dependency of $f_p$ from $b$ coefficient value of intensity shift.

**Figure 7.** Dependency of $t_p$ from $a$ and $b$ coefficients values of intensity shift.

**Figure 8.** Dependency of $f_p$ from $a$ and $b$ coefficients values of intensity shift.
7. Results
Considering the results of all experiments, studies in the field of image pre-processing methods for detecting duplicates are promising. The best solution in accordance with the values of $tp$ and $fp$ for all ranges of parameters of applied transformations is a method that combines the preliminary processing procedure with local binary patterns (as well as the best results for block-based approach [12]). With a sufficiently low computational complexity through the developed scheme, it is possible to achieve high detection quality in comparison with existing block-based and keypoints-based solutions that provide good results with much greater computational complexity.

8. References
[1] Access mode: https://www.brandwatch.com/blog/47-facebook-statistics-2016/
[2] Access mode: https://www.brandwatch.com/blog/37-instagram-stats-2016/
[3] Christlein V, Riess C, Jordan J and Angelopoulou E 2012 An Evaluation of Popular Copy-Move Forgery Detection Approaches IEEE Trans. Inf. Forensics Secur. 7 1841-1854
[4] Lin W, Khan S U, Yow K C, Qazi T, Madani S A, Xu C-Z, Kolodziej J, Khan I A, Li H and Hayat K 2013 Survey on blind image forgery detection IET Image Process 7 660-670
[5] Al-Qershi O M and Khoo B E 2013 Passive detection of copy-move forgery in digital images: state-of-the-art Forensic Sci.Int. 231 284-295
[6] Warif N B A, Wahab A W A, Idris M Y I, Ramli R, Salleh R, Shamshirband S and Choo K K R 2016 Copy-move forgery detection: Survey, Challenges, and Future Directions J. Network Comp. Apps 75 259-278
[7] Salleh S F M, Rohani M F and Maarof M A 2017 Copy-Move Forgery Detection: A Survey On Time Complexity Issues And Solutions Journal of Theoretical & Applied Information Technology 95(11) 2552-2566
[8] Popescu A C and Farid H 2004 Exposing Digital Forgeries by Detecting Duplicated Image Regions Dartmouth College, Computer Science 515 1-11
Kuznetsov A and Myasnikov V 2014 A fast plain copy-move detection algorithm based on structural pattern and 2D Rabin-Karp rolling hash Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8814(1) 461-468

Kuznetsov A and Myasnikov V 2016 A Copy-Move Detection Algorithm Using Binary Gradient Contours Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9730 349-357

Evdokimova N and Kuznetsov A 2017 Local patterns in the copy-move detection problem solution Computer Optics 41(1) 79-87 DOI: 10.18287/2412-6179-2017-41-1-79-87

Kuznetsov A and Myasnikov V 2016 A copy-move detection algorithm based on binary gradient contours Computer Optics 40(2) 284-293 DOI: 10.18287/2412-6179-2016-40-2-284-293

Acknowledgements
This work was supported by the Federal Agency of scientific organization (Agreement 007-GZ/43363/26) in parts "Copy-move detection schemes. Block-based and keypoint-based methods" and “Copy-move detection algorithm based on hash functions” and by the Russian Foundation for Basic Research (#17-29-03190 - ofi_m) in part "Experimental research".