Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

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

Regular Paper

A new copy move forgery detection technique with automatic threshold determination

Beste Ustubioglu, Guzin Ulutas *, Mustafa Ulutas, Vasif V. Nabiyev

Department of Computer Engineering, Karadeniz Technical University, 61080 Trabzon, Turkey

ARTICLE INFO

Article history:
Received 5 May 2015
Accepted 6 May 2016

Keywords:
Copy move forgery
Benford's generalized law
DCT phase

ABSTRACT

Ensuring authenticity of images has become an important issue recently. Copy move forgery is one of the most common tampering techniques used to modify images. Copy move forgery detection techniques in the literature divide the image into overlapping blocks and use various techniques to extract features from the blocks. Similarity between the feature vectors is a clue about the forgery. However, these techniques use a predefined threshold to test the similarity. Test images with different characteristics require various threshold values. Determination of the best threshold value can be troublesome because the range of the feature vector elements' cannot be predetermined. Therefore, many experiments must be realized to find the best threshold value. In this work, we utilize DCT-phase terms to restrict the range of the feature vector elements' and Benford's generalized law to determine the compression history of the image under test. The method uses element-by-element equality between the feature vectors instead of Euclidean distance or cross correlation and utilizes compression history to determine the threshold value for the current test image automatically. Experimental results show that the method can detect the copied and pasted regions under different scenarios and gives higher accuracy ratios/lower false negative compared to similar works.

1. Introduction

Nowadays, digital images have been used in many areas such as a evidence in a courtroom or for insurance claim, in a scientific fraud, within the medical patient history etc. Development of new powerful image editing software is also increasing parallel to increase on the usage of images in daily life. Ease of use of such software makes even a casual user an expert for image forgery. Thus, anyone can modify an image even if he has no know-how in that field and the forged image does not exhibit any visible clues about the modification. As a result, developing methods to prove the authenticity of an image becomes an active field of research. There are two techniques in the literature to authenticate images: Active and Passive Techniques.

Active techniques used in the literature can be classified into two sub groups: Digital Watermarking and Digital Signatures. Digital watermarking technique generates a watermark using some method and inserts it into image without any visible clues. However, either the digital device can realize watermark insertion at the time of the capture or by software after capture. This requirement is the main drawback of the watermarking approach. It necessitates either special device (high cost) or special software. The methods in the second subgroup use Digital Signatures to verify the digital images. A unique signature is generated from the image and it contains information about the original image. Other party must regenerate the signature during authentication and must compare the regenerated signature with the received one. If they are consistent, the image is authenticated. Digital signature based methods also require special software and necessitates signature generation for each image to be authenticated.

Passive Techniques become popular recently since they do not require any prior information. The methods in this group use the underlying statistics of the test image to detect the forged regions. Researchers in the literature presents two type forgery that can be applied on an image: Copy-move forgery and Image splicing. Image splicing technique obtains regions from other images and put them into an image to create the forged image. Copy move forgery technique uses a region from an image to hide a different region on the same image or uses a region to replicate it on the same image. Fig. 1 shows an example of copy move forgery operation. Fig. 1(a) is the original image whereas Fig. 1(b) shows the forged version of the original image.

* Corresponding author. Tel.: +90 533 2277990.
E-mail address: guzin@ieee.org (G. Ulutas).

http://dx.doi.org/10.1016/j.aeue.2016.05.005
1434-8411/© 2016 Elsevier GmbH. All rights reserved.

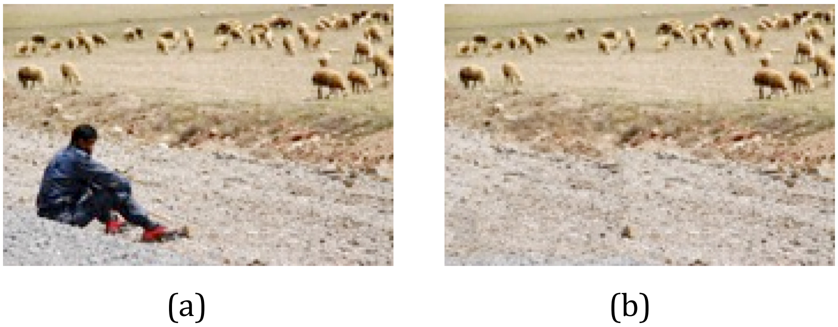

Fig. 1. (a) Original image. (b) Forged image.

Easy implementation of the copy move forgery operation makes it popular among the forgery techniques. Therefore, many works in the literature proposed a method to detect the copied and pasted regions on the forged image. Fridrich et al. suggested a method in 2003 to detect forgery in the literature for the first time [1]. Their method divides the test image into overlapping blocks. Discrete Cosine Transform is used to extract feature vectors. Similarity between the vectors is a clue about forgery. However, in most forged images, these clues are minimized by some post processing operations such as JPEG compression, noise addition, blurring, rotation, scaling etc. Using post-processing operations after forgery prevents the copied and pasted regions to be visible to human eyes. In this case, Fridrich et al.'s method does not work properly. Proposed methods in the literature from 2003 have two goals:

1. Make the feature vector to be smaller to reduce the complexity.
2. Find the robust feature extraction methods against the post processing operations.

Copy move forgery detection techniques in the literature can be divided into two subgroups: Block based and Key-point based methods.

Block based methods have similar algorithmic structure: Dividing the test image into overlapping sub-blocks, extracting features from the blocks, matching the feature vectors and marking forged regions. Methods in the literature are different from each other with the algorithm they used to extract feature vectors. Many frequency domain techniques (DCT, DWT, FMT, LPT, etc.) are utilized to extract features from the blocks. Some works to represent the blocks also uses image moments such as Hu, Zernike and Blur moments.

Key-point based methods extract descriptors from the forged image using key-point extraction algorithms such as SIFT and SURF. Methods in the literature proposed to use different clustering techniques to evaluate the key-point correspondence. The details of the methods in each subgroup will be given in the next section.

The proposed method is in the first group and utilizes DCT phase term to extract the feature vectors. The method generates feature vectors with elements that can take values from {−1, 0, 1}. DCT phase ensures restricted range of the elements of the feature vectors. Thus element-by-element matching can be used due to an element can only take three integer value. During the determination of the similarity between the vectors, the method uses element-by-element matching in case of using the spatial or frequency based measurement techniques. Methods in the literature use spatial or frequency based distance measurement techniques (Euclidean distance, Phase Correlation, etc.) to determine how two feature vectors are similar. So, their method must determine the appropriate threshold value to test the similarity. Many experiments must be realized to find the appropriate threshold value. In the proposed method, we use Benford's generalized law to judge the test image has been compressed before. If it is, quality factor will be estimated by the method. Threshold value will be set according to compression history of the test image. The threshold value indicates how many elements of two vectors must be equal to judge similarity. However the methods in the literature must specifically find the best value for the threshold. (e.g. If Euclidean distance between two vectors is smaller than 0.0025, they are same). Thus, proposed method automatically determines the best threshold value and eliminates extra tests to determine the exact value for the similarity threshold. When the method is compared to other works reported in the literature, it shows better results under various post-processing operations as shown in the experimental results.

The paper is organized as follows. Section 2 gives the related work in the literature. The details of the proposed method and experimental results will be given in Section 3 and Section 4 respectively. The conclusion is drawn in Section 5.

2. Related work

Researchers proposed many algorithms in the literature after Fridrich's paper to detect copy move forgery. The methods proposed between 2003 and 2008 share similar framework. They divide the image into overlapping blocks and use feature extraction techniques to extract vectors corresponding to blocks. Similarity between feature vectors indicates possible forgery among blocks corresponding to features. In 2008, Huang et al. used a key-point extraction algorithm to extract features from the whole image instead of using only one block [14]. After his work, the literature is divided into two categories: Block based methods and Key-point based methods. A brief introduction of both categories is given in the following sections.

2.1. Block based methods

In this section, block based copy move forgery detection methods are explained. They are distinguished by the feature extraction methods used.

Fridrich et al. proposed the first approach in the literature to detect copy move forgery [1]. Their method utilized DCT [2] to extract features from overlapping blocks. Extracted feature vectors are stored in a matrix and the matrix is lexicographically sorted to move the similar vectors closer. Each feature vector is compared to neighboring vectors and Euclidean distance is calculated for each pair. If the distance is smaller than a predefined threshold, a vector called shift vector between the upper left coordinates of corresponding blocks will be calculated. When the number of the same shift vectors exceeds a predefined threshold, blocks designated by these vectors are marked as forged. The main drawback of the