Copy-Move Forgery Detection in Digital Images by Reduced Features Ovelapping Block-Based Algorithm

EVA TUBA
University of Belgrade
Faculty of Mathematics
Studentski trg 16, Belgrade
SERBIA
etuba@acm.org

RAKA JOVANOVIC
Hamad Bin Khalifa University
Qatar Environment and Energy Research Institute
PO Box 5825, Doha
QATAR
rjovanovic@qf.org.qa

MILAN TUBA
John Naisbitt University
Faculty of Computer Science
Bul. umetnosti 29, Belgrade
SERBIA				tuba@ieee.org

Abstract: Digital image forgery is one of the problems that, besides many benefits, appears with widespread use of digital images in all areas of life. Unfortunately, with rapidly developing and increasingly more powerful hardware and software, not only is digital photography manipulation for legitimate goals easier, but also for the forgery. One well-known forgery of digital images is the so-called copy-move forgery where one part of the image is copied to another location in the same image. In this paper we proposed an overlapping block-based method for detection of copy-move forgery. It uses overlapping blocks of the size 16*16 and only three features extracted from such blocks. Huge amount of computation is reduced by using bin-sort algorithm. The proposed method was tested on standard benchmark images and in spite of its simplicity it proved to be very successful.

Key–Words: Digital image forensics, image forgery detection, copy-move forgery detection, block-based forgery detection algorithm

1 Introduction

A variety of information can be shared around the world in a very inexpensive way, through the Internet. The world economy relies entirely on communication via the Internet. The wide availability and low price of the equipment and the technology itself are very facilitating for everyday life. But that penetration to all aspects of human life has also some drawbacks. With the rapid expansion of Internet technology in the world, Internet crime also expanded. World wide web has enabled criminals to do all sorts of criminal activities that threaten all users on the Internet. There are variety of threats from identity theft, money theft to national security threats. No one is fully protected from these threats. Although many people think that there is a hundred percent protection, it is simply not so. For example, a hacker can get hold of discriminatory photos of a person that can be used for blackmail in different ways.

Digital forensics investigators need a reliable and stable tools to keep pace with criminals, and if possible to be one step ahead. Digital forensics encompasses many areas. The images are full of information and are very widespread on the Internet. Digital images can be quite easy to change because they do not have any form of protection. Due to the wide availability of software tools on all desktop and mobile platforms one can easily edit any image. Changing the image represents a threat to everyone because, with the help of tools, one can change the meaning and significance of the image very easily. When changed picture begins to spread on social networks that picture can make significant harm. It is therefore very important to be able to determine the originality of the picture. Determining the integrity and originality of the image content is the essence of research in digital image forensics. Digital image forensics in the last ten years has become very important. One of the important papers is the [1]. Due to the method which is reflected in the fact that the image is reversed and divided on circular blocks, the properties of the blocks are obtained by rotating the local binary patterns (LBP). The main things that digital forensics is trying to clarify is the determination of the authenticity of the source image and confirmation of the integrity of the picture. Authentication sources involves the interrogation techniques of the work whether on camera or through some algorithm. Integrity involves examining images by searching for any modification. Digital forensics is divided into active and passive.

Active digital forensics refers to control of digital signatures. The signatures are putted on the images during their creation. Any image change will also change the signature of a greater or lesser extent. Changes to the signature subsequently suggest that
the image has been altered. But to determine that the picture is really changed original image as well as image which is supposedly changed are needed. In addition, it is necessary to use very expensive equipment during the verification process.

Another way to protect the digital image is by putting the digital watermark. Applications for watermarking can be use for protecting the copyright of the digital image and also for protecting its integrity. In the case of protecting the copyrights, digital watermark is suppose to hide some number of bits but without significant changes of the cover. Even if some changes were made on image, image should be able to recover. On the other hand, if the goal of digital watermark is to protect the image integrity than in the case of forgery attempt, the image gets destroyed. This is a method to control in some way the image content manipulation. One of the main problems with this application is inability to differentiate malicious modifications and naive ones such as a fair compression.

Passive forensics does not require a comparison of the original and modified images and special equipment. In the paper [2] methods that are good in detecting fraud were listed. Passive forensics in each original image sees a scheme that is permanent. When a change occurs in the image this pattern or a scheme is changing along with the image. A number of methods was given in the paper [3], where an overview of the progress of ways to detect fraud was given.

Forensics of a digital image may be based on the cameras, pixels and others. Copy-paste detection falls under the forensics of pixels. The most widespread image change technique is copy-paste where one part of the image is copied to another part of the same image. Although copies are identical to the original part the copying process can be very skillfully done, so determining the use of copy-paste method is not easy. If there is no post production, this method is a little easier to detect. Discovering this fraud becomes more difficult when the copied part is rotated, cut, extended, etc. Copy-paste technique is also known as copy-move forgery.

In this paper we propose a method for detecting copy-move forgery based on features extracted from blocks of the image. Image is divided into blocks of size 16×16 and further more from each block three different features were extracted. Three different features from these blocks are used for detecting copy-move forgery by using the bin-sort algorithm.

This paper is organized as follow. In Section 2, literature review is presented. In Section 3 different digital image forgeries were presented. In Section 4 the description of copy-move forgery is given. In Section 5 the proposed method is described, while in Section 6 we present the experimental results. At the end in Section 7 conclusion of this paper is given.

2 Literature Review

Digital images are widely used and represent very important part of everyday life. This is a reason of numerous applications that deal with digital images. Some applications are used for character recognition [4], [5], some are used for image enhancement [6]. Multilevel image thresholding is used in many application, so this topic is intensively researched [7], [8], [9], [10]. Determination of optimal threshold values represents hard optimization problem. Swarm intelligence algorithms are one of the most common methods for solving hard optimization problems in the last years. Swarm intelligence algorithms are widely used and searched which resulted with many different proposed algorithms. Some of the most used swarm intelligence algorithms are bat algorithm [11], artificial bee colony [12], [13], [14], firefly algorithm [15], [16], [17], ant colony optimization [18], [19], cuckoo search [20], etc. Compression of images is also important topic when deals with digital images [21].

As mentioned before, digital images can be easily changed and that represents a threat. One of the possible changes is so-called copy-move forgery. One part of image is used to cover or changed another part. Many methods for detecting this kind of forgery are proposed.

In [22] method based on blocks and their features was proposed. Features based methods was used in order to improve the accuracy of detection of fraud. The proposed method uses DCT coefficients and properties of discrete Fourier transformation. Features were compared in order to detect fraud and also register the location of the region in the picture that are counterfeited.

In [1] a method where the image was divided into the round blocks was presented. The properties of round blocks were extracted by using rotating uniform local binary pattern. After that the vectors of properties were compared and rigged regions can be located by following the appropriate blocks. Experimental results show that this method was good not only for detecting the forgery on images with JPEG compression and blur, but also for images with regional exchange and flipping.

A method where the picture is divided into blocks with a fixed size and discrete cosine transform was used was proposed in [23]. Discrete cosine transformation was performed on each block and
DCT coefficients representing each block. Each cosine transform block was represented by a circular block and four properties have been pulled out to reduce the dimensions of each block. Vectors are lexicographically sorted, and duplicate blocks of the image are compared with the threshold value. Experiments show that the proposed scheme in addition to good results in copy-move fraud also have good results in blurring and has low system requirements.

In [24] an approach that was based on the blocks that use texture of blocks as the basis was considered. The aim of [24] was to examine whether the texture is appropriate for the specific application. Tests were conducted on not compressed as well as on the compressed JPEG images.

In [25] a blind forensic approach for detecting copy move scams was described. Used technique was based on the implementation of the discrete wavelet transform on the input image. The goal of this step was to help in reducing the dimensions. Then, the compressed image was divided into overlapping blocks. Criteria for determining similarity of these blocks, to determine duplicate is the phase correlation. Due to the use of discrete wavelet transform, detection is carried out at the lowest level of the image. This process significantly reduces the time of detection [25].

Another approach for copy-move detection was proposed in [26]. In that paper algorithm for precise detection of copy-move scam based on the rotational properties was proposed. Rotational properties were calculated in the image. Techniques dense fields proposed in the literature guarantee better performance with w.r.t but with the price which is reflected in a much greater time for implementation. To avoid this shortcoming patch match algorithm is used that is good for dense fields above pictures. The analysis, which used databases that are available online proves that the proposed technique was correct but also faster than other references dealing with the field [26].

Most of the techniques is not good enough to locate parts of the image where the fraud was made. Poor performance is even more the case when the show in the picture has a lot of great parts that are similar. Passive forensics for image copy-move forgery using a method based on DCT and SVD was proposed in [27]. A method that has a much higher percentage of detection with the images with such parts was described. Each image is divided into blocks of the same size and the DCT is applied to each. The results show that the proposed method can detect fraud in the figures with the same region, even when the image is further compressed.

In paper [28] a cellular automata as a way of detecting the fraud was proposed. The essence of this method is that for each block was known cellular automata rules which are related to the changing values of intensity. A similar method was also described in the paper [29]. Based on these rules, copies of blocks were found. For standard copy-move fraud a cellular automata rule was sufficient, however, if the change of the copied part of the primer block was rotated, cellular automata rules complicate significantly. But cellular automata can be used in such cases if filtering application was done before the start of the detection of the fraud [28].

In [30] Kumar et al. proposed using DCT binary vectors. Kumar et al. presented a method for fraud detection based on the contrast with the help of discrete cosine transform vectors. The image was divided into blocks and for each block DCT coefficients were calculated. Further vectors are created for each block on the basis of DCT components. The proposed method can detect fraud when the contrast of the image is changed [30].

Kirchner et al. in [31] proposed a method of detection of fraud where the picture instead of blocks is divided into triangles. Triangles are matched on the basis of their properties, specifically on the color and vectors. The method is designed to cope with the geometric information. Results are made simultaneously with testing the latest methods for comparing blocks [24]. Kirchner et al. underlined that rectangular blocks for detecting the copy-move forgery is not the only way to divide the image.

### 3 Digital Image Forgery

Detecting digital image forgery represents active research topic. Digital image is considered tampered if the content of the image was intentionally changed for malicious purposes. Image can be changed in two ways, by adding or removing the information. In order to remove information usually there is no need for using another digital image, while in case of adding information it is not the case. Information that is added on the digital image can be some part of the same image or the part of one or more images. Techniques for detecting digital image forgery are different for forgery that involves single image and for forgery that uses more then one digital image.

Deleting part, for example object or person, from the image is one of the most straightforward way to change the content of the digital image. After removing the part, it is necessary to that part fill with some content. This filling methods are opportunity for detecting the forgery. There is some well known
techniques for filling out the empty part of the image. The most simple way is to copy another part of the same image. The copied part should fit into the environment so it be as less as possible visible that is copied part. In order to be less visible or better to say to smooth transition between surround and copied part some image processing operation can be also done. The technique of coping the part of the image can be also used to duplicate some object as it is shown in Fig. 1. In literature, this type of forgery is known as copy-move forgery.

![Figure 1: Example of copy-move forgery](image1.png)

Another technique used for filling the space left after removing the object is the technique inspired with process of real painting restoration. This method is called in-painting technique. Idea is to fill the space through some number of iteration. At the beginning, the edges of the empty space are filling and the process of filling advances toward the center of the space. This method can provide good results in the case of simple texture of surround, but in the case of highly textured area, in-painting method will not give good results.

One of the latest filling technique is the seam carving method [32]. This algorithm was initially used for content-aware image resizing [33]. This algorithm is based on the concept of seam, where seam is consider sa a monotonic and connected path of pixels. This path include one pixel per row or column. The image is processed from the top to the bottom or in case when one pixel per column is included, left to the right. Seams are removed iteratively. Every seam remove matches to a horizontal or vertical resizing of one pixel. The final image is better than the one obtained by re-sampling. If the algorithm is limited to just one region of the image it can be successfully used for covering the removal forgery.

Beside adding and removing information from digital image, changing the content by some image processing operations such as histogram manipulation and contrast enhancement sometimes is also consider as image forgery. Moreover, brightness adjustment and scaling are often used for masking copy-move forgeries. Simple geometric transforms and filters can be used for compromising the forensics analysis. Applications like this have an aim to cover or delete traces of the forgery.

Second type of image forgeries is one that include more than one image. Combining different images and creating false visual message represents one of the most powerful tools for deceiving people. Unfortunately, today with all available techniques and softwares this is also very easy task. Fig. 2 shows one of the famous image forgery. This type of forgery is called splicing image forgery. In the case of this kind of tamper some blending and matting methods are also used for covering the boundaries of the spliced parts. On this way image will have more uniform look. In order to spliced parts or objects match the perspective of the original image, some geometrical transformation might need to be applied. By applying rotation, translation or scaling the spliced parts can be adjusted to look like they belong to one image.

![Figure 2: Example of splicing forgery](image2.png)

In some cases, inserted object does not come from natural digital image. It can be artificial, computer generated. Every improvements in technology can be used in illegal purposes, which is the case also with the computer graphics. Nowadays, numerous applications for modeling and rendering realistic three
dimensional objects exist and these objects can be used as material for splicing into digital image. One of the effective methods for detecting splicing forgery include shadow analysis. In recent years many papers explore methods for shadow analysis with aim to detect if the image was modified and result of splicing forgery or it is original image [34], [35], [36].

4 Copy-Move Forgery

Advanced cameras along with software tools allow anyone to easily modify the image. Community professionals dealing with photography began to try to prevent easy manipulation of images with digital signatures, but with no greater success. However the downside of technique with signature is that image needs to be preserved before it is printed because of the signature entered. Therefore, the application of signature are very limited. This scam can be detected in the following manner as in the work [37].

Copy-move forgery is an cunning way of changing the image in the sense that part of the image is copied and placed in another part of the image to hide the specific object. Because the copied part of the picture and all of its properties such as texture match the rest of the picture, it is very difficult to detect this kind of fraud. However, every day a new combination of methods appears that facilitate finding the fraud. Thus, the combination of methods dyadic wavelet transform (DyWT) and scale invariant feature transform (SIFT) gives good results as presented in the paper [38].

Copy move forgery is predominantly used for changing the content of pictures, usually to the contours of the area the one in the picture and replaced with some other element which is in the same picture. The simplest way to solve this problem is to use the extensive search, which means that the original image is compared with a cyclic version of the image. This technique requires a large hardware power. This method with blocks is quite prevalent method. The image is divided into blocks that overlap. Then, on the basis of common characteristics blocks that face one another are recognized. The analysis of the blocks is carried out and only those pairs of blocks that are the same distance are recorded. There are different methods to investigate copy-move fraud. One of them is the discrete cosine transform (DCT). The virtue of this method is that the signal energy is reserved for a few coefficients, and the compression operation, for example, should not affect the result greatly. DCT does not work, if the geometric transformation is applied to the duplicated regions. Principal component analysis (PCA) can also be used to display different blocks. This method has proven to be resistant to compression, but rotation can affect the final results. Also, this method can be investigated in a slightly different way [39].

4.1 Features

The basic step in the techniques of protection from copy-move scam is to divide images on the blocks. The next step is determining the properties of the original blocks to be able to compare them with potential copies to detect the slightest modification. Different properties or features of blocks were used in recent research.

Already mentioned discrete cosine transform (DCT) coefficient can be used. DCT is used because the signal power located on the first few coefficients, while others are smaller. As a result, operations such as the compression will not affect adversely on the first coefficients.

The next method is the principal component analysis. In this method the coefficients of the blocks in the matrix and calculates the appropriate matrix. There is a new base thanks to eigenvectors of matrix. In order to reduce the dimensions, blocks that go to basic vectors with higher values are used. These results are resistant to compression, however, the rotation would affect the final results.

The next method proposed is discrete wavelet transform (DWT), which divides the image into four smaller. It divides lower frequency components in the blocks that overlap in order to reduce their number and speed up the process of finding the same regions. In these regions singular value decomposition (SVD) are applied. Because the SVD and PCA are similar, SVD will behave in certain situations as well as PCA.

Another way that circumvents the methods used in the compression ratio is to use a method that is based on the colors of the blocks. These colors include red, green and blue. Another set of blocks is divided into two parts in 4 continents. The intensity of a block in relation to the total number of blocks is calculated. Experiments have shown that this method can well cope with JPEG compression, or with Gaussian blurring.

It can be also applied to the method of Fourier Mellin transform (FMT) on the blocks. First, find the value of Fourier Mellin transform for each block and then the results are transferred to coordinate. The vector can be calculated based on the polar values and these values are used for the properties. The tests proved that this method carries well with excellent compression ratio. There are also other methods that are listed in the paper [40]. Method that was presented allows making it easier to find the location
of deception. An interesting method was explained in the paper [26], where method uses a densely compacted neighboring fields. Method facilitates easier detection of fraud in the picture.

5 Our Proposed Method

As already mentioned copy-move detection is usually based on overlapping blocks. In [41] a similar method was proposed. After obtaining the vectors of features that represent the blocks, vectors are sorted lexicographically in order to facilitate detection. Blocks whose vectors are similar and close can be easily detected by shift vector.

Our proposed algorithm uses reduced vector of features and bin-sort algorithm instead of RADIX sort which was used in the references paper [41]. They proposed nine component feature vector. Image was divided into blocks of size 16×16 and each of these blocks were represented with the 9 characteristics. These characteristics were as follows. First, the average value of the intensity of the block was calculated. Then, each block was divided into four identical subblocks and the remaining eight characteristics concerning the relationship between block and subblock. Four characteristics mark the ratio of the average intensity of each of the washers block and subblock. Four characteristics represent the blocks, vectors are sorted lexicographically in order to facilitate detection.

Thus obtained characteristics are normalized to the rank of 0 to 255. This is done by the following formula:

\[
 f_i = \begin{cases} 
 f_i = \text{Ave}(B) & \text{if } i = 1, \\
 \text{Ave}(S_{i-1}) \div (4\text{Ave}(B) + \varepsilon_1) & \text{if } 2 \leq i \leq 5, \\
 f_i = \text{Ave}(S_{i-5}) - \text{Ave}(B) & \text{if } 6 \leq i \leq 9.
\end{cases}
\]

However, our experiments show that it is possible to detect forged regions using only three simple features, basic square 16×16, upper left corner subsquare size 4×4 and upper triangle, above and including main diagonal.

For each of this regions average intensity value is calculated and these quantities do not have to be normalized because they are already in the interval 0 to 255. With this three components, each in the range from 0 to 255, it is possible to have 256³ different feature vectors. For this number it is possible to use bin-sort algorithm since responding data structure is easily handled by todays computers. We can define an array of 16.7 million elements or we can use some kind of lists. In any case, for each pixel that belongs to the image of N×N size, we generate a vector and sign it to the corresponding place in the bin-sort vector together with \((x, y)\) coordinate of upper left corner of that square. In the same way we increase associated counter with each element of the bin-sort array. When features vectors are calculated and assigned for the whole image, we linearly go through bin-sort vector. Theoretically, at many points there will be no associated pixel, at some points there will be only one pixel and at some points there will be two pixels associated. Those places where two pixels from the image are associated with the same point in the bin-sort vector correspond to points with the same feature vectors so they should correspond to the places that belong to possible forgery areas. It is possible by the chance (since the set of used features is rather limited) to have the same feature vector values even though if they do not belong to forged areas, but such places will be isolated. In the reference paper [41], they use shift vector and corresponding counter to determine forged areas. However, our experiment showed that it is enough if we mark on a black image matrix points that have the same feature vector. After that apply the median filter for removing speckle noise which will result in removing those isolated points.

This approach depends on the number of selected features and the range for each feature. It can be extended to a larger number of features only by reducing the range for each feature.
6 Experimental Results

Experiments were performed on computer with following performances: Intel® Core™ i7-3770K CPU at 4GHz, 8GB RAM, Windows 10 Professional OS. Proposed method was implemented in Matlab version R2015a.

In this paper we used the database for a copy-moved forgery detection proposed in [42]. The dataset consists of 260 forged image sets. Every image set includes forged image, two masks and original image. Images are grouped in 5 categories according to applied manipulation: translation, rotation, scaling, combination and distortion. Also, post-processing methods, such as JPEG compression, blurring, noise adding, color reduction etc., are applied at all forged and original images. Examples of dataset are shown in Fig. 3.

![Figure 3: Example of dataset](image)

In Fig. 4, Fig. 5, Fig. 6, Fig. 7, Fig. ... and Fig. 8 the result of our proposed algorithm are shown. Fig. 4(a) is a picture of nature where the flower is copied. Copied parts are shown in Fig. 4(b). Our algorithm detected copied figures as it can been seen in Fig 4(c). Proposed algorithm was able to recognize copied regions. Fig. 5 represents second example where the car is copied. Similar to Fig. 4, Fig. 5(a) is an image that contain copy-move forgery, while in Fig. 5(b) copied regions are shown. Recognition of proposed algorithm is represented in Fig. 5(c). In this case our proposed algorithm was also shown as good.

![Figure 4: Experimental results](image)
Figure 5: (a) Original, (b) Mask, (c) Recognized regions

Figure 6: (a) Original, (b) Mask, (c) Recognized regions
Figure 7: (a) Original, (b) Mask, (c) Recognized regions

Figure 8: (a) Original, (b) Mask, (c) Recognized regions
The third example shows the copied pigeon and it is presented in Fig. 6. It can be seen that our proposed algorithm successfully detects image manipulation. Similar to the previous examples, experimental results of our proposed algorithm are shown in Fig. 7(c) and Fig. 8(c). Our proposed algorithm that uses only three simple features was able to successfully detect forged regions.

7 Conclusion

In this paper we proposed an algorithm for copy-move forgery detection. Proposed method is based on block analysis. It divides an image into blocks of the size $16 \times 16$ using the vector with three elements which are determined from that square, smaller block size $4 \times 4$ from upper left corner of the block and triangle above and including the diagonal. Proposed method was tested on standard benchmark images from [42]. In all cases it successfully detected copy-move forgery. One of the main limitation of the proposed method is the number of the features that can be used. It is possible to modify the algorithm so that we do not require that two points have exactly the same feature vector but we allow some tolerance component by component. Also it is possible to use different data structures rather then an array that would allow a larger number of features to be used.

Acknowledgement: This research is supported by the Ministry of Education, Science and Technological Development of Republic of Serbia, Grant No. III-44006.

References:


