COPY-MOVE FORGERY DETECTION TECHNIQUE WITH AUTOMATIC THRESHOLD DETERMINATION

Malti Puri
Research Scholar Dept. of CSE,
DAV Institute of Engineering & Technology,
Jalandhar, India.

Dr. Vinay Chopra
Assistant Professor Dept. of CSE,
DAV Institute of Engineering & Technology,
Jalandhar, India.

Abstract—Image authenticity is a major concern these days. Copy-move forgery is a very common type of digital image forgery. There are basically two methods for the detection of copy-move forgery. One is block-based method and another one is keypoint-based method. In this paper we study a method that automatically calculates threshold to detect copy-move forgery. It is a block-based method which uses DCT-phase to extract features from input image.

Keywords—Image forgery, copy-move forgery, CMFD, block-based, DCT.

Introduction
In present era we have very powerful image capturing tools which are inexpensive and very easily available. Moreover, almost everyone has smart phones which have very high quality cameras. So, there is wide amount of digital images all over the world. Images are used in many fields to convey information. Digital images have become integral part of every field. There are very easy to use and inexpensive image editing tools with the help of which it is very easy to modify images. Digital images are very easy to manipulate by any non-expert person. It is very easy to add or remove important features from an image. Any image modification or manipulation can become a forgery, if after modifying semantic of original image changes [2]. If after image manipulation semantic of an image does not get changed then we cannot call it a forgery. A forger can create a forgery to cover objects of an image, to make the image more pleasant for appearance or to hide something from image. Sometimes forger create forgery just for the fun or to showcase his skills. Whereas some forger may create forged image to do some kind of fraud. Various operations used while creating forged images are copy, move, paste, selection, resize, crop, rotate, scale, digital filters etc. Different kinds of forgeries can be created using these simple operations. Some main types of forgeries that can be created using these simple operations are Image retouching, Image enhancing, Image splicing, Image morphing and Copy-move. Image Retouching is done to reduce certain feature of an image and enhances the image quality to capture the reader's attention. [11]. Image enhancing is done to enhance an image with the help of Photoshop operation such as saturation, blur and tone etc. These enhancements don’t affect image meaning or appearance but effects the interpretation of an image. In image splicing different parts or elements from multiple images are pasted into a single image. At last, one image is obtained that is having content from different images. Image morphing can be defined as a digital technique that gradually transforms one image into another. Transformations are done using smooth transition between two different images [11]. Copy-
move Forgery is done by pasting content from an image to another part of the same image.

Copy-move forgery is the type of forgery in which one region is copied from an image and pasted onto another region of the same image. Therefore, source and the destination both are same [9, 11]. Operations used in copy-move forgery are: copy and paste. Since the copied area belongs to the same image, the properties of copied area like the color palette, noise components, dynamic range and the other properties too will be compatible with the rest of the image [12]. So, the human eye has difficulty in detecting copy-move forgeries. Also, forger may have used some sort of retouch or resample tools to the copied area so that it becomes even more difficult to detect copy-move forgery. Retouching involves compressing the copied area, adding the noise to the copied area etc. and resampling may include scaling or rotating the image.

Digital images have become an integral part of almost every area. So, image authenticity is a major concern [3]. Digital images play a very important role in areas like forensic investigation, intelligence services, medical imaging, journalism etc. Medical images are produced in most of the cases as proof for unhealthiness and claim of disease. In courtrooms digital images are used as evidence and proofs against various crimes. In e-commerce sites images are used to display the item to be sold. But the basic requirement to believe what we see is that the images should be authentic [3]. Digital image forgery detection techniques are mainly classified into two categories: one is active approach and another one is passive approach [2, 10]. Active approach requires a preprocessing step and suggests embedding of watermarks or digital signatures to images. Therefore require knowledge original image. Thus it limits their operation. Algorithm is used to embed the watermark. Any manipulation of the image will affect the watermark and examination of watermark’s condition will show if tampering has occurred. Whereas, in case of passive approach forgery detection, there is no requirement of knowledge of original image and does not rely on presence of Digital watermark. The passive approach is regarded as evolutionary developments in the area of tamper detection [11]. Passive techniques are also known as blind techniques because it uses the received image only for assessing its authenticity.

For the first time when a system was designed by Jessica Fridrich et al. to detect copy-move forgery, Discrete Cosine Transform (DCT) was used in the feature extraction step to extract features from an image. DCT transforms an image from spatial domain to frequency domain. Less space is required to represent the image features if it is transformed to DCT. It also removes redundancy between neighboring pixels.

In the system explained in this paper DCT-phase will be used instead of. DCT-phase is used to reduce the size of the image. For an image I of size N×M DCT phase is represented as I'_{DCT}. DCT-phase basically consists of signs corresponding to DCT values. It contains substantial edge information about the test image. DCT phase can take values from {-1, 0, 1} [3].

**Block-based Forgery Detection Methods**

Block based method is one of the two methods to detect copy-move forgery. It works on pixel level and gives detailed information about copied pixels. Image is firstly divided into several overlapping blocks. This process is also known as block tiling. For an image of size M×N and block size of b×b, the number of overlapped blocks is given by (M-b+1) × (N-b+1). After block tiling features are extracted from each block. Features like Local Binary Pattern (LBP), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Principle Component Analysis
(PCA), Zernike moments etc. are used. After feature extraction matching is done to detect forgery. As source and destination image is same in case of copy-move forgery, so similarity between different blocks help as a clue of copied blocks. Matching is done to detect the duplicated regions. High similarity between two feature vectors is interpreted as duplicated region. Methods used for matching can be Euclidian distance, correlation, Best-Bin-First search etc. [10]. Values from matching are compared with a threshold value. If its value exceed threshold value then forgery is detected and marked. Block-based methods are more accurate, at the same time takes more time and more computational load for processing. Works well in case of pure translation and also in case of complex scenes. It can detect multiple copied regions with high accuracy. There are a number of block-based methods to detect copy-move forgery since 2003.

Jessica Fridrich et.al [2003] investigated the problem of detecting the copy-move forgery and describes a reliable copy-move forgery detection method. The method can successfully detect the forged part even when the copied area is enhanced and when the forged image is saved in a lossy format, such as JPEG [7].

Babak Mahdian et.al [2007] proposed a method based on blur moment invariants. Image is firstly divided into overlapping blocks and these blocks are represented using blur invariants. Experimental results shows its robustness against various changes like blur degradation, additional noise [2].

M. K. Bashar et. al [2007], proposed a DWT based method to detect copy-move forgery. Multi-resolution wavelet decomposition is applied to blocks. Then, Normalized wavelet coefficients are stacked into a vector. Lexicographically sorting is done. And at last and then forgery is detected [9].

Sevinc Bayram et. al [2009] proposed a system based on FMT. Experimental results show that the proposed method can detect duplicated areas in the images even when various manipulations have been done to the copied area [11].

Zhang Ting, et. al [2009] proposed a method based on SVD for detecting copy-move forgery. It works by first extracting singular value SV features and then it is matched to its nearest neighbors in image. Matching is done using of k-d tree searching method. Experimental results shows that the proposed algorithm has low computational complexity [13].

Seung-Jin Ryu et. al [2010] proposed a detection method of copy-move forgery using Zernike moments. The proposed method can detect a forged region even if it is rotated [10].

Leida Li et.al [2013] proposed a method for detecting the copy-move forgery based on LBP. The image is first filtered and divided into overlapping circular blocks. Then the features of the circular blocks are extracted using LBP. Experimental results illustrate that the method is robust against JPEG compression, noise contamination, blurring [8].

Guzin Ulutas et.al [2013] proposed a system based on Color Coherence Vector (CCV) to detect copy-move forgery. CCV designates coherent pixels in images and use spatial relationship in color information. Experiments show that the method can detect forged regions even if the image is processed by Gaussian Blurring to hide forgery [6].

**PROPOSED SYSTEM**

It is a DCT-phased based method that can calculate threshold value using Benford’s law. This method works by applying DCT and quantization to the blocks of images [3]. Steps involved to detect a copy move forgery using this method are pre-processing, Block-tiling, Feature Extraction, threshold calculation,
Similarity search and marking the forged regions. All steps are explained below [3]:

A. Pre-processing:
Image is firstly converted into YCbCr color space. Y is luma component and is in greyscale, Cb is Blue-difference chroma components and Cr is red-difference chroma components.

B. Block-Tiling:
After converting image into YCbCr color space, block-tiling is done, Image is divided into overlapping blocks. Block size is taken to be 8x8. Each block of an image of size NxM is denoted as B'_i. Where i=1, 2…(N-7) (M-7).

C. Feature Extration:
Feature is extraction is done using DCT-phase, but requires some steps to follow one by one. Fig. 3.1. Represents these steps and these steps are discussed here also. First of all every block is transformed into frequency domain using DCT and quantization is done. Blocks will be represented as

B'_i=DCT(B'_i)/qt.

Now, corresponding to each block, sign blocks are obtained using eq. (1), (2) & (3).

\[ \forall B'_{ijk} > 0 \rightarrow S_{ijk} = 1 \] ........................ (1)
\[ \forall B'_{ijk} < 0 \rightarrow S_{ijk} = -1 \] ........................ (2)
\[ \forall B'_{ijk} = 0 \rightarrow S_{ijk} = 0 \] ........................ (3)

Once the sign blocks are generated, zigzag scanning is done. Elements having low frequency will be extracted to create the corresponding feature vectors. Number of these feature elements is predetermined and is 16. First 16 elements are selected and these elements are lower in frequency. Last three feature element are computed using YCbCr. Average value of Y, Cb and Cr is inserted at the end of 16 elements. So, at the end we got a 19 element feature set for each block. At last Feature vectors are lexicographically sorted and after sorting these feature vectors are stored in a matrix.

Input: Overlapped blocks

DCT and Quantization

Create the sign block

Zigzag scan the sign block

Feature vector generation using first 16 elements and YCbCr channel

Lexicographic sorting

Output: Feature vectors

Fig. 3.1 Steps of feature extraction [3]

Threshold Calculation:
Threshold value is calculated using quality factor and value of quality factor depends on whether input image has been compressed or not. Benford’s law is used to check whether input image has been compressed or not.

Benford’s law states that the probability distribution of the first digits in a set of natural numbers is logarithmic. According to Benford’s Law a data set will satisfy the following criterion [3]:

\[ p(x) = \log_{10}(1+1/x) \] .......................... (4)

Where x=1, 2…9 and p(x) is the probability of x.

In case input image is JPEG compressed, the distribution of the first digits of the DCT coefficients will satisfy the following criterion [3]:

\[ p(x) = N \log_{10}(1 + \frac{1}{s+q}) \] ............... (5)

Where x=1…9, N is normalization factor, s and q are model parameters.
For example, the input image is compressed with a quality factor of 100. If it follows the Benford’s Law, this is the first compression of the input image. If it doesn’t follow then input image has been compressed. Algorithm to calculate quality factor for compressed image is as follows [3]:

**Algorithm 1. Quality factor determination.**

**Input:** Test image I  
**Output:** Quality Factor q

- % Compress (a, b)→ a image file is JPEG compressed with quality factor b  
- % Ssfit_Curve (x, a) → First digit distribution of the JPEG coefficients of the a is fitted with Benford’s generalized law and the function returns the value of SSE

\[
j = 1; x = [1 \ldots 9]; \text{min} \_sse = 999;
\]

for \( qf = 30:5:100 \)

\[
I\_compressed[j] = \text{Compress}(I, qf);
\]

\[
j = j + 1;
\]

end

\[
j\_\text{end} = j; qf = 30;
\]

for \( j = 1: j\_\text{end} \)

\[
\text{temp} = \text{Ssfit\_Curve}(x, I\_\text{compressed}[j]);
\]

if \( \text{temp} < \text{min} \_sse \)

\[
\text{min} \_sse = \text{temp};
q = qf;
\]

end if

\[
qf = qf + 5;
\]

end

return q;

We will get a quality factor q at the end of Algorithm 1 and this value q will be used to calculate threshold (ts) using equation (6), (7) & (8) [3].

\[
\begin{align*}
    & \forall (q \leq 30) \rightarrow \text{ts} = 12/16 \\
    & \forall (q > 30) \land (q < 70) \rightarrow \text{ts} = 13/16 \\
    & \forall (q \geq 70) \rightarrow \text{ts} = 14/16
\end{align*}
\]

Threshold value (ts)

Fig. 3.2 Steps of Threshold calculation

**Similarity search:**

Similarity search is done by comparing each feature vector with another. The value compared with threshold value \( t_s \). If similarity between two feature vectors is greater than \( t_s \) and distance between same blocks is greater than \( t_d \), shift vector will be calculated and inserted into a list [3]. Value of shift vector (sh) is computed using equation (9) [3].

\[
\left( \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \geq t_d \land (r^i \geq t_s) \right) \rightarrow \\
sh_m = \{x_i, y_i, x_j, y_j, |x_i - x_j|, |y_i - y_j|\} \quad \ldots (9)
\]

**Marking the detected Regions:**

If count of shift vectors that were calculated in previous step exceeds \( t_{shift} \), then those blocks will be marked as forged regions [3].

**Results**

Fig.4.1(a)Input image[3]  
Fig.4.1(b)Experimental result [3]
Fig. 4.2 (a) Input image [3]  
Fig. 4.2(b) Experimental Result [3]

Fig. 4.1 (a) and 4.2(a) shows the forged images. Forged areas are indicated using blue line boundary. Fig. 4.1(b) and 4.2(b) shows the output that comes at the end of detection algorithm. Experimental results have shown that method can detect even multiple copied regions in an image.

**Conclusion**

In today’s world due to presence of low-cost, easy to use and high resolution photo capturing/editing tools it has become very easy to create digital image forgeries. Copy-move forgery is one of the most common digital forgeries. A lot of research has been done to detect copy-move forgery. This paper describes a system that can compute threshold value automatically. Threshold is the value that is used to detect similarity between different areas of the input image. Similarity is important to check because similarity between different elements of the image is the clue to detect forgery in a copy-move forged image. Experimental results have shown that this method gives high accuracy and less false negative results.

**References**


