A Study to Ascertain and Differentiate between Genuine and Transplanted Documents/Signatures

Manisha Mann*, Shukla SK and Shruti Gupta

Amity Institute of Forensic Sciences, Amity University, Noida, UP-201303, India

Abstract

Digital forgery is very common these days because the use of digital document is increasing day by day and readily available software's for manipulating the documents. It is very necessary to authenticate the integrity of digital document, whether it's genuine or not. Various secure detection systems are made which involves use of algorithm, but this paper deals with examining simple features present in image/document which can be used to check the authenticity of an electronic document, whether they are genuine or forged. These simple features are like Color variation, Font size difference, Pixelate resolution, Range of magnification-low, medium and high range of magnification and extracting hash value-MD4, MD5, SHA-1, RIPEMD-128 and RIPEMD-160. Results are highly reliable.

Keywords: Digital documents; Digital forgery; Image hashing; Pixels

Introduction

Now a days use of electronic documents have been increased tremendously which results in increase in digital forgery. A forgery is an unlawful act of forging a document or item for the reasons of misrepresentation or trickery. Digital forgery is very common now, because digital images are not difficult to manipulate and alter due to easily available image processing and editing software’s. These days, it is feasible to include or exclude any important characteristics from a picture without leaving any conspicuous hints of altering. Digital forensics is the current topic which has received attention recently. Digital images play an important role in depicting and transferring the data easily, therefore new techniques for detection of forgery in digital images have been investigated. There is a little difference between image forgery and digital image forgery, as the digital image forgery deals with the digital image as compared to photographs which are used in image forgery. There are many different computer graphic editing software are available like Adobe Photoshop, GIMP (GNU Image Manipulation Program), and Corel Paint Shop, etc.

Areas like legal, criminal, journalism, medical requires the digital document to be authentic. So there is high demand for a dependable, safe and secure detection system, which is capable to determine or check whether the digital image/document is real or altered. As due to presence of easily available image editing software’s alteration could be done to the digital document and some modification may be impossible to be seen by human eye, these modification results in some core statistics changes in the digital document which can be detected.

Forgery techniques in digital images are classified into three main groups

- Copy Paste Forgery (Image splicing).
- Image retouching.
- Copy Move Forgery (Image cloning).

Copy paste forgery also known as image splicing: In this type an altered duplicate copy or a document is prepared with the help of an original image along with some additional images, for instance including particular area of the additional image to the original one, just to hide or manipulate the image.

Image retouching: In this type the forger manipulates the image in a way so that the modification in the content of the image becomes unnoticeable.

Copy move forgery also known as Image cloning: In this type of forgery a distinct part of an image is copied and moved to another part of the same image.

Usually there are two types of detection techniques or approaches are used. Following are the two techniques

Active method includes features like watermarking which helps in detecting digital tampering like name, signature, date, etc.

Passive method in this method digital image forgeries are detected without taking any use of the features or information of the original image [1].

Image hashing/Image fingerprinting

Image hashing or image fingerprinting is a procedure of providing value that is specific to particular image by examining its content. This fingerprint is basically a string which is assimilated with other fingerprint for possible matches [2]. There are two types of image hashing techniques:

Perceptual hashing: Perceptual hashing a technique which produces fingerprint of different multimedia like audio, video or image file with the help of algorithm. Perceptual hash functions (PHF) are most commonly used in area of digital forensics and protection against copyright infringement. Through PHF one can assimilate and map source the data with the help of correlation between the hashes. For
instance, Wikipedia contains a database of hashes of online articles and books for which copyright is held by the authors, when any Wikipedia users upload any document containing data of online books or article registered in that database will have same hash value and will match with each other and this can be used in flagging or pointing out plagiarism [3].

**Cryptographic hashing:** Cryptography is a technique in which a message authentication code (MAC) is produced with the help of hash function. Even a slight change in input message results in distinct hash value. Following are mostly used hash algorithm:

- MD (Message Digest Algorithm) - MD4, MD5.
- SHA (Secure Hash Algorithm) – SHA-0, SHA-1.
- RIPEMD (RACE Integrity Primitive Evaluation) – RIPEMD-160 [4].

**Pixels:** Pixel is derived from a word "picture element". In a computer picture, a pixel is the simplest unit of programmable color. The size of the pixel relies on the resolution of the display screen. If the display is at its highest resolution, the physical size of pixel will be equal to dot pitch of display. If the resolution is less than the highest, then size of pixel will be larger than dot pitch [5].

Early studies on this topic were Weihai li; et al. developed a method for detection of copy paste forgery in manipulated JPEG pictures and also locates the position or the area of manipulation. This method works by extracting DCT block artefact grid and determining mismatch of grid [6]. Thirumagal, et al. proposed a forensic technique for detection of contrast enhancement (globally or locally applied) and by identifying the peculiarities of intrinsic fingerprint the histogram equalization in a picture can be detected [7]. Najah Muhammad, et al. proposed an effective non-intrusive technique for detection of copy move forgery. In this technique the image is segmented and the similarity is detected with the help of Dyadic Wavelet Transform (Dy WT) [8]. R. Venkatesan, et al. proposed an image indexing technique which is known as image hash function. Randomized signal processing is used by algorithm for a non-reversible compression of an image which results into arbitrary binary strings [9]. Kelsy Ramirez-Gutierrez, et al. proposed two algorithm to detect authenticity of an image, even if the image is affected by distortion like filtering, compression and other malevolent modification like geometric distortion. The algorithm can also detect tampering and also the localized tampered areas [10].

This paper deals with examining simple features present in image/document which can be used to check the authenticity of an electronic document, whether they are genuine or forged. These simple features are like Color variation, Font size difference, Pixelate resolution, Range of magnification: low, medium and high range of magnification and extracting hash value: MD4, MD5, SHA-1, RIPEMD-128 and RIPEMD-160.

**Material and Methods**

**Sample size**

25 samples of documents were created which contained signatures, dates, names and addresses, which have been transplanted from the originals on those documents. Doctor's prescription, list of student selected in any institution, stamp papers, certificates and appointment letters are the type of samples.

**Sample collection**

The 25 samples were collected from Google Images.

**Procedure adopted for analysis**

As original disputed documents cannot be gathered due to their authorization and confidentiality, which should be maintained by government forensic laboratories with due reason, such documents were prepared manually for the research with the help of software's. These samples were then analyzed in soft copy format. Then the signatures were cropped from the originals, copied and pasted on the documents to be forged with the help of MS Paint. These forged documents were then examined in soft copy format with the help of Picasa. Following are the features on the basis of the samples were examined:

- Colour variation.
- Font size difference.
- Pixelate resolution.
- Range of magnification: low, medium and high range of magnification.
- Hash value: MD4, MD5, SHA-1, RIPEMD-128 and RIPEMD-160.

**Software's and application used for analysis**

- Microsoft Paint or MS Paint is a simple graphics program that has been included with all versions of Microsoft Windows. This program can be in colour mode or two-colour, black-and-white, but there is no grayscale mode.
- Picasa 3.9 is an image organizer and image viewer for organizing and editing digital photos plus an integrated photo-sharing website, originally created by a company named Lifescape in 2002 and owned by Google since 2004.
- Fileformat.info is a website which reveals the all types of hash value of any type of file like audio, video or any document. Basically it's an online hash value calculator.

**Samples (observed features)**

Since the number of samples created and observed for this study is very large, it would not be convenient to attach all the samples and pictures of the features observed in each sample. Therefore, only a few samples are being attached with the zoomed in images of their parts showing some difference from the whole of the document, indicating different sources of origin (Figures 1-4).

**Result and Discussion**

The following 25 samples were examined on basis of different features in Picasa Software:

- Range of magnification: low, medium and high range of magnification.
- Colour variation.
- Font size difference.
- Pixelate resolution.

Above mentioned features are given in Table 1.

**Range of magnification**

All the samples were first analyzed by magnifying them. The range
| Features/Samples          | Sample 1 | Sample 2 | Sample 3 | Sample 4 |
|--------------------------|----------|----------|----------|----------|
| Color Difference         | Present  | Present  | Present  | Present  |
| Font Difference          | -        | -        | -        | -        |
| Pixelate Difference      | Present  | Present  | Present  | Present  |
| Low-range magnification (%) | 105%    | 152%    | 231%    | 26%      |
| Mid-range magnification (%) | 160%    | 266%    | 352%    | 40%      |
| High-range magnification (%) | 369%    | 465%    | 535%    | 106%     |

| Features/Samples          | Sample 5 | Sample 6 | Sample 7 | Sample 8 |
|--------------------------|----------|----------|----------|----------|
| Color Difference         | Present  | Present  | Present  | Present  |
| Font Difference          | -        | -        | Present  | -        |
### Pixelate Difference

| Magnification (%)      | Sample 21 | Sample 22 | Sample 23 | Sample 24 |
|------------------------|-----------|-----------|-----------|-----------|
| Low-range              | Present   | Present   | Present   | Present   |
| Mid-range              | Present   | Present   | Present   | Present   |
| High-range             | Present   | Present   | Present   | Present   |

### Features/Samples

| Sample | Color Difference | Font Difference | Pixelate Difference |
|--------|------------------|------------------|---------------------|
| Sample 21 | Present | Present | Present |
| Sample 22 | Present | Present | Present |
| Sample 23 | Present | Present | Present |
| Sample 24 | Present | Present | Present |
| Sample 25 | Present | | |

### Magnification Levels

- **Low-range magnification (%)**
  - Sample 21: 132%
  - Sample 22: 45%
  - Sample 23: 34%
  - Sample 24: 119%
  - Sample 25: 157%

- **Mid-range magnification (%)**
  - Sample 21: 175%
  - Sample 22: 68%
  - Sample 23: 59%
  - Sample 24: 208%

- **High-range magnification (%)**
  - Sample 21: 465%
  - Sample 22: 181%
  - Sample 23: 157%
  - Sample 25: 419%
Table 1: Shows the observation made on examination of samples with transplanted signatures on basis of Range of magnification, Color variation, Font size difference and Pixelate resolution.

| Sample | Original Sample | Simulated Sample |
|--------|----------------|------------------|
| MD4    | 34e60c92e5cc5b80ff0c86d0f636cc9 | MD4 | 195522f2158896931bcc957710df7f9da |
| MD5    | 547626b3ad63062fa47c0b63ea4bd8f | MD5 | cc0399c13d221a29634606b79a30068 |
| RipeMD128 | 2fc0a2e98719f0b1823d9e079daa8f | RipeMD128 | caa0ab9adeb53aece753062f5970729d0 |
| RipeMD160 | a443a85d7f3865a8ca87abcf156070f96d83f8 | RipeMD160 | 906b54d66d1d6f5863e915be4c4f17306785 |
| SHA-1  | c22c1e103c9737651c4c12b5f651d30174d4f09 | SHA-1 | 6027cc87585551cd0134a1ce6ef4a4fc7e259125 |
| MD4    | c9f910c73a6a67bca05fa3324 | MD4 | 69f7abece148018daa686be4b707658e |
| MD5    | 62723c5e2b9f7ca4c4f823744a | MD5 | f3a4d3e0d1bb34e5b2a332b5d69 |
| RipeMD128 | f1b04d6b145b980ade9470126f249 | RipeMD128 | ca295b7a0bfe6db710adac4f280978f |
| RipeMD160 | 190eb7e274303777c716b49519bd3f4646cb79 | RipeMD160 | 4f53ed926262a9d2dab58c33416385b003ba |
| SHA-1  | 59313a9a1e2f889d977d47966e22777d2f797750 | SHA-1 | 5200700ab4963b9767562a1d332bc8e10bbed |
| MD4    | 8585091a42ccdf590615436e0287e1 | MD4 | b0d64247495690a2a945e52a44c57 |
| MD5    | 1184a3e186b001930355aba08135b194 | MD5 | e95238ed12f4b4e67675f26b0e5969 |
| RipeMD128 | 029e1c3da8b699a16cf84e364e2c3 | RipeMD128 | 625eb0d620704900c533ba4cd53c5c |
| RipeMD160 | e25186f2b0459f118b0f938a3e3ad98f5aee63 | RipeMD160 | c5c7af9656cd74e4ffab113c4d4113685b545 |
| SHA-1  | eaaf2ca6a054be2d07910e7d70a30af6999db | SHA-1 | 033818b2025b37e7af29b7afe2b4f49b392771746 |
| MD4    | e48f60718a66e142107a1955466148a | MD4 | bcd64247495690a2a945e52a44c57 |
| MD5    | 8ed4c475204c71e71f9eac985095565f9 | MD5 | e95238ed12f4b4e67675f26b0e5969 |
| RipeMD128 | 0e2f01f95888b4f467c36536c135c3fc | RipeMD128 | 625eb0d620704900c533ba4cd53c5c |
| RipeMD160 | e074d0e8b80dcb3c1d57ca7431585a533817 | RipeMD160 | c5c7af9656cd74e4ffab113c4d4113685b545 |
| SHA-1  | 4311a59f99f976147b162f669adbed2c45785346 | SHA-1 | 033818b2025b37e7af29b7afe2b4f49b392771746 |
at which transplantation was first observed was considered as lowest range of magnification for the specific sample. Simultaneously the range of magnification was increased and it was observed that on increase of magnification the transplantation was more prominently seen. All the samples were analyzed between 26 to 1252 ranges of magnification.

### Color variation and font size difference

All the samples were analyzed to check color variation in two aspects: color variation in paper background and color variation in ink. Difference in the font size of the samples was examined and observations were noted down in the Table 2. The samples which were transplanted had noticeable difference in their background/ink color as well as font size.

### Pixelate resolution

All the samples were analyzed to detect difference in pixelate resolution as compared to the document on which they were transplanted. Pixelate resolution as compared to the document is authentic or not. The results are extremely beneficial and reliable as after an appropriate magnification, color variation can be seen between transplanted area and the original document. Along with it on increasing the range of magnification the differences in font size and pixelate resolution are very much evident. The range of magnification was found to be within 26-1252%, which clearly determines the difference between the transplanted area and original document. The five types of hash values were also calculated to determine the difference between simulated document and the original one, the hash values were: MD4, MD5, RIPEMD-128, RIPEMD-160 and SHA-1. By adopting the material and methods used in this research, experts can be benefitted in dealing with softcopy transplanted forgery cases up to a huge extent.

**References**

1. Hajeel S, Sulong G (2013) State of the art of copy-move forgery detection techniques: a review. International journal of computer science 10: 174-183.
2. Rosebrock A (2014) Fingerprinting images for near-duplicate detection.
3. Zauner C (2010) Implementation and benchmarking of perceptual image hash functions.
4. Wen C, Yang K (2006) Image authentication for digital image evidence. Forensic science journal 1-11.
5. Li W, Yuan Y, Yu N (2008) Detecting copy-paste forgery of jpeg image via

---

**Table 2:** Shows the observation made on examination of samples with transplanted signatures on basis of Hash values.

| MD4    | MD5    | RipeMD128 | RipeMD160 | SHA-1 |
|--------|--------|-----------|-----------|-------|
| 1.     | 1.     | 1.         | 1.         | 1.    |
| 2.     | 2.     | 2.         | 2.         | 2.    |
| 3.     | 3.     | 3.         | 3.         | 3.    |
| 4.     | 4.     | 4.         | 4.         | 4.    |
| 5.     | 5.     | 5.         | 5.         | 5.    |

**Conclusion**

The present research reveals some simple methods which can help document experts as well as common man to establish whether the document is authentic or not. The results are extremely beneficial and reliable as after an appropriate magnification, color variation can be seen between transplanted area and the original document. Along with it on increasing the range of magnification the differences in font size and pixelate resolution are very much evident. The range of magnification was found to be within 26-1252%, which clearly determines the difference between the transplanted area and original document. The five types of hash values were also calculated to determine the difference between simulated document and the original one, the hash values were: MD4, MD5, RIPEMD-128, RIPEMD-160 and SHA-1. By adopting the material and methods used in this research, experts can be benefitted in dealing with softcopy transplanted forgery cases up to a huge extent.
block artifact grid extraction. International workshop on local and non-local approximation in image processing.

7. Thirumaga S, Allwin D (2012) Image manipulation detection using intrinsic statistical fingerprints. International journal of advanced research in computer science and software engineering 2: 207-212.

8. Muhammad N, Hussain M, Muhammad G, Bebis G (2011) Copy-move forgery detection using dyadic wavelet transforms. Computer graphics, imaging and visualization (cgiv), 2011 eighth international conference 103-108.

9. Koon RV, Jakubowski M, Moulin P (2000) robust image hashing. Image processing. Proceedings. international conference on, 3: 664-666.

10. Ramirez-gutierrez K, Nakano-miyatake M, Meana H (2013) Image authentication using perceptual hashing. Academic journals 8: 447-455.