Document Forgery: The State of Art

Surbhi Gupta[1], Parvinder S. Sandhu[2]

Abstract: The modern era is all about digitization and its usage. Over the past years, digital images have been widely used in the Internet and other applications. Whilst image processing techniques are developing at a rapid speed, tampering with digital images without leaving any obvious traces becomes easier and easier. This may give rise to some problems such as image authentication. In this paper, we will discuss this new methodologies used for forensics technology for documents and give an overview of the prior literatures. Some concluding remarks are made about the state of the art and the challenges in this novel technology.

Keywords: image authentication, passive/blind forensics, document tampering.

I. Introduction

The modern era is all about digitization and its usage to all the areas having impact on our life. Be it Medical field, entertainment field, office needs or customer services, digitization is everywhere. As the role has became so important none of the field is left untouched. Due to this the schemes for manipulating the digital versions of the information has emerged. People are daily looking for the methods to play with the digital inputs and coming out with edited output. Now that output could be of two types: one that is the improved version of the input and second the altered version of the input.

In this paper, we have considered the improved version as that version which is edited for the enhancement purpose only and will not change the original content of the input; whereas, the altered version has the changes which also effects the content of the input.

These types of digital forgeries are mainly performed on data, documents, Currency or Cheques, images, and video. Each such kind of manipulation which changes the content of any of these becomes illegal. If the information is digital than it's very easy to manipulate it, but if its not, for example in the case of documents or images the editor has to first digitize it and then manipulate it. There are many tools in the market which are readily available for this purpose and the most common among them is the scanner. In this paper we have concentrated on the document forgery attempts mainly.

II. Different Types of Tampering in Images

Document Tampering

This is most common type of tampering in which the content of the document is changed so as to make it more beneficial for a personal use. In this process the document is first scanned at a high resolution and then some of its content is changed for some specific purpose. The common example is change in academic certificates to show better awards.

Punjab Agricultural University

Certified that Surbhi Gupta
Admission No. L-2002-AE-230-M
daughter of Shri A. S. Gupta and Smt. Rama Gupta
having successfully completed the programme,
has obtained the degree of Master of Technology
in the discipline of Computer Science & Engineering
with Overall Credit Point Average 9.20 out of 10.00
(Eight point two zero out of Ten) in November 2004.

Given under the seal of the University

Ludhiana February 2, 2005 — Registrar — Vice-Chancellor — Chancellor

Fig1. Example of Document tampering

Steganography

Steganography is an alteration made to an image or video to hide some confidential information in it. The original image should look untouched but it must carry the message intended. This is done by altering the lower order bits of the image as they carry less meaningful data pertaining to the image. These bits are loaded with new bits that together can convey a meaningful message to the intender. Many variations of the above schemes have been implemented to make the identification of that message difficult. The aim is that the study of the image can never reveal the message or the presence of the message in it.

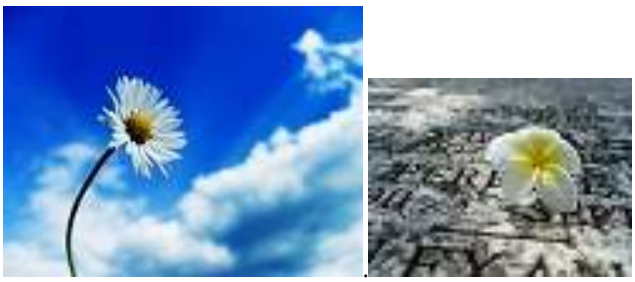

Fig 2. Vessel image, Hidden image

Ms. Surbhi Gupta is Research Scholar at Punjab Technical University, Jalandhar in the department of Computer Science Engineering.(email: royal_surbhi@yahoo.com)
Dr. Parvinder Singh Sandhu is with RBIEBT, Sahauran in the department of Computer Science. (e-mail:parvinder.sandhu@gmail.com)

Fig 3. Stego image

Image Inpainting

Image inpainting is a technique which is utilized for recovering and editing pictures or paintings. In this technique the required area of the images is edited and filled with desired information. Using this technique we can recover the missing parts of the image, we can make some objects invisible in the image and can also add some new objects in the image.

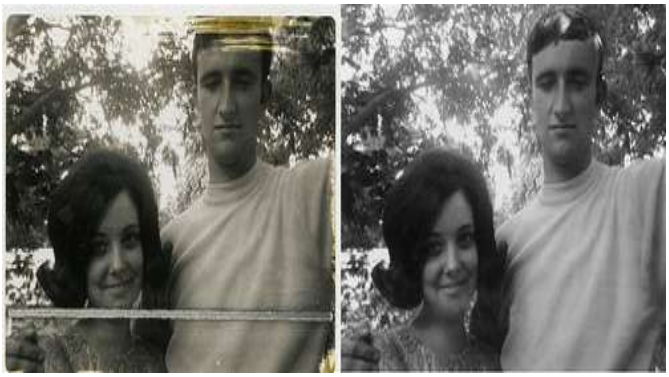

Fig 4. Example of Image Inpainting

Digital forgery

This type of digital tampering is quite similar to digital inpainting in sense of technique bur differs in sense of intention. The aim in digital inpainting is usually to improve the image or remove the unwanted parts in the image but in digital forgery the aim is to misguide the viewer with wrong information provided in the image. Sometimes the face of the person is changed or some other person is introduced to present the wrong facts.

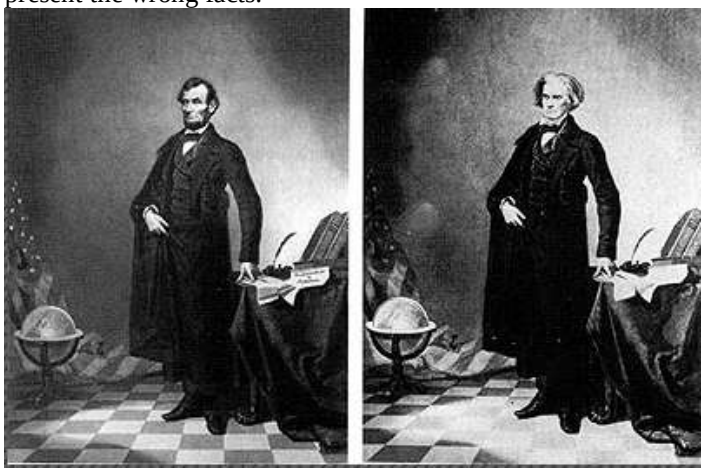

Fig 5. Example of Digital Forgery

III. Method to prevent Digital forgery

- Active: active methods can prevent forgery by embedding some data in the image which is difficult to replicate. Two main variations of the same are:
  - Watermarking
  - Digital signatures

This technique is the most prevalent one to prevent the document from forgery. The idea is to embed an image or signature or symbol inside the image which will authenticate the image. Any tampering in the image will cause the watermark to be altered that can be easily identified.

- Passive: Passive techniques are applied when no active technique has been used in the image. In the absence of active technique there is no direct method to recognize weather the image is altered or not.
  - Identify Image source: the main aim of the technique is identify the source of image i.e.
    - Camera
    - Photo Editor
    - Printer

    Once the source is identified then image origin can be identified. If the origin is camera then it can be concluded that it's an original image. If the source is printer then further the printer model needs to be identified. As the printer model for the original document is usually known, the knowledge of suspected document printer model will clarify whether it is original or not. If the source is photo editor the image is known to be edited.
  - Detect tampering: this methodology tries to detect the tampering in image by analyzing the image features.
    - Global analysis
      - Jpeg Quantization
      - DCT coefficient analysis:
      - Interpolation analysis
    - Local analysis: after global analysis the images is further looked upon for areas of manipulation which can be identified by local analysis.

IV. Major Contributions

The work on passive techniques is gaining the interest day by day. A lot of researchers has contributed immensely in this filed but still many area are open and a number of question are unanswered. Many authors aimed at implementing generalized forensic techniques for documents and images based on deterioration of their quality or change in image features. Much work has been done to identify the source of the image as camera, printer or the scanner.

Mikkilineni et al. (2005) discussed the printer identification based on gray level occurrences. The author first estimated the Gray-Level Co-occurrence Matrix (GLCM) based on Region of Interest [1]. Then he derived 20 parameters from the GLCM. The first four are the marginal means and variances The next seven features are the energy of the normalized GLCM, three entropy measurements, the

maximum entry in the GLCM, and two correlation metrics. Four features are obtained from the difference histogram. They are the energy, entropy, inertia, and local homogeneity. The last five features are obtained from the sum histogram. They are the energy, entropy, variance, cluster shade, and cluster prominence. The last two are the variance and entropy of the pixel values in the ROI. Although the matrix performs considerably well in identifying the printer, it is important to note that the technique requires the prior information about the printers in question. If the unknown document was printed by a printer which is not included in the classifier training data set, then the result will point to one of the considered printers.

Moreover the technique will not work with multiple font sizes, font types, and also different Characters.

Then Gupta et al. (2006) stated that overall features of document can be studied using Video Spectral Comparator (VSC) and document forgery can be detected but it requires the specialized hardware. The author made important observations about the variation in grey levels present in the original document image and the edited document image at the white spaces and the edges of text. The technique fixes the printer and scanner combination used to produce the fraudulent document. The technique works only if a characteristic database for all the combinations is first created [5].

Khanna et al.(2009) identified features based methods to identify scanner . The work proposed methods for source scanner identification for scanned text documents using texture features [3]. Author has shown by the experiments that the proposed method is robust to JPEG compression and gives 100% classification accuracy for classifying A4 size text documents and more than 95% classification accuracy for classifying smaller blocks of size 512 × 512. The limitation was that the effect of variation in font sizes and fonts needs to be tested and investigated further for one or more scanners. As with every other method for image forensic, this scheme also fails to work in certain circumstances such as heavy post-processing of scanned documents.

Ryu et al. (2009) has presented the fraud document identification scheme using 17 image quality measures. From C1 to C5, they measures similarity between two images. C1, C2, and C3 represent structural content, normalized cross correlation, and image fidelity, respectively. The absolute mean and variance of the angles between the pixel vectors are C4 and C5.6 measures from S1 to S6 come from frequency domain. S1, S2, and S3 represent MSE of magnitude, phase, and combined measures, respectively. H1, H2, and H3 are measures related to human visual system (HVS). Even though they used 17 features there were some wrong classifications made by the classifier. It means that selection of proper features is the key to achieve accurate results. The need is to improve on these measures to obtain a more accurate set to classify the documents correctly [4].

V. Conclusion

There is a growing need for digital image forensics techniques, and many techniques have been proposed to address various aspects of digital image forensics problem. Although many of these techniques are very promising and innovative, they all have limitations and none of them by itself offers a definitive solution. Ultimately, these techniques have to be incorporated together to obtain reliable decisions. However, there is still one major challenge to be met in case of document forensics i.e. to find the actual location of alteration on the document. Till the time this problem remains unsolved, the mystery remains as such.

References

[1.] K. Mikkilineni, O. Arslan, P.-J. Chiang, R. M. Kumontoy, J. P. Allebach, and G. T.-c, "Printer forensics using svm techniques," in Proceedings of the IS&T's NIP21: International Conference on Digital Printing Technologies, vol. 21, Baltimore, MD, October 2005, pp. 223–226.

[2.] Khanna, N., Mikkilineni, A.K., Martone, A.F., Ali, G.N., Chiu, G.T.-C., Allebach, J.P., Delp, E.J.: A survey of forensic characterization methods for physical devies. Digital Investigation 3, 17–28 (2006).

[3.] N. Khanna, A. K. Mikkilineni, and E. J. Delp, "Scanner identification using feature-based processing and analysis," Trans. Info. For. Sec.vol. 4, no. 1, pp. 123–139, 2009.

[4.] Ryu SJ, Lee HY, Cho IW, Lee HK (2008) Document forgery detection with SVM classifier and image quality measure. Lect Notes Comput Sci 5353:486–495.

[5.] Gupta, G., Mazumdar, C., Rao, M.S., Bhosale, R.B.: Paradigm shift in document related frauds: Characteristics identification for development of a non-destructive automated system for printed documents. Digital Investigation 3, 43–55 (2006).

[6.] Gupta, G., Saha, S.K., Chakraborty, S., Mazumdar, C.: Document Frauds: Identification and Linking Fake Document to Scanners and Printers. In: Proc. ICCTA 2007 (2007).

[7.] Avcibas, I., Sankur, B., Sayood, K.: Statistical evaluation of image quality measures. J. Electron. Imag. 11, 206–223 (2002).

[8.] Avcibas, I., Memon, N., Sankur, B.: Steganalysis using image quality metrics. IEEE Trans. Image Processing 12(2), 221–229 (2003).