International Journal of Information Technology and Knowledge Management
July-December 2011, Volume 4, No. 2, pp. 555-557

PRINTER BASED DOCUMENT COUNTERFEIT DETECTION TECHNIQUES-A REVIEW

Jagdeep Kaur

This paper discusses the various techniques to identify the printer responsible for generating a fraudulent document. The emphasis is on texture based, HSV color space and grey level co-occurrence based matrix methods. For the purpose of anti-counterfeit other techniques viz. intelligent Document Image Preprocessing Algorithms, Geometric Layout Analysis and Page Segmentation & Logical Layout Analysis and Reading Order Detection are suggested.

1. Introduction

In the last years, the number of observed forgeries of official documents has increased strongly. But a strong growth of forged documents can be observed in the area of identification, like passports, visa, drivers' licenses, cheques, invoices or certificates of authenticity. Digital imaging techniques have evolved to a level where forgeries can be created within seconds, e. g. using color photocopy machines, that are indistinguishable from the original for the untrained human eye. Thus, there is a large demand for automatic systems that can decide if a document is genuine or not, and the interested parties range from governmental organizations overlarge companies like banks and insurances down to small companies like pharmacies and even end users. The detection of counterfeit in printed documents is currently based mainly on built-in security features or on human expertise. A classification system is needed that supports non-expert users to distinguish original documents from PC-made forgeries by analyzing the printing technique used. Generating fraudulent document using a scanner and printer for malicious and unlawful gain has seen a paradigm shift. Some more work is required to be done for the problem of detecting a fraudulent document and then fixing it to color laser printers or color inkjet printers. For each printer in question study has found characteristics specific to printing technology and same varies among same model, same company printers. For the forensic purpose the linking of a document to a scanner/ printer and printing technology is extremely useful. The Printed Anti-counterfeit detection system deals with various issues like handwriting & signature Identification, Detection of alterations, Deciphering obliterations, alterations, erasures, Comparison of inks and identification of type of writing instrument, Printer Identification of the document. This paper surveys the last two issues.

Computer Science & Engineering Department, itm University Sector 23 A, Gurgaon, Haryana, India
E-mail: jagdeep@itmindia.edu

2. Current Approaches

Printed document examination plays an important role in public security. In principle there are two ways that the problem of counterfeit detection for documents can be accessed: model based or generically.

2.1 The model-based approach requires pre-knowledge on characteristic features of a document to be checked and then searches specifically for them. Often, the document are already created with the possibility for such checks in mind by including security features that are easy to check for later, either by the human eye or by using special devices. Typical examples are banknotes and credit cards, which contain watermarks, holograms or special ink, which is only visible in ultraviolet light. Correctly applied, these methods provide the highest level of security and many approaches to model-based forgery detection exist. However, model-based systems have the drawback that only those documents can be checked, for which a model of the security features is available, e. g. from a database. Many classes of documents, e. g. stamps, come in such a great variety of characteristics that a database of all such models is impractical. Other important documents can be generated by anybody on-the-fly, e. g. invoices, making a database of all originals impossible. Some of these drawbacks are avoided in the alternative, generic way.

2.2 Here, in generic approach a general selection of features is extracted from a document, and the decision if a document is genuine is based only on a class membership of a document and statistical information of the expected features. Because of this limited knowledge, generic counterfeit detection systems show a larger rate of error than model-based, but have the advantage of being applicable for a wider class of documents. Typically, the generic approach does not really test if a document is genuine, but rather if a certain method of forgery has been applied.

In the generic selection many techniques have been applied so far viz.

2.2.1 Video Spectral Comparator, an imaging device which works on the concept of separation of wavelength of light of spectra, can only detect the genuineness of a document, but fixing it to the machine/tool used in its generation is not possible. It works in a broad range of electromagnetic spectrum ranging from ultra violet to infrared. It helps in the detection of forgeries by producing spectra based on absorption, reflection and transmission of various combinations of light arrangements viz. spot light, sidelight, ultraviolet light, fluorescent light etc. The high cost, cumbersome size and the skills required to operate such an instrument makes it inappropriate for use in common financial institutions like banks, airports, customs department etc; they can only be used by experts. In addition, this instrument does not offer much assistance for fixing the origin and authorship of the questioned documents, which is crucial in stopping the menace of abuse of technology for malicious and unlawful gain.

2.2.2 The Thin Layer Chromatography (TLC), methods are used to differentiate between various kinds of inks, and thus can only say whether the ink components of the questioned document are different from that of the original. They require high investment and expertise and laborious off-line processes, making them inappropriate for real time forgery detection and fixing. Identifies characteristics specific to inject printers and put forwards a technique to detect and fix the origin of fraudulent document from two scanners and printers based unique color present, percentile of dark color and pattern of dots distributed in an magnified questioned image.

2.2.3 Some other techniques to identify the printer involves use of textures based features & pixel distribution in color cube. [1] This approach assumes that original document for verification is available. The Table 1 provides the observations made in this context.

Table 1

S.No	Parameter	Indication
1.	Variance of Intensity	It provides an idea of amount of variability in the original and fraudulent document
2.	Total number of unique color Count	It varies considerably in fake documents because of impurity gets introduced in their generation process
3.	Gray Level Co-occurrence Matrix Uniformity	This parameter is technology dependent (Laser or inkjet based) is used.

2.2.4 A printer identification system is presented for the purpose of identifying the printer which created a suspect printed document [2]. This approach was tested for laser printer only. As shown in Figure 1. It is composed of image acquisition, image preprocessing and character matching. An optical instrument is developed for document image observation and acquisition under high resolution. Normalized printed characters are obtained and recognized in Preprocessing. Distance transform algorithm is applied to character image matching and distance between printed documents is calculated. Finally the minimum distance classifier is applied to printer identification. Experiments are carried out in a database of Laser printers and very promising results are achieved.

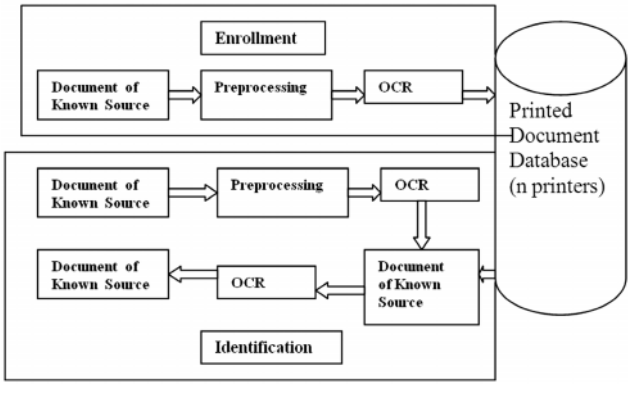

Fig. 1: The Printer Identification System

3. CONCLUSIONS

The present system for identification of printer & printing technique takes into consideration one or two parameters only. Future work may be concentrated on integrating the

various factors like change in printing styles over a long term, individual character & print quality variations in different times will be considered. Beyond the digitalization of pure textual information with Optical Character Recognition (OCR) algorithms, the interpretation of the layout is also crucial for understanding and digitizing the content of a document. Some basic layout analysis technologies including Intelligent Document Image Preprocessing Algorithms, Geometric Layout Analysis and Page Segmentation & Logical Layout Analysis and Reading Order Detection can be incorporated.

REFERENCES

[1] Gaurav Gupta, Sanjoy Kumar Saha, Shayok Chakraborty, Chandan Mazumdar, "Document Frauds: Identification and Linking Fake Document to Scanners and Printers", ICCTA 07, IEEE, pp 497-501, 2007.

[2] Wei Deng, Qinghu, chen, Feng, Yaun, YuchenYan, "Printer Identification Based on Distance Transform", Proceedings of the International Conference on Intelligent Networks and Intelligent Systems, ICINIS' 09, IEEE, pp 565-568, 2008.