Printer ballistics through character's texture analysis

Adriano Ruggero*
Gabriel Rodrigues†
Mário Brito‡
Maurício Perez§
Anderson Rocha¶

Abstract

We describe a technique for ballistics of printed documents, that is, link a printed document to a specific printer. The principle of this technique is the analysis of character's texture, by extracting some properties from the characters of the scanned images from the printed documents, and relate this properties through a co-occurrence matrix. This matrix can be used to create a "fingerprint" for the characters related to the same printer, what allows us to identify the specific printer device that printed these characters.

1. Introduction

In August of 2013, a russian man wrote his own small print in a credit card contract [1]. The credit card's administrator bank didn't read the amendments made by the client, and just signed and certified the document. The changes included unlimited credit line, 0 percent interest rates and no fees. When the bank decided to terminate the man's credit card, because overdue payments, he sued them for more than 24 million rubles (US$ 727.000). How could the bank prove the falsification?

Altough we are living in a digital era, printed documents still are a significant part of our day by day. Likewise, with the constant reduction in prices and increase in quality of printing equipment, forgeries become increasingly commonplace.

Legal aspects aside, a way to verify if a document, or a part of it, came from a specific device can be through character's texture analysis.

Our approach for the analysis of character's texture is as follows. From printed pages scanned at high resolution, selected characters were extracted. From these characters, we obtained its properties of contrast, correlation, energy and homogeneity, creating with them a co-occurrence matrix. This matrix can be called a "fingerprint" of the character. This "fingerprint" of characters is closely related to the printing device which originated it, and can be used to identify which printer was responsible for printing it.

However, slight imperfections may occur during the printing and/or scanning process of documents. To handle these small errors (or variations), characters were selected from different areas of scanned document, their properties were obtained and then classified using machine learning algorithms.

2. State-of-the-Art

Related work: [2]

3. Proposed Solution

Our solution consist in getting the image of characters selected from scanned documents in grayscale, extract its properties of contrast, correlation, energy and homogeneity - creating a co-occurrence matrix - and cluster them by machine learning algorithms.

3.1. Printers dataset

The documents used for this work were obtained from the Wikipedia site [3], and were written in English, some of them containing pictures, some not. They were printed by printers listed in Table 3.1.

These documents were scanned at high resolution and saved as Tiff file format (Tagged Image File Format), forming a database. These files were made available through an FTP site [4]. An example of a typical document used in this

*Institute of Computing, University of Campinas (Unicamp). **Contact**: arruggero@lasca.ic.unicamp.br

†Institute of Computing, University of Campinas (Unicamp). **Contact**: gabriel_rodrigues@aol.com

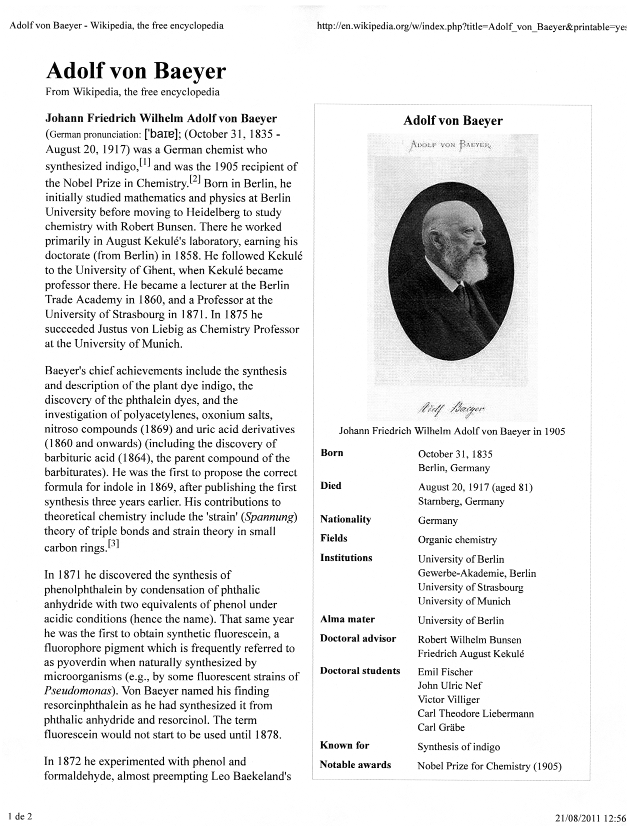
‡Institute of Computing, University of Campinas (Unicamp). **Contact**: britomar@aedu.com

§Institute of Computing, University of Campinas (Unicamp). **Contact**: mauriciolp84@gmail.com

¶Institute of Computing, University of Campinas (Unicamp). **Contact**: anderson.rocha@ic.unicamp.br

Table 1. Printers used in this work

Printer	Documents
Brother-HL4070CDW	28
Canon-D1150	28
Canon-MF3240	28
Canon-MF4370DN	28
HP-CLJ-CP2025A	28
HP-CLJ-CP2025B	28
HP-JL-CP1518	28
Lexmark-E260D	28
OKI-C330	28
Samsung-CLP315	28


Adolf von Baeyer - Wikipedia, the free encyclopedia http://en.wikipedia.org/w/index.php?title=Adolf_von_Baeyer&printable=ye:

Adolf von Baeyer

From Wikipedia, the free encyclopedia

Johann Friedrich Wilhelm Adolf von Baeyer (German pronunciation: [ˈbaɪɐ]; (October 31, 1835 - August 20, 1917) was a German chemist who synthesized indigo,[1] and was the 1905 recipient of the Nobel Prize in Chemistry.[2] Born in Berlin, he initially studied mathematics and physics at Berlin University before moving to Heidelberg to study chemistry with Robert Bunsen. There he worked primarily in August Kekulé's laboratory, earning his doctorate (from Berlin) in 1858. He followed Kekulé to the University of Ghent, when Kekulé became professor there. He became a lecturer at the Berlin Trade Academy in 1860, and a Professor at the University of Strasbourg in 1871. In 1875 he succeeded Justus von Liebig as Chemistry Professor at the University of Munich.

Baeyer's chief achievements include the synthesis and description of the plant dye indigo, the discovery of the phthalein dyes, and the investigation of polyacetylenes, oxonium salts, nitroso compounds (1869) and uric acid derivatives (1860 and onwards) (including the discovery of barbituric acid (1864), the parent compound of the barbiturates). He was the first to propose the correct formula for indole in 1869, after publishing the first synthesis three years earlier. His contributions to theoretical chemistry include the 'strain' (*Spannung*) theory of triple bonds and strain theory in small carbon rings.[3]

In 1871 he discovered the synthesis of phenolphthalein by condensation of phthalic anhydride with two equivalents of phenol under acidic conditions (hence the name). That same year he was the first to obtain synthetic fluorescein, a fluorophore pigment which is frequently referred to as pyoverdin when naturally synthesized by microorganisms (e.g., by some fluorescent strains of *Pseudomonas*). Von Baeyer named his finding resorcinphthalein as he had synthesized it from phthalic anhydride and resorcinol. The term fluorescein would not start to be used until 1878.

In 1872 he experimented with phenol and formaldehyde, almost preempting Leo Baekeland's

Adolf von Baeyer

Adolf von Baeyer

Johann Friedrich Wilhelm Adolf von Baeyer in 1905

Born	October 31, 1835 Berlin, Germany
Died	August 20, 1917 (aged 81) Starnberg, Germany
Nationality	Germany
Fields	Organic chemistry
Institutions	University of Berlin Gewerbe-Akademie, Berlin University of Strasbourg University of Munich
Alma mater	University of Berlin
Doctoral advisor	Robert Wilhelm Bunsen Friedrich August Kekulé
Doctoral students	Emil Fischer John Ulric Nef Victor Villiger Carl Theodore Liebermann Carl Gräbe
Known for	Synthesis of indigo
Notable awards	Nobel Prize for Chemistry (1905)

1 de 2 21/08/2011 12:56

Figure 1. Typical document used in this work [3].

study can be seen in Figure 1.

The characters chosen for this work were ”e” and ”t”, both in lowercase, because they are, respectively, the first and second most common letters in texts written in English [5].

Due to the fact that all scanned documents come from laser printers, it was necessary to take some precautionary measures. Laser printers are known as ”page printers”, while dot matrix printers and inkjet printers are called ”line printers”. This is a crucial difference, and should be considered for more careful study. Line printers print documents line by line from the top of the sheet, keeping a characteristic pattern which periodically repeats for each paper feed. That is, any line printed by this printer model will have basically the same characteristics, regardless of their vertical location in the sheet of paper. Page printers, instead, do not print documents line by line. In this case, an image of the entire page is ”printed” on the photoreceptor drum by a laser unit. This image attracts toner particles, and then transfers it to the paper sheet. Finally a fuser unit heats the paper, so the toner melts and attach it. Figure 2 shows a default page printer schema.

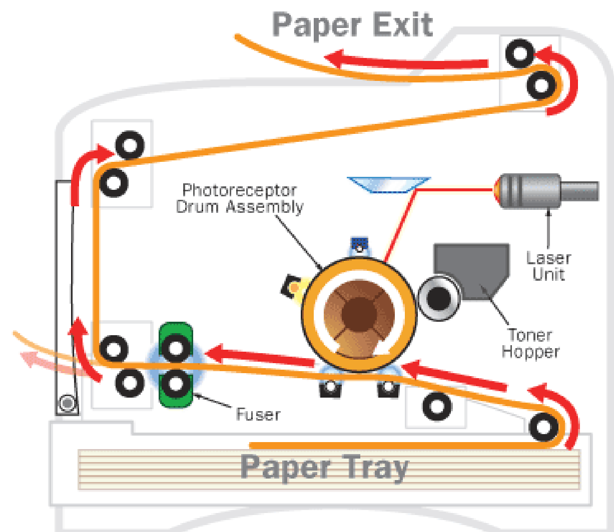


Figure 2. Default page printer schema [6].

4. Experiments and Discussion

5. Conclusions and Future Work

References

[1] RT. 700k windfall: Russian man outwits bank with handwritten credit contract. http://rt.com/business/man-outsmarts-banks-wins-court-221/, last access in November 17, 2013. 1

[2] Orhan Bulan, Junwen Mao, and Gaurav Sharma. Geometric distortion signatures for printer identification. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009, 19-24 April 2009, Taipei, Taiwan*, pages 1401–1404. IEEE, 2009. 1

[3] Wikipedia. Wikipedia - the free encyclopedia. http://www.wikipedia.org/, last access in November 19, 2013. 1

[4] Giuliano Pinheiro. Index of / giulianorp/printer_dataset. http://www.recod.ic.unicamp.br/

~giulianorp/printer_dataset/, last access in November 19, 2013. 1

[5] University of Notre Dame. Letter frequencies in the english language. http://www3.nd.edu/~busiforc/handouts/cryptography/Letter.html, last access in November 17, 2013. 2

[6] Forensic Document Examination Services. Letter frequencies in the english language. http://www.forensicdocumentexaminer.co.uk/Document_Analysis.html, last access in November 17, 2013. 2