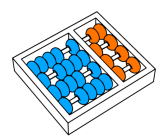
Printer Ballistics Through Texture Analysis of Characters

Adriano Ruggero, Gabriel Fernandes, Mário Brito, Maurício Perez

Institute of Computing - Unicamp

November 29, 2013



Outline

Motivation

- We (still) live in a ”paper era”
- Documents forgery has become common
- There is a way to relate a document to a specific printer?

Printer attribution

A way to do this is called ”Printer Attribution”

Methods

- Geometric distortion
- Texture analysis of characters

Geometric distortion

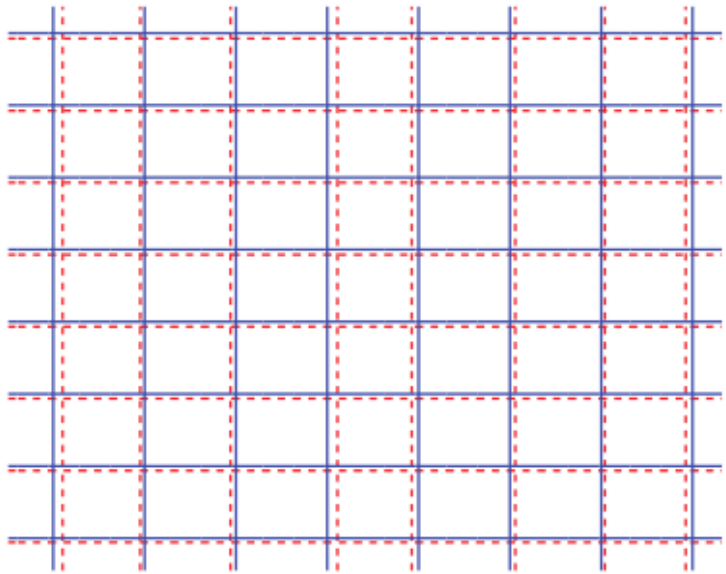
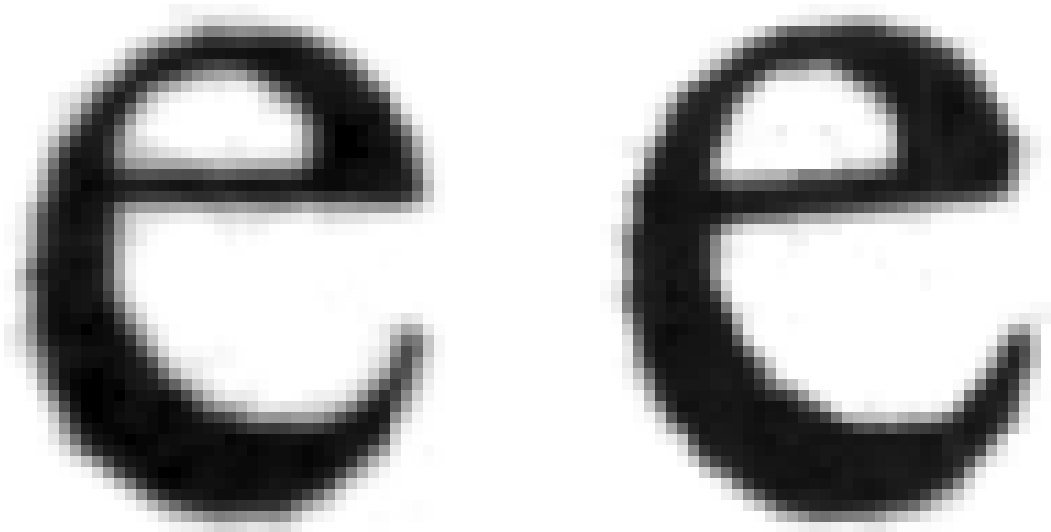


Figure 1: Geometric distortion [1]

[1] Geometric Distortion Signatures for Printer Identification[1]

Texture analysis of characters



State-of-the-art

- **Bulan et al[2]:** a method for analyzing geometric distortions introduced during the printing process of electrophotographic printers (EP).
- **Kee and Farid[3]:** a method of geometric modeling of degradation caused by the printer.

[2] Geometric Distortion Signatures for Printer Identification[1]

[3] Printer Profiling for Forensics and Ballistics[2]

Proposed solution

- Get the image of characters selected from scanned documents (grayscale)
- Create a co-occurrence matrix
- Extract its properties (contrast, correlation, energy and homogeneity)
- Create a feature vector from this properties
- Use machine learning algorithms to classify them

Printers

Table 1: Printers used in this work

Printer	Documents	Characters ”e”	Characters ”t”
Brother-HL4070CDW	28	252	252
Canon-D1150	28	252	252
Canon-MF3240	28	252	252
Canon-MF4370DN	27	252	252
HP-CLJ-CP2025A	28	250	250
Lexmark-E260D	30	636	629

Characters

- Characters ”e” and ”t” (most common in English texts)
- Same size, same font, no texts effects
- Misaligned characters were summarily discarded

Differences between aligned and misaligned characters

Table 2: Differences between an original character and a rotated character.

Property	Original character	Rotated character (-4º)
Contrast	3.2443 - 2.2905	5.1617 - 4.6504
Correlation	0.7869 - 0.8502	0.6967 - 0.7264
Energy	0.1608 - 0.1744	0.1106 - 0.1216
Homogeneity	0.6946 - 0.7462	0.6610 - 0.6995

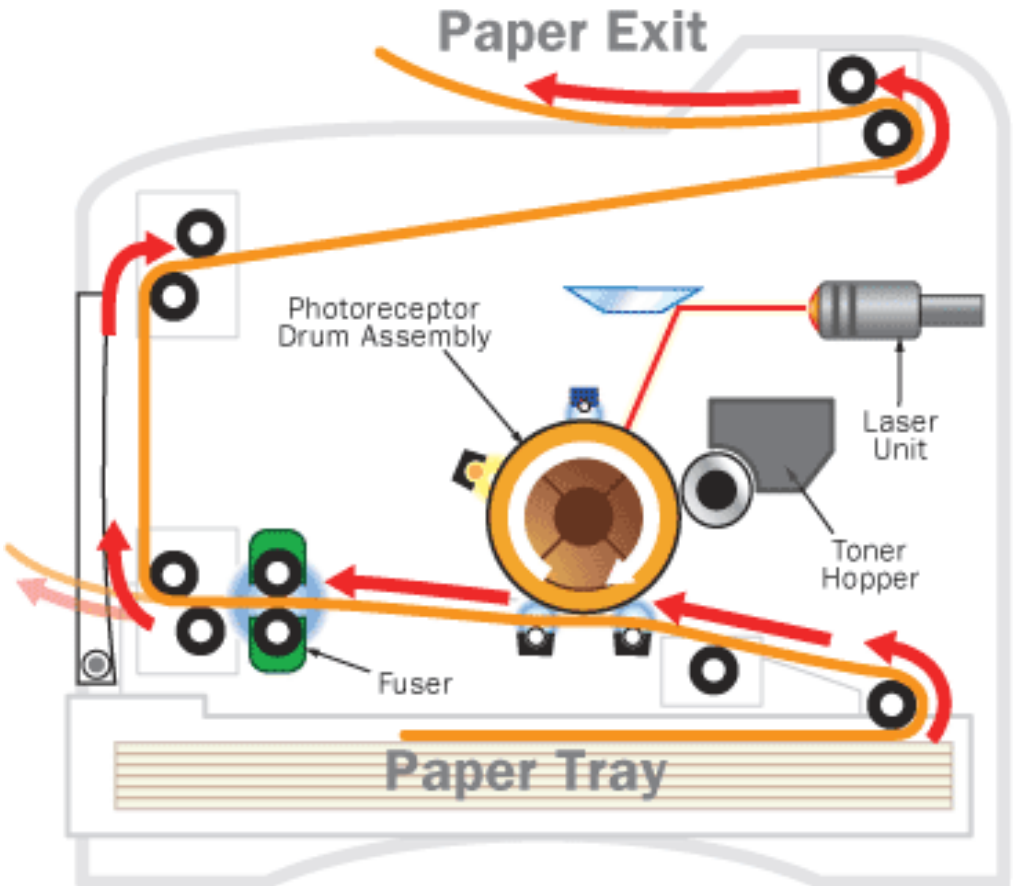
Something about printers

- All documents came from laser printers, so...

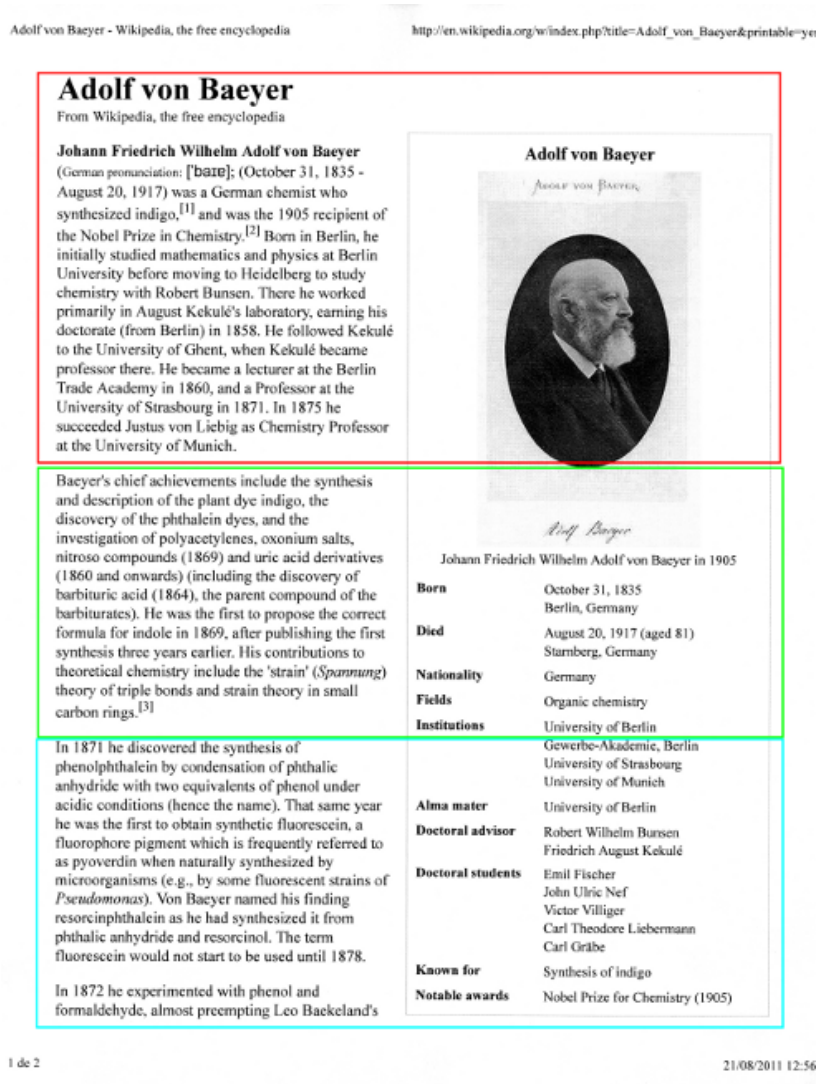
Something about printers

- All documents came from laser printers, so...
  - we need to understand how they works!

Default laser printer schema



Division of document areas



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: ['baɪɐ]; (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

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

Gray level co-occurrence matrix

The primary use of the co-occurrence matrix is characterized texture in an image from a set of statistics for instances of each gray level in different pixels along different directions[4].

In other words...

- A matrix of relative frequencies $P(i, j, d, \theta)$
  - p represents the pixel-of-interest
  - i and j represents the properties
  - θ represents the distance

[4] Classificação de texturas a partir de vetores de atributos e função de distribuição de probabilidades[3]

Character's selection and extraction

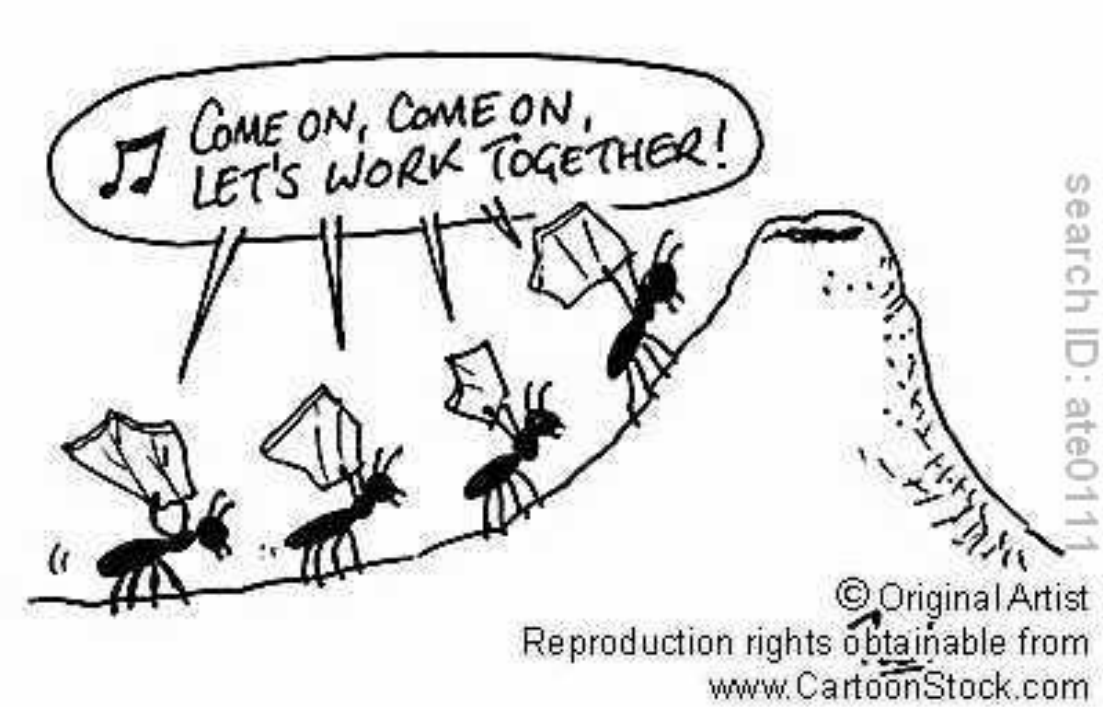


Figure 2: A ant's work![5]

[5]The Wifey Journals[4]

Neighborhood

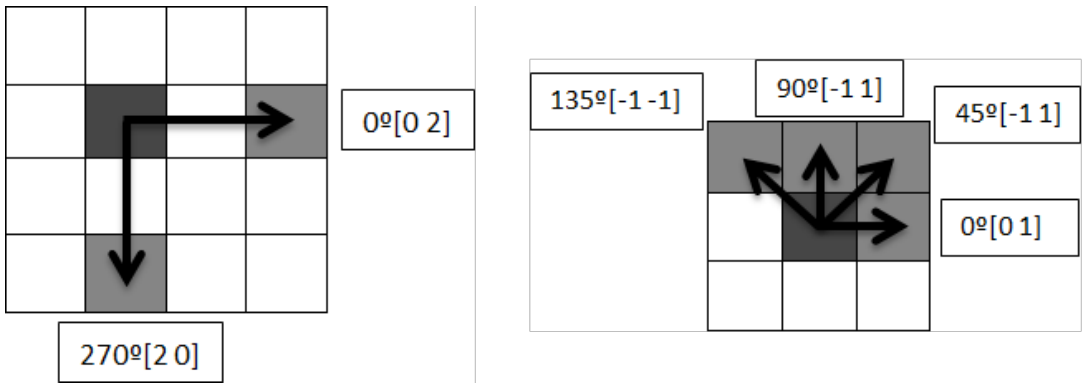


Figure 3: Neighborhood A, leftmost, and B, rightmost, used in properties extraction.

Algorithms versus correct classification

Table 3: Percentage of correct classifications of printers.

	Neighborhood A		Neighborhood B	
Method	Chars e's	Chars t's	Chars e's	Chars t's
Logistic	81	81.3	85	84.6
KStar	77.6	83	72	79.6
RotationForest	83.1	85	81.7	85.7
NNge	74.1	80.2	72.2	67.8
LMT	83.8	84.6	82.7	85.5

Conclusions

Acknowledgements

- Professor Anderson Rocha
- Giuliano Pinheiro

Bibliography I

Orhan Bulan, Junwen Mao, and Gaurav Sharma.
Geometric distortion signatures for printer identification.
In *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, ICASSP '09, pages 1401–1404, Washington, DC, USA, 2009. IEEE Computer Society.

Eric Kee and Hany Farid.
Printer profiling for forensics and ballistics.
In *Proceedings of the 10th ACM Workshop on Multimedia and Security*, MM&Sec '08, pages 3–10, New York, NY, USA, 2008. ACM.

Bibliography II

Anderson de Rezende Rocha and Neucimar Jerônimo Leite.
Classificação de texturas a partir de vetores de atributos e função de distribuição de probabilidades.

The Wifey Journals.
Ants!
`http://www.thewifeyjournals.com/2010/09/ants.html`, last access in November 28, 2013.

Thanks

Thanks!

Adriano R. Ruggero, Gabriel Rodrigues, Mário F. Brito,
Maurício L. Perez