

Printer Ballistics Through Texture Analysis of Characters

Adriano Ruggero, Gabriel Rodrigues, Mário Brito, Maurício Perez

Institute of Computing - Unicamp

November 29, 2013



Outline

Motivation

Introduction

State-of-the-art

Proposed solution

Experiments and discussion

Conclusions and future work

Acknowledgements

References



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?

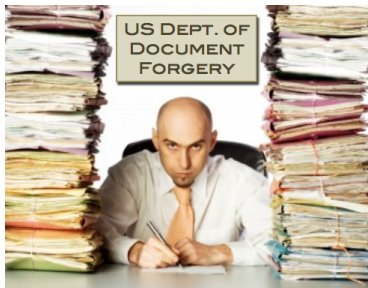


Figure 1: Document forgery¹

¹The Infothority[1]



Printer attribution

A way to do this is called "Printer Attribution"



Methods

- ▶ Geometric distortion
- ▶ Texture analysis of characters



Geometric distortion

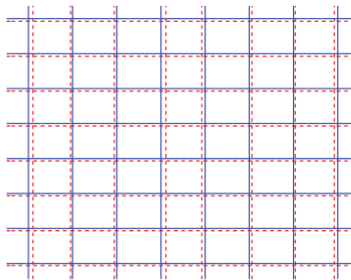
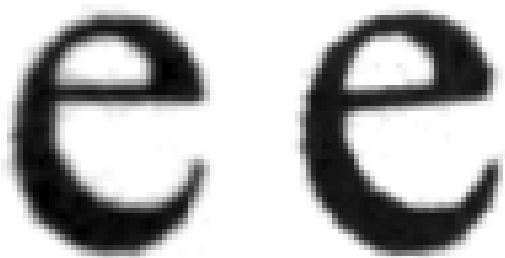


Figure 2: Geometric distortion ²

²Geometric Distortion Signatures for Printer Identification[2]



Texture analysis of characters



State-of-the-art

- ▶ **Bulan et al**³: a method for analyzing geometric distortions introduced during the printing process of electrophotographic printers (EP).
- ▶ **Kee and Farid**⁴: a method of geometric modeling of degradation caused by the printer.

³Geometric Distortion Signatures for Printer Identification[2]

⁴Printer Profiling for Forensics and Ballistics[3]

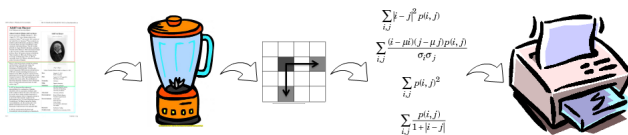


Proposed solution

- ▶ 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



Process overview



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	252	252
Lexmark-E260D	28	252	252



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 (-4°)
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

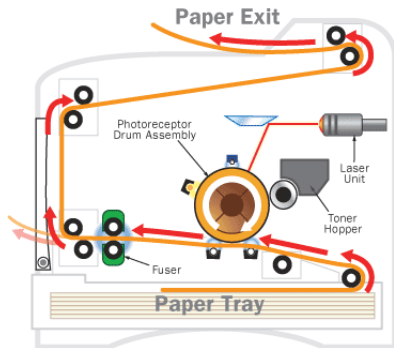


Figure 3: Default laser printer schema⁵

⁵Forensic Document Examination Services[4]

Division of document areas

Adolf von Baeyer - Wikipedia, the free encyclopedia

http://en.wikipedia.org/w/index.php?title=Adolf_von_Baeyer&printable=yes

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.

Adolf von Baeyer



Baeyer's chief achievements include the synthesis and description of the plant dye indigo, the discovery of the phthalic dyes, and the investigation of polyacetylenes, oxonium salts, nitroso compounds (1869) and alic 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 pyoverdinin when naturally synthesized by microorganisms (e.g., by some fluorescent strains of *Pseudomonas*). Von Baeyer named his finding resorcinphthalic 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

<i>Adolf Baeyer</i>	
 <div>Johann Friedrich Wilhelm Adolf von Baeyer in 1905</div>	
Born	October 31, 1835 <div>Berlin, Germany</div>
Died	August 20, 1917 (aged 81) <div>Strasbourg, Germany</div>
Nationality	Germany
Fields	Organic chemistry
Institutions	University of Berlin <div>Gewerbe-Akademie, Berlin</div> <div>University of Strasbourg</div> <div>University of Munich</div>
Alma mater	University of Berlin
Doctoral advisor	Robert Wilhelm Bunsen <div>Friedrich August Kekulé</div>
Doctoral students	Emil Fischer <div>John Ulric Nef</div> <div>Victor Villiger</div> <div>Carl Theodore Liebermann</div> <div>Carl Gräbe</div>
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 directions⁶.

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 (gray level values)
 - ▶ d represents the distance
 - ▶ θ represents the angle

⁶Classificação de texturas a partir de vetores de atributos e função de distribuição de probabilidades[5]



Character's selection and extraction

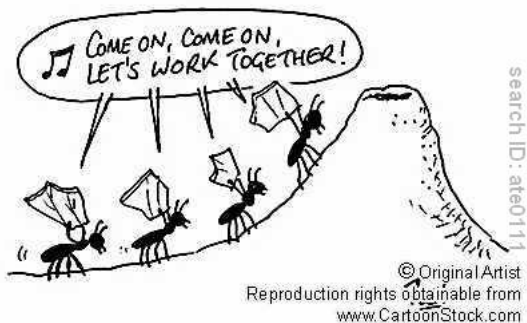


Figure 4: An ant's work!⁷

⁷The Wifey Journals[6]

Neighborhood

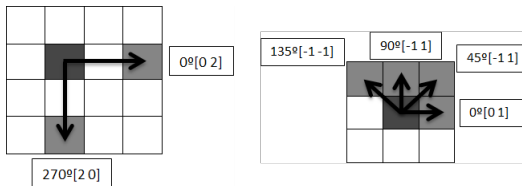


Figure 5: Neighborhood A, leftmost, and B, rightmost, used in properties extraction.

Algorithms versus correct classification

Table 3: Percentage of correct classifications of printers.

Method	Neighborhood A		Neighborhood B	
	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



Printer attribution results

Table 4: Percentage of correct printer attribution.

Method	Chars e's	Chars t's	Chars e's and t's
Logistic	21/24 = 87.5%	22/24 = 91.7%	21/24 = 87.5%
RotationForest	21/ 24 = 87.5%	21/24 = 87.5%	22/24 = 91.7%
LMT	21/24 = 87.5%	21/24 = 87.5%	22/24 = 91.7%



Conclusions

- ▶ Is possible assign a document to a printer analyzing the texture of its characters
- ▶ Greater feature vectors doesn't mean a better classifier
- ▶ Character choice would have probably low impact on the results



Future work

- ▶ Extract more varied or more characters
- ▶ Try other texture analyzers, such as HOG (Histograms of Oriented Gradients Extract) or LBP (Local Binary Patterns)
- ▶ A (semi) automated method to extract the characters would be of great benefit to researchers in this field



Acknowledgements

- ▶ Professor Anderson Rocha
- ▶ Giuliano Pinheiro



Bibliography I



The Infothority.

The infothority.

<http://theinfothority.com/>, last access in November 28, 2013.



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.



Bibliography II



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.



Forensic Document Examination Services.

Letter frequencies in the english language.

http://www.forensicdocumentexaminer.co.uk/Document_Analysis.html, last access in November 17, 2013.



Anderson de Rezende Rocha and Neucimar Jerônimo Leite.

Classificação de texturas a partir de vetores de atributos e função de distribuição de probabilidades.



Bibliography III



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

