Forensic Science International 247 (2015) 105–125

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Laser printer attribution: Exploring new features and beyond

Anselmo Ferreira [a,*], Luiz C. Navarro [a], Giuliano Pinheiro [a], Jefersson A. dos Santos [b], Anderson Rocha [a]

[a] Institute of Computing, University of Campinas, Av. Albert Einstein, 1251, Cidade Universitaria, Campinas, SP CEP 13083-852, Brazil
[b] Universidade Federal de Minas Gerais, Department of Computer Science, Av. Antônio Carlos 6627 – Prédio do ICEx – Pampulha, Belo Horizonte, MG 31270-010, Brazil

ARTICLE INFO

Article history:
Received 26 August 2014
Received in revised form 26 November 2014
Accepted 27 November 2014
Available online 23 December 2014

Keywords:
Printer forensics
Texture patterns
Banding

ABSTRACT

With a huge amount of printed documents nowadays, identifying their source is useful for criminal investigations and also to authenticate digital copies of a document. In this paper, we propose novel techniques for laser printer attribution. Our solutions do not need very high resolution scanning of the investigated document and explore the multidirectional, multiscale and low-level gradient texture patterns yielded by printing devices. The main contributions of this work are: (1) the description of printed areas using multidirectional and multiscale co-occurring texture patterns; (2) description of texture on low-level gradient areas by a convolution texture gradient filter that emphasizes textures in specific transition areas and (3) the analysis of printer patterns in segments of interest, which we call frames, instead of whole documents or only printed letters. We show by experiments in a well documented dataset that the proposed methods outperform techniques described in the literature and present near-perfect classification accuracy being very promising for deployment in real-world forensic investigations.

1. Introduction

The massive use of printers is now giving rise to questions about authenticity of printed documents. Today, unknown contractual terms can be added easily and a forged correspondence can be linked to an innocent. Also, documents related to crimes such as child pornography photos, fake travel tickets, terrorist plots, fake money, pirated copies of books and illegal drug selling accounting are constantly printed everywhere. Identifying the source printer of these documents is an important clue to pinpoint their owner.

To understand the clues given by these printers and use them to identify the printer source, it is paramount to understand how they work. One of the most used printer devices currently is the Laser Printers (LPs). These devices work by using electromagnetic energy created by a laser canon onto fix the toner to a paper. As described by Chiang et al. [1], identifying the source of a printed document involves two strategies: the first, known as finding the *extrinsic signatures*, is an active procedure and involves embedding a signature on the printed page. This is done by modifying the document before it is sent to the printer or by encoding identification information, such as the device's serial number. The second, and most used way of identifying the source printer, is finding the *intrinsic signatures*. This is a passive strategy which is used on a scanned version of the document. It requires an understanding and modeling of the device mechanism to find clues in the printing pattern that are present on the scanned image. Most techniques applied to identification of laser printers take into account an artifact commonly caused by the printer manufacturing process: the *banding*. These techniques investigate how the texture in letters of text behaves and link it to a specific printer. Most of them [2–7] select a common letter in the text and describe the texture on it.

In this paper, we propose three solutions aimed at the identification of the source printer of a document that explore these intrinsic signatures. The proposed solutions do not need very high resolution digital versions of documents and take into account that this problem requires multidirectional and multiscale analysis, because of different printing patterns yielded by different manufacturing processes. The proposed solutions described in this paper are:

1. Two descriptors, based on multidirectional and multiscale properties of texture micro patterns. These descriptors are

* Corresponding author at: Av. Santa Isabel 1125, Barão Geraldo, Campinas, São Paulo CEP 13084643, Brazil. Tel.: +55 1933080321.
E-mail addresses: anselmoferreira@ic.unicamp.br, anselmo.ferreira@gmail.com (A. Ferreira).

http://dx.doi.org/10.1016/j.forsciint.2014.11.030
0379-0738/

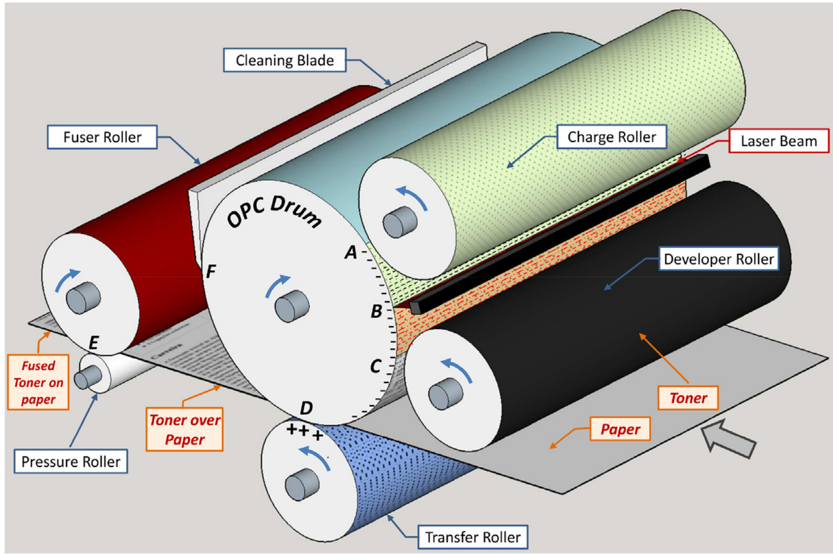

Fig. 1. Steps of LP workflow: (A) charging, (B) exposure, (C) development, (D) transfer, (E) fusing, (F) cleaning.

applied in text letters or regions of interest. These descriptors are focused on the inner part of printed letters.

2. Another descriptor, here described as the convolution texture gradient filter (CTGF). The CTGF is built as a histogram of low-level gradient filtered textures. We use filters of one or more scales, which are focused on filtering inner and outer parts of printed letters and figures.
3. The investigation of texture artifacts on segments of a document, called frames. With this approach, we can recognize the printing source of a document even if parts of it are unavailable or with problems. If the whole document is available, we can use this approach allied with fusion strategies, which provides even more reliable results.

We perform experiments in a well documented printed document benchmark, which is a very difficult one containing different letter sizes, styles and figures.

The dataset was created within the scope of this work and is freely available through FigShare [1] along with the source code of the proposed methods available on GitHub. [2]

Finally, we show that the presented techniques are very competitive and have important properties when compared to other ones in the literature.

2. How laser printers work

To understand the intrinsic signatures and how it can be detectable for laser printer attribution, the Laser Printer process must be known first. Laser printers basically use the attraction of opposite electrical charges in the printing process. The main component of the LP system is a revolving drum or cylinder. This assembly is made of photo-conductive material, which is discharged by light photons of a laser beam. As described by Chiang et al. [1], Laser Printers works in six steps:

1. **Charge:** the revolving drum that rotates at a constant angular velocity is positively charged by a roller or wire having electrical current moving through it.
2. **Exposition:** as the drum revolves, the printer uses a laser beam reflected by a mirror to discharge certain points on the drum, which will be the letters and images to be printed.
3. **Development:** after the pattern has been created on the drum, the printer coats these areas with positively charged ink (or toner) particles.
4. **Transferring:** the printing is done by moving the positive toner particles on the drum to a sheet of paper negatively charged, which moves on a belt below it.
5. **Fusion:** a fuser uses pressure and heat to fuse toner onto the paper.
6. **Cleaning:** to print the next page, a blade cleans the drum to eliminate any residual toner.

Fig. 1 depicts how LPs work.

In black and white printed documents, colors are represented by grayscale using standard conversion formulas to preserve visual perception characteristics, such as luminance. As laser printers have only one ink that is darkest black, grayscale intermediate tonalities are achieved using density variation from black and white small areas (above human eyes resolution) using halftones. Halftones are an old printing technique consisting in black small dots with different diameters over a white surface, which creates a grayscale visual illusion. Common halftone algorithms are error diffusion [8] and clustered dot halftone [9].

As laser printers are electromechanical devices with moving parts, there are many small physical differences on LPs such as motor drifting and gear precision that can be seen on printed pages. These informations patterns can be used as intrinsic signatures of these devices. The *banding* [10,11] is an artifact detectable on scanned printed images that can be used to identify the source printer. It is defined as nonuniform light and dark lines perpendicular to direction in which the paper moves through the printer.

Different printing devices have almost unique banding frequencies, depending of model and brand. To recognize this property, several techniques proposed in the printer attribution literature analyze the frequency domain of one dimensional signal of large halftone regions of the document. Studying the Fourier transform of the printed material can be useful to identify the frequencies at which printers work. But those features are only detected at higher resolutions, where variations on distances of halftones can be measured properly. In text documents, whereby only the black color is visible, the absence of halftone areas makes it difficult to perform the Fourier analysis of a signal. In this case, the banding can be seen as textures in specific characters and

[1] http://dx.doi.org/10.6084/m9.figshare.1263501
[2] https://github.com/anselmoferreira/printer_forensics_source_code