

COLOR LASER PRINTER IDENTIFICATION USING PHOTOGRAPHED HALFTONE IMAGES

*Do-Guk Kim**, *Heung-Kyu Lee*†*

* Graduate School of Information Security, KAIST † Department of Computer Science, KAIST

ABSTRACT

Due to the spread of color laser printers to the general public, numerous forgeries are made by color laser printers. Printer identification is essential to preventing damage caused by color laser printed forgeries. This paper presents a new method to identify a color laser printer using photographed halftone images. First, we preprocess the photographed images to extract the halftone pattern regardless of the variation of the illumination conditions. Then, 15 halftone texture features are extracted from the preprocessed images. A support vector machine is used to be trained and classify the extracted features. Experiments are performed on seven color laser printers. The experimental results show that the proposed method is suitable for identifying the source color laser printer using photographed images.

Index Terms— Digital forensics, Printer identification, Discrete Fourier Transform

1. INTRODUCTION

The price competitiveness of the color laser printer has been promoted by the advancement of manufacturing technologies. Therefore, people can easily acquire high quality printed documents by using color laser printers. This phenomenon has brought not only benefits in everyday life, but also negative effects associated with forgery. Document forgeries have been reported constantly and it is hard to distinguish forged documents from genuine documents with the naked eyes. To prevent forgeries, it is important to identify the source printer of printed documents.

Several methods to identify the source printer have been presented in previous studies. Mikkilineni et al. [1] suggested a printer identification method using the banding frequency extracted from the halftone printed images. Each printer has a different banding frequency so that it can be used as a fingerprint of the source printer. However, since it is difficult to extract the banding frequency from text documents, they suggested another printer identification method using gray-level co-occurrence features [2]. On the other hand, Deng et al. [3] introduced a printer identification method that uses the distance transform features of monochrome laser printed text documents. Choi et al. [4] analyzed a color laser printed im-

age and suggested noise features from the statistical analysis based on a discrete wavelet transform. Bulan et al. [5] used geometric distortion of the halftone dots as a feature of the source printer. Ryu et al. [6] presented a feature by analyzing the halftone texture of color laser printed images. They constructed histograms from angle values of line components in halftone patterns for each CMYK color channel and applied it as a feature. Tsai et al. [7] extracted texture features using a gray-level co-occurrence matrix and a discrete wavelet transform. Then, they used feature selection techniques to find the optimal feature subset.

Existing printer identification techniques use a scanner to acquire images of the printed material. Most scanners are not portable, so it is hard to prevent damage caused by forgery effectively using existing printer identification techniques. The present paper suggests a color laser printer identification method that uses photographed images of the printed material. In the digital forensic field, there has been no study on color laser printer identification using photographed images. Since photographing devices, including smartphones, are portable and widespread, the proposed technique can be an effective countermeasure against forgeries such as banknote forgery. The proposed method analyzes halftone images in the discrete Fourier transform (DFT) domain. After the analyzing, 15 features are extracted and used to train the support vector machine (SVM). The trained SVM is used to identify the source printer. Experiments to verify the performance of the proposed method were conducted, and a smartphone was used to acquire images.

The rest of the present paper is organized as follows. The photographing environment and halftone texture are analyzed in Section 2. The overall process of the proposed method is presented in Section 3. Experimental results are summarized in Section 4. Finally, the conclusion and future work are presented in Section 5.

2. BACKGROUND

2.1. Analysis of the photographing environment

There are several differences between a scanning environment and a photographing environment. Most of all, difference in the uniformity of the illumination is the most important. The



Fig. 1. Close-up photographed halftone image

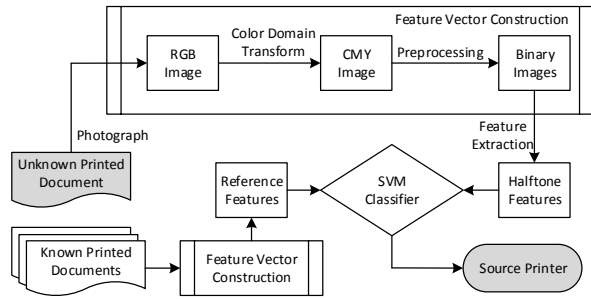


Fig. 2. Overall process of color laser printer identification

intensity of the illumination is uniform in scanned images. Therefore, it is possible to extract high frequency components, such as color noises, and identify source printers based on them [4, 7].

On the other hand, it is hard to extract noise features from photographed images because the intensity of the illumination is not uniform. Color noise features are based on the fact that the printed color of the specific color is different in each printer. Nonuniform illumination changes the color of pixels according to their position in the images even when the images are printed from the same printer. Therefore, it is hard to extract color noise features exactly from photographed images.

The other sort of features that can be used to identify a color laser printer is halftone texture features. In scanning, a halftone texture image can be acquired by high resolution scanning. Similarly, a halftone texture image can be acquired by close-up photography. An example of a close-up photographed halftone image is shown in Fig. 1. Halftone dots appear clearly in the close-up photographed halftone image. In the proposed method, halftone textures of cyan, magenta, and yellow toner are analyzed to extract features.

2.2. Halftone texture features

Currently known halftone texture feature [6] is not sufficient to identify the source color laser printer precisely. It is hard to distinguish color laser printers precisely with only the known feature. Therefore, we analyzed halftone texture to find addi-

tional features to identify the source color laser printer.

There are three sorts of features of the halftone texture: printing angle, printing resolution, and detail texture. The feature printing angle was used in [6]. In the proposed method, all three sorts of features are used, and the accuracy of the proposed system is improved then using only one sort of feature. For each color channel, we used one feature for the printing angle, one feature for the printing resolution, and three statistical features for the detail texture.

3. PRINTER IDENTIFICATION METHOD

3.1. Overall process

The overall process of the proposed method is shown in Fig. 2. First, the photographed RGB image is transformed to a CMY image. Each color channel in the CMY image is pre-processed to extract halftone dots, and halftone texture binary images for each color channel are obtained. Then, halftone features are extracted from the binary images in the discrete Fourier transform (DFT) domain. In the proposed system, five features for each CMY color channel are used for identification: the printing angle feature, the printing resolution feature, and three statistical features of DFT coefficients. Therefore, a total 15 features are used to train the SVM classifier. Reference features that are extracted from the known printed documents are used in training. After the training, classification is carried out with trained SVM classifier.

3.2. Preprocessing

The preprocessing consists of two steps: adaptive image thresholding and morphological opening. Adaptive image thresholding is used for extracting halftone dots regardless of the illumination variation. As shown in Fig. 3(a), the intensity of illumination is not uniform in photographed images. Fig. 3(b) shows that the conventional image thresholding cannot extract halftone dots from the image. The conventional image thresholding uses a global threshold for all pixels in the entire image so that it does not work well when the intensity of illumination is not uniform. However, adaptive image thresholding changes the threshold dynamically over the entire image [8].

The details of the adaptive image thresholding process are summarized as follows. For each pixel, the mean of the pixel values in the 20×20 block that is centered at the corresponding pixel is calculated. Then, the value that is 0.96 times of the mean is used for the threshold of the corresponding pixel. After the thresholds for all the pixels are determined, the pixels are binarized by comparison with the determined thresholds. If the pixel value is higher than the threshold, the pixel is binarized to one which means the halftone dot pixel. Otherwise, the pixel is binarized to zero which means the background pixel. Fig. 3(c) shows the extracted halftone dots by the adaptive image thresholding.

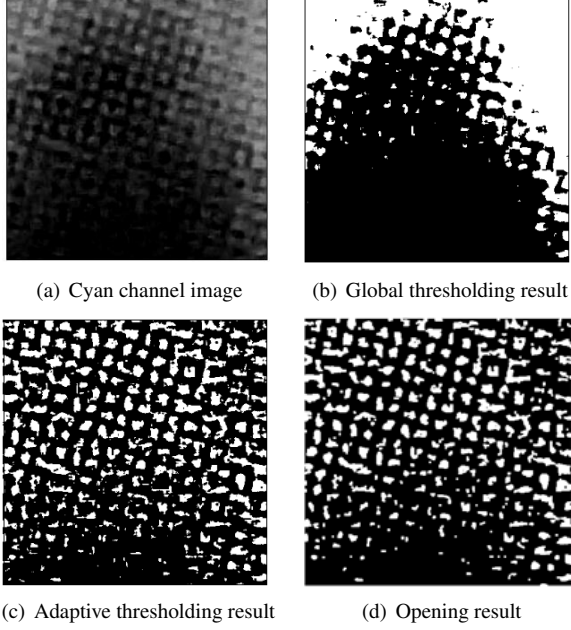


Fig. 3. Preprocessing

The extracted binary image contains noises that can affect the analysis result. Therefore, they are removed by morphological opening. The morphological opening is the morphological dilation of the morphological erosion:

$$A \circ B = (A \ominus B) \oplus B, \quad (1)$$

where \circ denotes opening, \ominus denotes erosion, and \oplus denotes dilation [9]. The noises are removed by erosion, and eroded halftone dots are recovered by dilation. An example of the fully preprocessed image is in Fig. 3(d).

3.3. Feature extraction

In the feature extraction process, five features for each CMY color channel are extracted: the printing angle feature, the printing resolution feature, and three statistical features. These features are extracted in the discrete Fourier transform (DFT) domain. In the ideal case, there is one halftone pattern in a binary image of one color channel. However, the observation showed that two halftone patterns can exist in a binary image of one color channel. In this case, it is hard to analyze the halftone texture in the global DFT domain because the peak of one halftone pattern can be masked by the other halftone pattern. Therefore, the binary images are analyzed in the blockwise DFT domain.

The blocks are overlapped as half of their size which was set as 384×384 . Then, the DFT of each block is computed, and the log-scale DFT magnitude is filtered with the high-pass

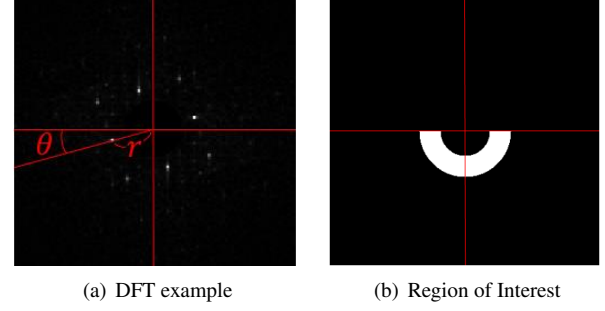


Fig. 4. Feature extraction

filter $H(r)$:

$$H(r) = \begin{cases} 0, & \text{if } 0 \leq r \leq 10 \\ 1, & \text{otherwise,} \end{cases} \quad (2)$$

where r is the distance from the center. After the high-pass filtered log-scale DFT magnitudes of every block are computed, the block that has the maximum peak value of all the blocks is selected for halftone texture analysis. That block is appropriate for analysis because it is the block in which the periodic halftone pattern appears most clearly.

The selected block is analyzed in the DFT domain, and an example is shown in Fig. 4(a). The θ represents the printing angle, and the r represents the printing resolution. The θ is computed and set as the first printing angle candidate of the corresponding color channel. As described before, there can be halftone patterns of two color channels in one binary image. Therefore, the second printing angle candidate is computed by the following method.

For the remaining blocks, the printing angles are computed in the blocks of which the peak value is more than 90% of the maximum peak value. If there is a printing angle different from the first candidate by more than 10° , it is set as the second printing angle candidate. If there is no such angle, the corresponding color channel has only one candidate.

Printing angle candidates for all CMY color channels are computed by this process. There can be three to six candidates for three color channels, and the printing angle for each channel must be determined before extracting the remaining features. The printing angles are determined by the process described below.

- **Case 1: Three candidates**

The printing angle of each color channel is directly determined as the candidate of each channel.

- **Case 2: Four candidates**

In this case, the printing angles of two color channels are directly determined but there are two candidates for the other color channel. The candidates are compared with the determined printing angles, and the duplicate one is eliminated from the candidates. The remaining angle is determined as the printing angle.

- **Case 3: Five candidates**

The printing angle of the only one color channel is directly determined in this case. The determined printing angle is compared with the other four candidate angles.

- If there is a candidate angle the same as the determined angle, the other candidate angle of the same color channel is determined as the printing angle. Then, there are four candidates and it becomes the same as Case 2.
- If there is no candidate angle the same as the determined angle, the printing angle of the one color channel is determined heuristically. The observation showed that the probability for the first candidate angle to be a real printing angle is high in order of magenta, cyan, and yellow color channel. Therefore, the first candidate of the magenta channel is determined as the printing angle if there are two candidates in the magenta channel. Otherwise, the first candidate of the cyan channel is determined as the printing angle. The printing angle of the last color channel can be determined by the Case 2 method.

- **Case 4: Six candidates**

The first candidate of the magenta channel is determined as the printing angle. Then, there are five candidates and it can be solved the same as Case 3.

The printing angles of three color channels are determined by the described method. Then, other features are extracted from the blocks from which the printing angles are extracted. One of them is the printing resolution feature which can be obtained by computing r in Fig. 4(a). That value is the frequency of the dominant periodic signal in the halftone pattern. Thus, it is the feature related to the halftone printing resolution.

The remaining three features for each color channel are statistical features: standard deviation, skewness, and kurtosis [10]. These features are extracted from the DFT coefficients in specific band B :

$$B = \{r | r_p - 5 \leq r \leq r_p + 5\}, \quad (3)$$

where r_p is the distance between the peak and the center. An example of the region of interest is shown in Fig. 4(b). Half of the B are used for statistical analysis due to the diagonal symmetry of the DFT magnitude. Finally, five features for each color channel are extracted by the described process.

4. EXPERIMENTAL RESULTS

In this section, experimental results to verify the performance of the proposed system are presented. A total of seven printers from three brands were used for the experiments. The color laser printers are listed in Table 1. For each printer, 238 images were photographed, and half of them were used for training while the other half were used for classifying. In

Label	Brand	Model
H1	HP	HP 4650
X1	Xerox	700 Digital Color Press
X2	Xerox	Docu Centre C6500
X3	Xerox	Docu Centre C450
K1	Konica Minolta	Bizhub Press C8000
K2	Konica Minolta	Bizhub Press C280
K3	Konica Minolta	Bizhub Press C280

Table 1. A list of printers used in experiments

total, 1666 images were used in the experiments. The test images were photographed from the LG Optimus G smartphone equipped with a Kakuyo KC-1 close-up lens. The size of images was the 3120×4208 . The photographing distance and angle were maintained equally during the photographing. The experiments were conducted with two identification methods using the same image set: the proposed method and Ryu’s method [6].

First of all, an experiment to verify whether the proposed method could identify the brand of the source printer was carried out. The results are summarized in Table 2. The average accuracy of brand identification was 94.4%. These results confirm that the proposed method can identify the brand of the source printer with high accuracy.

On the other hand, the average accuracy of Ryu’s method was 70.3%, despite the halftone printing angles of the three brands being different from each other. That means that Ryu’s method cannot precisely extract the feature from the photographed images. Since Ryu’s method concerned only scanned input images, the result shows that the method is not robust to nonuniform illumination and noise.

Next, a printer device identification experiment was conducted to confirm the performance of the proposed method. Table 3 shows the results of identifying seven different color laser printers. The average accuracy of brand identification was 76.0%, which exceeds the average accuracy of Ryu’s method (41.5%). It is proven that the proposed method works better with photographed images than the existing method, but it is hard to identify the source printer exactly. The results demonstrate the possibility of identifying a source printer precisely by improving the proposed method.

5. CONCLUSION

In the present paper, a source color laser printer identification method using photographed images was proposed. To our knowledge, this is the first attempt to identify a source printer using photographed images in the digital forensic field. To identify the source printer, we suggested 15 halftone texture features that were extracted in the DFT domain. These features were used to train an SVM classifier, and the trained classifier was used to identify the source color laser printer of

		Proposed method(%)			Ryu's method(%)		
Real \		HP	Xerox	Konica	HP	Xerox	Konica
HP		90.0	9.2	0.8	79.0	10.1	10.9
Xerox		2.5	94.1	3.4	7.6	72.3	20.1
Konica		0.0	0.8	99.2	14.2	26.1	59.7

Table 2. Brand identification results

		Proposed method(%)							Ryu's method(%)						
Real \		H1	X1	X2	X3	K1	K2	K3	H1	X1	X2	X3	K1	K2	K3
H1		84.1	8.4	2.5	0.8	3.4	0.0	0.8	52.1	10.9	3.4	4.2	21.0	6.7	1.7
X1		2.5	67.2	25.3	2.5	0.0	0.8	1.7	14.3	37.8	11.8	6.7	5.9	10.1	13.4
X2		0.0	33.6	54.6	2.5	3.4	1.7	4.2	7.6	22.7	42.0	6.7	2.5	12.6	5.9
X3		0.0	5.9	1.7	73.9	0.8	2.5	15.2	2.5	8.4	10.1	49.5	7.6	16.0	5.9
K1		1.7	5.0	3.4	0.8	76.5	9.2	3.4	20.2	9.2	5.9	5.0	40.4	12.6	6.7
K2		0.0	0.8	2.5	0.0	18.5	78.2	0.0	14.3	16.8	10.1	11.8	10.1	23.5	13.4
K3		0.0	0.0	0.0	2.5	0.0	0.0	97.5	6.7	11.8	5.9	14.3	5.9	10.1	45.3

Table 3. Printer identification results

unknown documents. The experimental results showed that the proposed method was appropriate for identifying source printers using photographed images.

For future work, we concentrate on finding additional features to improve the identification accuracy. Then, we will test the improved method for various photographing environments. Finally, future work for acquiring robustness for the various photographing distances and angles will be performed.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. 2012R1A2A1A05026327)

REFERENCES

- [1] A. K. Mikkilineni, G. N. Ali, P. J. Chiang, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "Signature-embedding in printed documents for security and forensic applications," in *Proc. of the SPIE*, 2004, vol. 5306, pp. 455–466.
- [2] A. K. Mikkilineni, P. J. Chiang, G. N. Ali, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "Printer identification based on graylevel co-occurrence features for security and forensic applications," in *Proc. of the SPIE*, 2005, vol. 5681, pp. 430–440.
- [3] W. Deng, Q. Chen, F. Yuan, and Y. Yan, "Printer identification based on distance transform," in *Proc. of the ICINIS*, 2008, pp. 565–568.
- [4] J. H. Choi, D. H. Im, H. Y. Lee, and H. K. Lee, "Color laser printer identification by analyzing statistical features on discrete wavelet transform," in *Proc. of the ICIP*, 2009, pp. 1505–1508.
- [5] O. Bulan, M. Junwen, and G. Sharma, "Geometric distortion signatures for printer identification," in *Proc. of the ICASSP*, 2009, pp. 1401–1404.
- [6] S. J. Ryu, H. Y. Lee, D. H. Im, J. H. Choi, and H. K. Lee, "Electrophotographic printer identification by halftone texture analysis," in *Proc. of the ICASSP*, 2010, pp. 1846–1849.
- [7] M. J. Tsai and J. Liu, "Digital forensics for printed source identification," in *Proc. of the ISCAS*, 2013, pp. 2347–2350.
- [8] N. Milstein, "Image segmentation by adaptive thresholding," pp. 1–38, 1998.
- [9] W. K. Pratt, *Digital image processing*, John Wiley & Sons Inc, 2001.
- [10] A. J. Hayter, *Probability and statistics for engineers and scientists*, Cengage Learning, 2013.