

Printer Identification from Micro-metric Scale Printing

Thong Q. Nguyen *, Y. Delignon *, L. Chagas #, F. Septier *

* *Institut Mines Telecom, LAGIS UMR CNRS 8146, TELECOM Lille1, rue Guglielmo Marconi, 59653 Villeneuve d'Ascq, France*

ILGP2 UMR CNRS 5518, 461 rue de la papeterie, 38402 Saint-Martin d'Herès, France

Abstract—Microscopic analyses of paper printing show some regularly spaced dots whose the shape depends on the technology and the tuning of the printer as well as on the paper properties. The modeling and the identification of paper and ink interactions are required for qualifying the printing quality, for controlling the printing process and also for authentication issues. In this paper, we propose to model the micrometric scan of document printing by a binary response model whose the parameters depend on the location and the shape of dots. A maximum likelihood identification algorithm is provided, its performance is assessed through simulations and true data. Furthermore, we illustrate the benefice of a such model and estimation algorithm in the case of authentication of printer from micro-tag made of four dots.

I. INTRODUCTION

Preventing and discouraging unauthorized printed materials is the expected achievement of the authentication methods. Papers dealing with this subject can be categorized in two strategies. The first one consists in embedding an extrinsic signature such as secure tags in a printing [1], [2] while the second one, which is the purpose of this paper, consists in characterizing the intrinsic features of the printer [3]. As examples of intrinsic features, let us mention signature consisting on the banding artifact coming from fluctuation of the optical photoconductor angular velocity in case of laser printers [4], dimple effect which are specific to inkjet printers or also texture features [5] [6] in case of both technologies, or the analysis of the unique print quality signature to differentiate one printer technology/supplier from another [7]. The impact of the channel model for authentication systems based on graphical codes was also investigated [8]. Because the digital printing consists in dots regularly spaced according to the resolution of the printer, we are able to consider the dot shape as the intrinsic feature of the printing process. At the microscopic scale, each dot is a random pattern whose shape depends on the technology, the setting of the printer, the ink quality and/or the paper properties. From the statistical point of view, we show that the digital acquisition of these random dots can be modeled as a binary response model based on an exponential power kernel which depends on location parameters as well as shape parameters. Because the latter enable to discriminate the printer, they can be used for authentication purpose. In this paper, we developed a maximum likelihood estimation algorithm of

these parameters and authentication performance is obtained thank to the unsupervised identification algorithm.

The paper is organized as follows. In section II, the exponential power binary response model is described. From a Newton-Raphson based optimization method with constraints, an unsupervised maximum likelihood estimation is developed in section III. Results are given in section IV. First, the performances of the new algorithm are analyzed through simulations. Then, the benefice of a such model and estimation algorithm is illustrated in the case of authentication of printer from micro-tag made of four dots. Conclusion is given in section V.

II. EXPONENTIAL POWER BINARY RESPONSE MODEL

A few documents about image degradation models have been proposed in the literature [9], [10], [11]. All of them depend on the scale of modeling and the involved application. Our work takes part in the micro-metric scale authentication of documents so we are concerned by the modeling of micro-metric scale document printing. As a consequence, large scale distortions such as blurring, dot inversion and non-uniform spreading of the ink are not considered. Moreover, since the gray level variation of inked area is not informative at the micro-metric scale, only the spatial distribution of the ink is taken into account.

An ideal printing is a set of points distributed in the image whose the resolution depends on the printer characteristics. At the micro-metric scale, each dot is a cluster of ink particles whose the shape and the number of particles depend on both the technology of the printer and its tuning. This model had been proposed in our paper [12] for the simpler case. For the following, let's note $U = \{U_s\}$, $s \in S$, a random field made of $N \times N$ pixels that take their values in $\{0, 1\}$, where 0 and 1 stand respectively for black and white pixels.

U is the superposition of K dots distributed in S . Each dot k can be considered as a set of $N \times N$ independent random variables distributed according to a Bernoulli distribution. From a statistical point of view, U is a binary response model [13] with column and row indexes as explanatory variables and an inverse link function $p_{s,k}$:

$$P(U_{s,k} = 0) := p_{s,k} = nf(s|\theta_k) \quad (1)$$

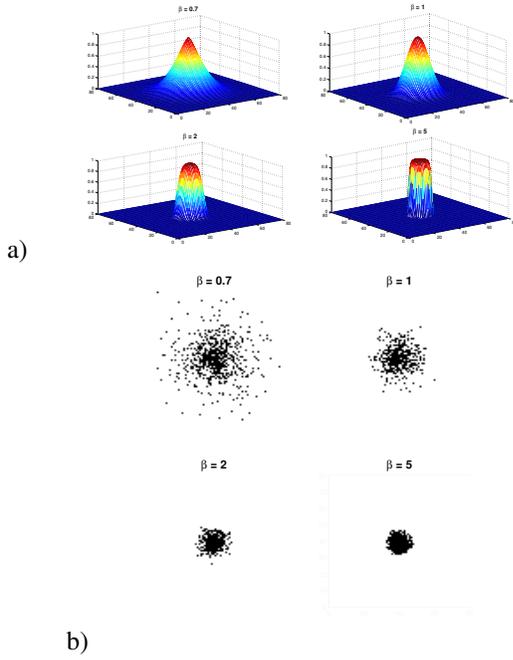


Figure 1. a) $p_{s,k}$ with $\sigma^2 = 50$, $n = 2^{\frac{1}{\beta}} \pi \sigma^2 \Gamma(1 + \frac{1}{\beta})$ and $\beta \in \{0.7; 1; 2; 5\}$ b) realisations of the binary response model with the same parameters.

when f has the properties of a probability density function (pdf), the sum of $p_{s,k}$ over S is the expected number of black pixels n . In order to encompass a large variety of shape, $f(s|\theta_k)$ is chosen as the exponential power density of mean μ_k , scale parameter Σ and shape parameter β . $f(s|\theta_k) = EP_2(s|\mu_k, \Sigma, \beta)$ is defined as follows :

$$f(s|\mu_k, \Sigma, \beta) = \frac{\exp\left\{-\frac{1}{2}\left((s - \mu_k)' \Sigma^{-1}(s - \mu_k)\right)^\beta\right\}}{\pi|\Sigma|^{\frac{1}{2}} 2^{\frac{1}{\beta}} \Gamma(1 + \frac{1}{\beta})} \quad (2)$$

Σ , β and n depend on the printer, its technology and its tuning. $\{\mu_k\}$ is the location of the dot k , n plays the role as the tone of ink while the scale parameter Σ decides the largeness of the droplets and also their shape. On the other hand, β controls the density of the black particles, it is more uniform with larger β .

For simplicity, the exponential power distribution $p_{s,k}$ is assumed to be circular, i.e. $\Sigma = \sigma^2 I_2$ with I_2 the 2-dimension identity matrix. Moreover $n \in \left[0, 2^{\frac{1}{\beta}} \pi \sigma^2 \Gamma(1 + \frac{1}{\beta})\right]$ in order to constraint $p_{s,k}$ to be a probability. At a site s , the Bernoulli parameter $p_{s,k}$ depends on its distance from the center of the dot μ_k . Fig. 1 shows both $p_{s,k}$ and dots in case of four different values of β . Let us note that $f(s|\theta_k)$ is the normal density if $\beta = 1$.

Turning now to a print made of many dots. Let us denote U a such binary response model and U_s the pixel at site s . $U_s = 1$ if no ink particle emanating from one of the K dots impinges the site s , that is $U_{s,k} = 1, \forall k \in \{1, \dots, K\}$. On another hand, a pixel U_s is 0 if at least one $U_{s,k} = 0, \forall k \in$

$\{1, \dots, K\}$. As a consequence, U_s can be modeled as :

$$U_s = \prod_{k=1}^K U_{s,k} \quad (3)$$

so that

$$P(U_s = 0) = 1 - \prod_{k=1}^K (1 - p_{s,k}) \quad (4)$$

Assuming that the random field is composed of independent pixels, the resulting distribution is given by :

$$p(u) := \prod_s \left(1 - \prod_{k=1}^K (1 - p_{s,k})\right)^{\mathbb{I}_{S_0}(s)} \left(\prod_{k=1}^K (1 - p_{s,k})\right)^{\mathbb{I}_{S_1}(s)} \quad (5)$$

with $S_i = \{s|u_s = i\}$ and $\mathbb{I}_A(\cdot)$ the indicator function of the set A . The parameters of the binary response model are denoted $\theta = \{\mu_1, \dots, \mu_K, \sigma^2, \beta, n\}$.

III. ESTIMATION

In this section, the maximum likelihood estimator is described. Since U is made of independent random variables, the log-likelihood of all pixels is given by :

$$L(\theta) = \sum_{s \in S_0} \ln \left(1 - \prod_{k=1}^K (1 - p_{s,k})\right) + \sum_{s \in S_1} \sum_{k=1}^K \ln(1 - p_{s,k}) \quad (6)$$

where σ^2, β and n have to be consistent with:

$$g_1(\theta) = n - 2^{\frac{1}{\beta}} \pi \sigma^2 \Gamma\left(1 + \frac{1}{\beta}\right) \leq 0 \quad (7)$$

$$g_2(\theta) = -n \leq 0 \quad (8)$$

$$g_3(\theta) = -\beta \leq 0 \quad (9)$$

Because the maximum likelihood is not tractable, we resort to an iterative algorithm consisting in successive estimation of each parameters. The initialization of the algorithm is a rough approximation of θ obtained by the k-mean++ method [14] for the means. The parameters n , σ^2 and β are respectively initiated by.

$$n = \frac{1}{K} \sum_{s \in S} \delta_{u_s} \quad (10)$$

$$\sigma^2 = 0.8 \times \frac{n^g}{\pi} \quad (11)$$

$$\beta = \frac{\log 2}{\log n^g - \log(\pi \sigma^{2g})} \quad (12)$$

with δ the Kronecker symbol.

At each iteration of the algorithm, the mean are estimated from the quasi Newton algorithm (QNA) based on the Broyden-Fletcher-Golfarb-Shanno (BFGS) method. β , σ and n are estimated from the QNA applied to the augmented Lagrangian method [15]. At each iteration i , the estimation of σ^2 consists in minimizing the dual function $C(\sigma^2, \lambda, \eta)$, while the other parameters are obtained from their previous estimation :

$$C(\sigma^2, \lambda, \eta) = -L(\sigma^2) + \lambda g_1(\sigma^2) + \frac{\eta}{2} \max(0, g_1(\sigma^2))^2 \quad (13)$$

where $\lambda \geq 0$ and $\eta \geq 0$ are respectively the Lagrangian multiplier and the penalization parameter. The initialization of this step is such that $\eta^{(0)} > 1$ and $\lambda^{(0)} \geq 0$. The iteration j of the minimisation algorithm proceeds as follows:

- **New approximation:** Minimizing the augmented Lagrangian $C(\sigma^2, \lambda^{(j)}, \eta^{(j)})$ with respect to $\sigma^{2(j)}$ from the QNA with BFGS method.
- **Stopping criterion:**

$$\begin{aligned} | -\partial_{\sigma^2} L(\sigma^{2(j)}) + \lambda^{(j)} \partial_{\sigma^2} g_1(\sigma^{2(j)}) | &< \epsilon, \\ | \sigma^{2(j)} g_1(\sigma^{2(j)}) | &< \epsilon \end{aligned} \quad (14)$$

- **Update:**

$$\begin{aligned} \lambda^{(j+1)} &= \max\{0, \lambda^{(j)} + \eta^{(j)} g_1(\sigma^{2(j)})\} \\ \eta^{(j+1)} &= c\eta^{(j)} \end{aligned}$$

with c is usually chosen in the range $[4, 10]$. At each iteration, the estimation of the other parameters follows the same method, only the dual function is modified to take into account the specific constraints of each parameter.

The clusters blind identification algorithm is summarized in Al. 1.

Algorithm 1 Exponential power binary response model identification algorithm

Initialization

Initialize $\theta^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)}, \dots, \mu_K^{(0)}, \sigma^{2(0)}, \beta^{(0)}, n^{(0)})$ by the K-mean++ algorithm, equations (10), (11) and (12)

Choose ϵ small.

while $|L(\theta^{(j-1)}) - L(\theta^{(j)})| > \epsilon$

Estimate $\{\mu_k^{(j)}\}$ by BFGS method using $(\sigma^{2(j-1)}, \beta^{(j-1)}, n^{(j-1)})$

Estimate $\sigma^{2(j)}$ by augmented Lagrangian method with $(\{\mu_k^{(j)}\}, \beta^{(j-1)}, n^{(j-1)})$ and constraint (7)

Estimate $\beta^{(j)}$ by augmented Lagrangian method with $(\{\mu_k^{(j)}\}, \sigma^{2(j)}, n^{(j-1)})$ and constraints (7), (9)

Estimate $n^{(j)}$ by augmented Lagrangian method with $(\{\mu_k^{(j)}\}, \sigma^{2(j)}, \beta^{(j)})$ and constraints (7), (8)

end while

IV. EXPERIMENTAL RESULTS

We highlight in the first subsection the estimation performances of the cluster blind identification algorithm. In the second subsection, we deal with authentication performance.

A. Simulated images

The performances of the identification algorithm are assessed for various shapes, scales, relative position of clusters and various expected number of particles per cluster. $N = 100$ simulations of 100×100 images of 2 dots are obtained for $\sigma^2 = 50, \beta \in \{2, 3, 5\}$, and $n \in \{50, 100, 165\}$. The accuracy is analyzed through the bias and the standard deviation of the estimator.

Tab. I and Tab.II show the performances of respectively the location estimator (5) and σ^2 . The larger n and β are, the more accurate is the algorithm. As it can be guessed, estimators are more accurate with higher value of β and the expected number of particles n firstly because particles are more concentrated around the centers with larger β (see Fig. 1) and secondly, the mean number of particle is higher.

n	50		100		165	
β	Avg.	Std.	Avg.	Std.	Avg.	Std.
2	0.82	0.33	0.55	0.26	0.39	0.24
3	0.63	0.40	0.40	0.15	0.31	0.26
5	0.51	0.33	0.34	0.38	0.21	0.11

Table I
THE AVERAGE AND STANDARD DEVIATION OF THE DISTANCES BETWEEN ESTIMATED CENTERS AND TRUE ONES.

n	50		100		165	
β	Bias	Std.	Bias	Std.	Bias	Std.
2	6.72	11.54	2.36	6.49	1.34	3.93
3	2.03	7.13	0.66	3.85	0.41	2.09
5	0.31	4.84	0.26	3.03	-0.05	1.98

Table II
THE BIAS AND STANDARD DEVIATION OF THE ESTIMATORS OF σ^2 .

β	2		3		5	
n	Bias	Std.	Bias	Std.	Bias	Std.
50	0.74	1.98	0.90	2.28	5.64	5.55
100	0.16	0.34	0.21	0.55	3.71	2.5
165	0.08	0.202	0.05	0.34	2.50	0.721

Table III
THE BIAS AND STANDARD DEVIATION OF THE ESTIMATORS OF β .

About the performances of the estimator of β given in table III, the best accuracy is obtained for the largest n . Moreover, the estimator of β is better when it is small. This result is explained from the fact that two kernels are quite similar to each other when their β are large.

n	50		100		165	
β	Bias	Std.	Bias	Std.	Bias	Std.
2	0.8842	4.46	0.8336	5.69	1.1884	6.18
3	0.33	4.17	0.8776	5.19	1.4147	5.34
5	-0.1923	4.25	-0.0178	5.32	1.9496	6.54

Table IV
THE BIAS AND STANDARD DEVIATION WHILE APPROXIMATING n .

The bias and the standard deviation of the estimator of n are given in table IV. It seems that the estimator is less accurate with large n . Nevertheless, when focusing on the relative error, the algorithm performs better with a lot of particles, which is also true for estimating other parameters.

B. Authentication

In this section, printer authentication is performed from micro-tag consisting in four printed dots. In this experiment, we have considered 40 tags printed on a same paper with the same ink, the half coming from the Ricoh Aficio MP 6001 printer and the remain coming from the Ricoh Aficio MP C2800 laser printer. The shape parameters (σ^2, n, β) of these 40 tags have been estimated by the algorithm 1. The estimations are displayed in Fig. 2 which shows respectively n vs. σ^2 and σ^2 vs. β . Each printer is a 3D Gaussian cluster in the space of parameters. Their location and shape are respectively given by their mean and their covariance matrix (Tab. V). Basic authentication consists in classifying a tag in one of the two printer classes. In our case, the normal distributions of the shape parameters enable to use the maximum likelihood classification method. After calculation, we obtain 0.035 for the probability of error, that is a probability of correct classification 0.965. This result cannot be generalized to any printers, it only points out the relevance of the exponential power binary response model for printed dots at the microscopic scale.

	β	σ^2	n
MP 6001	3.10	1165.9	3525.2
MP C2800	2.356	449.1	720.67

7.4e-3	0.207	-0.40	0.1	27.6	68.81
0.207	55377	27222	27.6	14681	21329
-0.402	27222	84358	68.81	21329	51628.9

Table V

THE MEAN VALUE OF THE ESTIMATORS AND THEIR COVARIANCE MATRIX, MP 6001 (RIGHT) AND MP C2800 (LEFT).

V. CONCLUSIONS

In this paper, we have proposed an original model for the microscopic scale printing and a maximum likelihood unsupervised identification algorithm. Performances of the estimator have been assessed through simulations and we have show the accuracy of the exponential power based binary response model to describe printing at the microscopic scale. Shape parameters strongly depends on the printer technology, its tuning as well as the paper and the ink properties. Authentication of printers from micro-tags has been performed thanks to the accuracy of both the binary response model and the maximum likelihood algorithm.

REFERENCES

- [1] J. Picard, C. Vielhauer, and N. Thorwirth, "Towards fraud-proof id documents using multiple data hiding technologies and biometrics," in *Electronic Imaging 2004*. International Society for Optics and Photonics, 2004, pp. 416–427.
- [2] R. Villan, S. Voloshynovskiy, O. Koval, and T. Pun, "Multilevel 2 d bar codes: toward high-capacity storage modules for multimedia security and management," in *Proc. SPIE*, vol. 5681, 2005, pp. 453–464.
- [3] P.-J. Chiang, N. Khanna, A. Mikkilineni, M. Segovia, S. Suh, J. Allebach, G. Chiu, and E. Delp, "Printer and scanner forensics," *Signal Processing Magazine, IEEE*, vol. 26, no. 2, pp. 72–83, march 2009.
- [4] G. N. Ali, A. K. Mikkilineni, J. P. Allebach, E. J. Delp, P.-J. Chiang, and G. T. Chiu, "Intrinsic and extrinsic signatures for information hiding and secure printing with electrophotographic devices," in *NIP & Digital Fabrication Conference*, vol. 2003, no. 2. Society for Imaging Science and Technology, 2003, pp. 511–515.
- [5] A. Mikkilineni, O. O. Arslan, P.-J. Chiang, R. M. Kumontoy, J. P. Allebach, G. T.-C. Chiu, and E. J. Delp, "Printer forensics using svm techniques," 2005, pp. 223–226.
- [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 *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, march 2010, pp. 1846–1849.
- [7] J. Oliver and J. Chen, "Use of signature analysis to discriminate digital printing technologies," in *NIP & Digital Fabrication Conference*, vol. 2002, no. 1. Society for Imaging Science and Technology, 2002, pp. 218–222.
- [8] A. T. Phan Ho, B. A. Mai Hoang, W. Sawaya, and P. Bas, "Document authentication using graphical codes: impacts of the channel model," in *Proceedings of the first ACM workshop on Information hiding and multimedia security*. ACM, 2013, pp. 87–94.
- [9] T. Kanungo and Q. Zheng, "Estimating degradation model parameters using neighborhood pattern distributions: an optimization approach," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 4, pp. 520–524, april 2004.
- [10] H. S. Baird, "The state of the art of document image degradation modelling," in *Digital Document Processing*. Springer, 2007, pp. 261–279.
- [11] A. Vongkumhae, J. Yi, and R. Wells, "A printer model using signal processing techniques," *Image Processing, IEEE Transactions on*, vol. 12, no. 7, pp. 776–783, july 2003.
- [12] Q.-T. Nguyen, Y. Delignon, L. Chagas, F. Septier *et al.*, "Printer technology authentication from micrometric scan of a single printed dot," in *IS&T/SPIE Electronic Imaging*, 2014, pp. 1–7.
- [13] J. Nelder and P. McCullagh, "Generalized linear models." London: Chapman and Hall, 1989.
- [14] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.
- [15] P. Pedregal, *Introduction to optimization*. Springer, 2003, vol. 46.

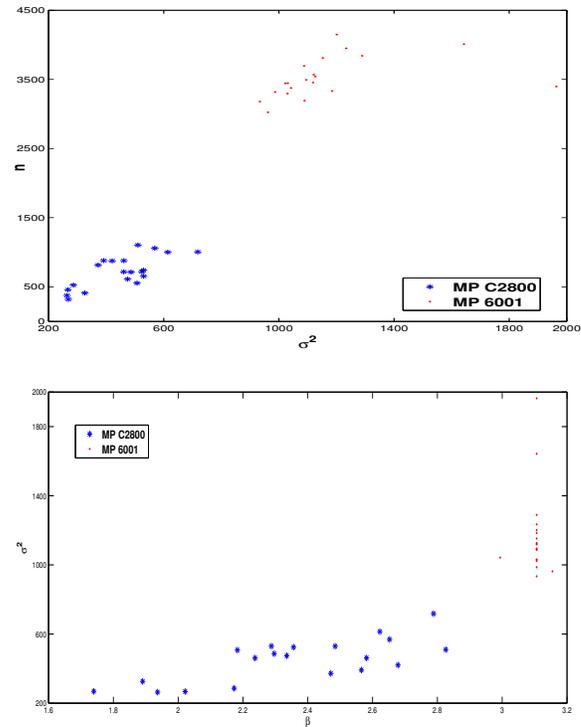


Figure 2. The relation between the estimated parameters.