Using genetic algorithms for spectral-based printer characterization

Silvia Zuffi°, Raimondo Schettini°§, Giancarlo Mauri§♣

°Tecnologie Informatiche Multimediali, ITC-CNR, Milano, Italy
§DISCO, Università degli Studi di Milano Bicocca, Milano, Italy

ABSTRACT

In recent years, many methods have been proposed for the spectral-based characterization of inkjet printers. To our knowledge, the majority of these are based on a physical description of the printing process, employing different strategies to deal with mechanical dot gain and the physical interaction among inks. But our experience tells us that as printing is a physical process involving a large number of effects and unpredictable interactions, it is not unusual to be unable to fit a mathematical model to a given printer. The question becomes, therefore, whether it is feasible, and to what degree, to employ an analytical printer model even if it appears to be incapable of describing the behavior of a given device. A key objective of our work is to obtain a procedure that can spectrally characterize any printer, regardless of the paper and the printer driver used. We consider in fact the printers RGB devices, and incorporate the printer driver operations, even if they are unknown to us, into the analytical model.

We report here our experimentation on the use of genetic algorithms to tune a spectral printer model based on the Yule-Nielsen modified Neugebauer equation. In our experiments we have considered three different inkjet printers and used different kinds of paper and printer drivers. For each device the printer model has been tuned, using a genetic algorithm, on a data set of some 150 measured reflectance spectra. The test set used for model evaluation was composed of 777 samples, uniformly distributed in the RGB color space.

Keywords: spectral-based printer characterization, inkjet printers, Neugebauer, genetic algorithms.

1. INTRODUCTION

In recent years, many methods have been proposed for the spectral-based characterization of inkjet printers\(^1\). Most of them are based on the Neugebauer equation, and employ different strategies to deal with mechanical dot-gain and inks interaction\(^2\)\(^3\)\(^4\). The model’s parameters are usually computed with error minimization procedures on a training set of data; their simultaneous estimation is, in general, not possible. The Neugebauer equation, modified with the Yule-Nielsen coefficient, can be applied successfully to describe printer behavior if the training set of reflectance data exhibits some characteristics of regularity that make it possible to fit the model to the printer simply by using least-square estimated parameters. But this is all too often not the case. For example, the concentration of inks may appear to be wavelength-dependent (Fig. 2), or the ink dots on the paper may have different shapes even with the same ink (Fig. 1), hindering the application of methods based on ink spreading shape analysis\(^5\).

The question becomes, therefore, whether it is feasible, and to what degree, to employ an analytical printer model even if it appears to be incapable of describing the behavior of the device considered.

The main contribution of our work is the proposal of an analytical printer model that can be used regardless of the characteristics of the considered device. The model is based on the Yule-Nielsen Spectral Neugebauer equation, formulated with a large number of degrees-of-freedom, in order to account for dot-gain, inks interaction and printer driver operations. To estimate the model’s parameters we use genetic algorithms.

Genetic algorithms are a general method for solving optimization problems, inspired by the mechanisms of evolution in biological systems. In the basic genetic algorithm (GA), every candidate solution for the optimization problem is represented by a sequence of binary, integer, real, or even more complex values, called an individual or chromosome. A
small number $n$ of individuals (with respect to the cardinality of the whole solution space) is randomly generated as an initial population $P$. Then, the GA iterates a procedure that produces a new population $P'$ from the current $P$, until a given "STOP" criterion is satisfied. Each iteration consists of the following steps:
- fitness evaluation: for every individual $x$ in $P$, the value $f(x)$ of a suitable "fitness" function is computed;
- selection: $n/2$ pairs of individuals are randomly selected from the population $P$; the probability of selection is higher for individuals of greater fitness;
- crossover: two new individuals (sons) are obtained by cutting the two elements of each pair (parents) at a randomly chosen point and interchanging the four parts thus obtained;
- mutation: the value of each position of the elements in $P'$ is changed with a given probability $p_m$.

The main advantages of using the genetic approach are that it allows the simultaneous computation of many parameters, and can deal with irregular training data sets. The disadvantages are that it cannot guarantee an optimal solution, and that it is, in general, also difficult to tune the genetic algorithms’ free parameters.

2. PRINTER MODEL

The model of the printer is based on the well-known Yule-Nielsen Spectral Neugebauer (YNSN) equation. According to the YNSN model, the spectrum of a $N$-inks halftone print is the weighted summation of $2^N$ different colors, called Neugebauer primaries, given by all the possible overprints of inks. The weight of each Neugebauer primary is the area it covers in the halftone cell. The YNSN model for a 4-ink halftone print is:

$$R_{\text{pr int,} \lambda} = \sum_{p=0}^{15} a_p R_{p, \lambda}^n \gamma^n \lambda = 1.8$$

where $R_{\text{pr int,} \lambda}$ is the reflectance of the printed color, $n$ is the Yule-Nielsen factor, $R_{p, \lambda}$ is the reflectance of the $p$-th Neugebauer primary, and $a_p$ is the primary area coverage. In our work, reflectance spectra were sampled at intervals of 40 nm in the range from 400 nm to 680 nm, producing 8 samples.

The Yule-Nielsen factor is an empirical correction that has been applied to the Neugebauer model to account for the optical dot gain effect due to the scattering of light in the paper. The area coverage is the percentage of the halftone cell covered by the Neugebauer primary, and can be computed as expressed in Table 1, where $c = [c, m, y, k]$ represents the concentration of inks for printing a given color.
Table 1. Calculation of the area coverage for each Neugebauer primary given the concentration of inks.

We consider each printer an RGB device, therefore assumptions must be made on how the printer driver computes the CMYK amounts, given an RGB value. Usually the black replaces the “gray component” of the tone. The gray component is a neutral gray that results from removing the smallest of the CMY components together with the amount of the other two colors necessary to form, all together, a neutral gray. This process is named Gray Component Replacement. We model the driver according to Equation 2, where the gray component is composed of equal amounts of cyan, magenta and yellow.

\[
c = c' - U \cdot k, \\
m = m' - U \cdot k, \\
y = y' - U \cdot k, \\
k = R \cdot \min(c', m', y'), \\
0 \leq U \leq 1; R = 1
\]

where \( c' = 1 - r, m' = 1 - g, y' = 1 - b \).

If we consider the print of a single ink on paper, the YNSN model is simplified to:

\[
R_{\text{print}, \lambda} = \frac{1}{c_{\text{ink}}} R_{\text{ink}, \lambda} + \frac{1}{c_{\text{ink}}} R_{\text{ink}, \lambda}
\]

where \( c_{\text{ink}} \) is the ink concentration.

If we measure a ramp of samples with ink concentration ranging from 0 to 1, and compute the effective concentration of ink using the previous equation with \( n=1.0 \), the effective concentration appears to be dependent on the wavelength (Fig. 2), indicating that the Yule-Nielsen factor may be far from being a constant and, consequently, that area coverage in our model must be a function of wavelength.
The real size of a dot printed on a substrate is larger than its theoretical size, due to the spread of the ink on paper. The area coverage, therefore, cannot be computed from the theoretical amount of ink. To account for this phenomenon, a non-linear function relating the theoretical concentration of inks with the effective concentration is employed for any possible overprint of inks. With the aim of using the smallest possible number of parameters for the model, we looked for a function that could describe dot gain and be tuned with only one parameter. The function used is:

$$c_\lambda = \frac{c_t}{(1-C_\lambda)\cdot c_t + C_\lambda}$$  \hspace{1cm} (4)$$

where $c_t$ is the theoretical concentration of ink, computed from RGB values using Equation 2, and $C_\lambda$ is the wavelength-dependent parameter.

Dot gain functions are usually estimated to model the spreading of inks on paper, nevertheless the spread of ink on paper may differ from its spread on a previously deposited ink. Different strategies can be employed to account for this phenomenon\textsuperscript{2,3,5}; we chose to take into account the interaction of the inks, providing a different dot gain function for any overprinting. No assumptions have been made in the model about the sequence of the printing.

Table 2 lists the parameters used to compute the effective concentration of inks, using Equation 4, to obtain the area coverage for any Neugebauer primary according to Table 1.
Table 2. Dot gain parameters for computing effective ink concentration using Equation 4.

Subscripts in the dot gain parameters refer to the type of ink present in the Neugebauer primary. There are 32 different dot gain parameters in the model; each one is a vector, the dimension of which is the number of wavelengths considered. The area coverage of paper is computed as the difference between the sum of the area coverage of the inks and their overprints, with the constraint to be positive.

\[
a_0 = 1 - \sum_{p=1}^{15} a_p
\]  

(5)

To effectively tune the model, the training set must be customized to employ all the parameters. The training set consists therefore of ramps of eleven patches, ranging from the absence of ink to full ink coverage of cyan, magenta, yellow, red, green, blue, black, cyan with black, magenta with black, yellow with black, red with black, green with black and blue with black, for a total of 143 samples (Fig. 3).

Figure 3. (a) Speciment for the training set and Neugebauer primaries measurement. It consist of 133 patches. (b) Test set speciment; it consist of 777 patches.
3. THE GENETIC ALGORITHM

A genetic algorithm is used to estimate the printer model parameters described above. The choice of the representation of candidate solutions (individuals) and the fitness function used to evaluate individuals are crucial factors in the effectiveness of the genetic approach.

The genetic material of each individual, called the genome, must consist of the minimum amount of data requested to represent a solution to the problem.

In our work, a genome consists of an array of real numbers. In the printer model, we introduced:
- the Yule-Nielsen factor (Equation 1),
- U for the printer driver model (Equation 2),
- and 32 parameters for the dot gain functions (Table 2).

The dot gain parameters are wavelength-dependent, giving us a total of 258 real numbers.

As the genome is an array of real numbers, a range must be specified. We consider that the dot gain functions do not alter the theoretical value of ink concentration by more than some 30%: consequently the range for real parameters has been set at [0.3; 3.0]. The theoretical value of the Yule-Nielsen factor ranges from 1.0, corresponding to the absence of scattering, to 2.0, corresponding to Lambertian or perfectly diffused scattering, with the assumption that the dots are rectangular in cross section. In reality, the dots have soft transitions, and in cases of high frequency rotated screens, or error diffusion, much of the paper is covered by transitory regions. In these cases, if the Yule-Nielsen factor is experimentally computed, it may exceed the theoretical limit of 2. In our experiment, we have considered a range [1.0; 12.0] for the Yule-Nielsen factor.

The fitness function is computed as:

\[
\text{fitness} = \frac{1}{S} \left( \sum_{s=1}^{S} \frac{1}{\Gamma} \left( \sum_{\lambda=1}^{\Gamma} \left( R_{\text{print},\lambda,s} - R_{\text{meas},\lambda,s} \right)^2 \right) \right) \]

(6)

where \( S \) is the number of elements in the training set (\( S=143 \)), \( \Gamma \) the number of wavelengths (\( \Gamma = 8 \)), and \( R_{\text{print},\lambda,s} \) is computed with Equation 1.

We have used the ‘simple’ genetic algorithm in the Galib library. It employs non-overlapping populations: at each generation the algorithm creates an entirely new population of individuals by selecting from the previous population, and then mating to produce the new offspring. The best individual from each generation is also carried over into the next generation (elitism). The probability of mutation is set at 0.002, and crossover at 0.9. Selection is roulette wheel. The initial population of 12 individuals is randomly selected; only the initial value for the parameter U in the conversion from RGB to CMYK (Equation 2) is initialized at 1. The stopping criteria is the number of iterations performed: we have considered 4000 iterations.

4. EXPERIMENT AND RESULTS

We have applied our model for the characterization of three printers:
- Epson Stylus Color
- HP 2000C
- Epson Stylus C80

With the Epson Stylus Color we used a driver that employed Floyd Steinberg dithering, while for the Epson Stylus C80 and the HP 2000C, we used the drivers from the printer manufacturers, disabling any color management or color enhancement. Different types of papers were used (plain paper and Epson Photo Quality paper).

The characterization procedure started with the printing and measurement of the Neugebauer primaries and the training set (Fig. 3). The Neugebauer primaries were obtained by measuring the printed inks at full coverage, and their overprints, by successive prints on the same sheet.

Measurements of the spectra were taken with a Gretag Spectrolino, considering values in the wavelength range from 400 to 680 nm with a step of 40 nm. Reflectance spectra were in the range \([0;100]\).

The results are reported in terms of color difference in CIELAB \( \Delta E_{ab} \) and CIELAB \( \Delta E^*_{94} \), hue and lightness difference\(^{11}\), and root mean square error (Table 3 and 4).
This is particularly true using a stochastic method for parameter estimation. If this method were correct, we would expect to obtain a much better performance if we were able to reduce, or eliminate the effect of print enhancement that we hold attributable to the printer’s driver.

We note that the \( \Delta E^*_{94} \) values are small and quite similar among printers for the training set of data, while the results for the test set are more varied. The best performance was obtained with the Epson Stylus Color printer, the worst, with the HP 2000C. The results for the Epson Stylus C80 with different types of paper are, on the training and test set, quite similar. The model’s performance, we see, may be strongly influenced by the printer driver, to the point that it may be impossible for the training set to provide a full description of the printer’s behavior.

In a previous work\(^3\), we modeled the Epson Stylus Color printer following a different approach, based on the least-square estimation of the model’s parameters. Using that method, we obtained, for the same test specimen used here, a mean CIELAB \( \Delta E^*_{94} \) of 1.78 with maximum of 8.84; and a mean CIELAB \( \Delta E^*_{94} \) of 1.53 with a maximum of 6.41. The results of the two methods are therefore comparable.

## 5. CONCLUSION

In this work we experimented a novel approach to the spectral characterization of inkjet printers. Our objective was to see whether it is feasible to approach the complex problem of printer spectral modeling by introducing a set of parameters into the mathematical framework of the Yule Nielsen Spectral Neugebauer equation. Considering the complex interaction between these parameters, we also tested to see whether an acceptable solution could be computed using a stochastic optimization method. If this were so, using our method we would be able to spectrally characterize a generic inkjet printer with only about 100 measurements, performed manually by an operator, without any of the equipment needed to measure the large number of samples required for colorimetric printer characterization based on interpolation methods.

The results reported indicate that a characterization is possible, with mean \( \Delta E^*_{94} \) ranging from 1.64 to 4.09, and a maximum hue difference of 12.2.

Some key aspects must be noted. First, we have assumed that any overprinting of inks can be predicted on the basis of the model optimization performed on a small set of samples. If this method were not correct, we would expect to obtain large shifts in hue in the test set prediction. This is particularly true using a stochastic method for parameter estimation. Second, we have made no assumption concerning the sequence of printing. Third, the printers were treated as RGB devices, and we therefore included the printer-driver operations in our model. This has certainly penalized the results. We could expect to obtain a much better performance if we were able to reduce, or eliminate the effect of print enhancement that we hold attributable to the printer’s driver.

### Table 3. Statistics of color distances and spectra differences for the Training Set.

<table>
<thead>
<tr>
<th></th>
<th>Epson Stylus Color photo quality paper</th>
<th>HP 2000C plain paper</th>
<th>Epson Stylus C80 photo quality paper</th>
<th>Epson Stylus C80 plain paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta E^*_{ab} )</td>
<td>mean 1.49 max 6.47 s.dev. 1.15</td>
<td>mean 2.29 max 8.02 s.dev. 1.67</td>
<td>mean 2.19 max 5.47 s.dev. 1.49</td>
<td>mean 1.67 max 5.93 s.dev. 1.21</td>
</tr>
<tr>
<td>( \Delta H )</td>
<td>mean 0.70 max 3.67 s.dev. 0.74</td>
<td>mean 0.87 max 3.29 s.dev. 0.82</td>
<td>mean 0.93 max 4.71 s.dev. 0.86</td>
<td>mean 0.78 max 3.33 s.dev. 0.71</td>
</tr>
<tr>
<td>( \Delta L )</td>
<td>mean 0.52 max 3.27 s.dev. 0.56</td>
<td>mean 0.79 max 3.56 s.dev. 0.73</td>
<td>mean 0.84 max 3.32 s.dev. 0.86</td>
<td>mean 0.92 max 5.24 s.dev. 1.11</td>
</tr>
<tr>
<td>( \Delta E^*_{94} )</td>
<td>mean 1.18 max 4.85 s.dev. 0.91</td>
<td>mean 1.71 max 5.77 s.dev. 1.25</td>
<td>mean 1.70 max 4.17 s.dev. 1.14</td>
<td>mean 1.47 max 5.85 s.dev. 1.18</td>
</tr>
<tr>
<td>RMS</td>
<td>mean 0.65 max 2.34 s.dev. 0.41</td>
<td>mean 1.12 max 3.23 s.dev. 0.82</td>
<td>mean 0.92 max 2.66 s.dev. 0.55</td>
<td>mean 0.96 max 4.69 s.dev. 0.68</td>
</tr>
<tr>
<td>fitness</td>
<td>mean 0.59 max 1.92 s.dev. 1.15</td>
<td>mean 1.15 max 1.15</td>
<td>mean 1.38 max 1.38</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4. Statistics of the color distances and spectra differences for the Test Set.

<table>
<thead>
<tr>
<th></th>
<th>Epson Stylus Color photo quality paper</th>
<th>HP 2000C plain paper</th>
<th>Epson Stylus C80 photo quality paper</th>
<th>Epson Stylus C80 plain paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta E^*_{ab} )</td>
<td>mean 1.92 max 7.99 s.dev. 1.22</td>
<td>mean 5.4 max 18.5 s.dev. 3.12</td>
<td>mean 3.79 max 13.5 s.dev. 2.04</td>
<td>mean 2.74 max 8.51 s.dev. 1.45</td>
</tr>
<tr>
<td>( \Delta H )</td>
<td>mean 0.87 max 6.31 s.dev. 0.97</td>
<td>mean 2.01 max 12.2 s.dev. 1.88</td>
<td>mean 2.04 max 10.3 s.dev. 1.86</td>
<td>mean 0.97 max 6.00 s.dev. 0.76</td>
</tr>
<tr>
<td>( \Delta L )</td>
<td>mean 0.9 max 3.41 s.dev. 0.71</td>
<td>mean 1.74 max 6.62 s.dev. 1.31</td>
<td>mean 1.48 max 5.11 s.dev. 1.22</td>
<td>mean 1.74 max 7.98 s.dev. 1.56</td>
</tr>
<tr>
<td>( \Delta E^*_{94} )</td>
<td>mean 1.64 max 6.03 s.dev. 0.95</td>
<td>mean 4.09 max 11.0 s.dev. 1.95</td>
<td>mean 3.17 max 9.48 s.dev. 1.61</td>
<td>mean 2.51 max 8.45 s.dev. 1.42</td>
</tr>
<tr>
<td>RMS</td>
<td>mean 0.90 max 5.82 s.dev. 0.56</td>
<td>mean 2.24 max 12.2 s.dev. 1.34</td>
<td>mean 1.47 max 4.45 s.dev. 0.61</td>
<td>mean 1.72 max 5.37 s.dev. 1.15</td>
</tr>
</tbody>
</table>
REFERENCES